13 # Estadística bidimensional

En context (pàg. 315)

Resposta oberta a manera de reflexió, que pot servir d'introducció a l'estadística bidimensional. Les preguntes es poden contestar observant el gràfic i analitzant la notícia que s'explica.

Llenguatge matemàtic (pàg. 321)

— El coeficient de correlació de Pearson no ens ha de donar informació sobre la mida de les nostres variables, sinó sobre el grau en què s'assemblen, per això és important que no siguin proporcionals a les variables ni tinguin unitats.

Amplía (pàg. 322)

— Si les variables tenen dependència funcional, $r = \pm 1$.

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \pm 1 \Rightarrow \sigma_{XY} = \pm(\sigma_X \cdot \sigma_Y)$$

Ara, a partir de les rectes de regressió tenim:

$$\left.\begin{aligned} y - \overline{Y} &= \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \\ x - \overline{X} &= \frac{\sigma_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \left.\begin{aligned} y - \overline{Y} &= \frac{\pm(\sigma_X \cdot \sigma_Y)}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \\ x - \overline{X} &= \frac{\pm(\sigma_X \cdot \sigma_Y)}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \left.\begin{aligned} y - \overline{Y} &= \frac{\pm\sigma_Y}{\sigma_X} \cdot \left(x - \overline{X}\right) \\ x - \overline{X} &= \frac{\pm\sigma_X}{\sigma_Y} \cdot \left(y - \overline{Y}\right) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \left.\begin{aligned} \sigma_X \cdot \left(y - \overline{Y}\right) &= \pm\sigma_Y \cdot \left(x - \overline{X}\right) \\ \sigma_Y \cdot \left(x - \overline{X}\right) &= \pm\sigma_X \cdot \left(y - \overline{Y}\right) \end{aligned}\right\}$$

Observem que les dues rectes coincideixen.

— Si les variables són independents, $r = 0$:

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = 0 \Rightarrow \sigma_{XY} = 0$$

Ara, a partir de les rectes de regressió tenim:

$$\left.\begin{aligned} y - \overline{Y} &= \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \\ x - \overline{X} &= \frac{\sigma_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \end{aligned}\right\} \Rightarrow \left.\begin{aligned} y - \overline{Y} &= 0 \cdot \left(x - \overline{X}\right) \\ x - \overline{X} &= 0 \cdot \left(y - \overline{Y}\right) \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \left.\begin{aligned} y &= \overline{Y} \\ x &= \overline{X} \end{aligned}\right\}$$

Observem que les rectes resultants són paral·leles als eixos de coordenades i, en conseqüència, perpendiculars entre elles.

— Veiem que, si $r = \pm 1$ les rectes de correlació són coincidents, mentre que, si $r = 0$ les rectes de correlació són perpendiculars entre si. Podem observar que, a mesura que augmenti el valor de r (o disminueixi entre 0 i −1) les dues rectes de regressió s'aniran apropant.

Calculadora (pàg. 323)

— Resposta suggerida per a una calculadora del tipus Casio fx-82ES. En el vídeo següent s'explica com es fa servir aquesta calculadora per a resoldre problemes d'estadística bidimensional:

http://links.edebe.com/gbxruc

Problemes resolts (pàgs. 324 a 326)

1. Calculem les marginals de X i de Y, en què Y = edat i X = temps. En aquestes taules, calcularem també les marques de classe.

Marginal de X:

	0,5	1,5	2,5	3,5	4,5
X	0-1	1-2	2-3	3-4	4-5
f_i	31	30	22	11	6

Marginal de Y:

	7,5	17,5	22,5	27,5	32,5	37,5	57,5
Y	0-15	15-20	20-25	25-30	30-35	35-40	40-75
f_i	15	14	14	14	14	14	15

Ara, calculem la covariància:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i \cdot f_i =$$

$$= \frac{(0,5 \cdot 31 + 1,5 \cdot 30 + 2,5 \cdot 22 + 3,5 \cdot 11 + 4,5 \cdot 6)}{100} =$$

$$= \frac{181}{100} = 1{,}81$$

$$\overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j \cdot f_j =$$

$$= \frac{7{,}5 \cdot 15 + 17{,}5 \cdot 14 + 22{,}5 \cdot 14 + 27{,}5 \cdot 14 +}{100}$$

$$\frac{+\, 32{,}5 \cdot 14 + 37{,}5 \cdot 14 + 57{,}5 \cdot 15}{100} = \frac{2\,900}{100} = 29$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n}\sum_{j}^{m} x_i \cdot y_j \cdot f_{ij} - \overline{X} \cdot \overline{Y} =$$

$$= \frac{1}{100}(0,5 \cdot 7,5 \cdot 7 + 0,5 \cdot 17,5 \cdot 5 + ... + 4,5 \cdot 57,5 \cdot 0) -$$

$$-1{,}81 \cdot 29 = -0{,}115$$

La covariància és negativa i de valor petit. Podem concloure que els grups amb més edat dediquen menys temps a l'ordinador.

2. a) Les rectes de regressió es tallen en el centre de gravetat d'una distribució, és a dir, en el punt $(\overline{X}, \overline{Y})$. Plantegem el sistema següent i el resolem:

$$\left.\begin{array}{l} y = 1{,}5x - 1 \\ x = 0{,}3y + 0{,}5 \end{array}\right\} \Rightarrow x = 0{,}36;\ y = -0{,}45$$

Per tant, la solució del sistema anterior coincideix amb les mitjanes de X i de Y. És a dir: $\overline{X} = 0{,}36$; $\overline{Y} = -0{,}45$.

b) Suposem que la recta r és la regressió de Y sobre X ja que tenim aïllada la y. D'altra banda, suposem que s és la regressió de X sobre Y ja que en aquesta equació està aïllada la x. Si això és cert, tindrem que:

$$\frac{\sigma_{XY}}{\sigma_X{}^2} = 1,5;\ \frac{\sigma_{XY}}{\sigma_Y{}^2} = 0,3$$

El coeficient de Pearson és:

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \Rightarrow r^2 = \frac{\sigma_{XY}{}^2}{\sigma_X{}^2 \cdot \sigma_Y{}^2} \Rightarrow r = \sqrt{\frac{\sigma_{XY}}{\sigma_X{}^2} \cdot \frac{\sigma_{XY}}{\sigma_Y{}^2}}$$

Substituïm les dades que hem obtingut:

$$r = \sqrt{\frac{\sigma_{XY}}{\sigma_X{}^2} \cdot \frac{\sigma_{XY}}{\sigma_Y{}^2}} = \sqrt{1{,}5 \cdot 0{,}3} = 0{,}67 < 1$$

Com que aquest valor és més petit que 1, aleshores el que havíem suposat és cert. És a dir, la recta r és la regressió de Y sobre X i la recta s és la regressió de X sobre Y.

c) El coeficient de correlació és el valor que hem calculat en l'apartat b), així que, $r = 0{,}67$. Com que aquest valor és més proper a 1 que a 0, podem concloure que la regressió és bona.

3. a) Coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{10{,}5}{6} = 1{,}75;\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{22}{6} = 3{,}\hat{6}$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{19{,}75}{6} - 1{,}75^2 = 0{,}229\,1\hat{6}$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{138}{6} - 3{,}\hat{6}^2 = 9{,}\hat{5}$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{46{,}5}{6} - 1{,}75 \cdot 3{,}\hat{6} = 1{,}\hat{3}$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{1{,}\hat{3}}{\sqrt{0{,}229\,1\hat{6} \cdot 9{,}\hat{5}}} = 0{,}9$$

b) Recta de regressió de X sobre Y:

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \Rightarrow x - 1{,}75 = \frac{1{,}\hat{3}}{9{,}\hat{5}} \cdot \left(y - 3{,}\hat{6}\right) \Rightarrow$$

$$\Rightarrow x - 1{,}75 = 0{,}14 \cdot \left(y - 3{,}\hat{6}\right) \Rightarrow x = 1{,}24 + 0{,}14y$$

c) Sigui $y = 5$; introduïm aquesta dada en l'equació resultant de l'apartat b).

$$x = 1{,}24 + 0{,}14 \cdot 5 = 1{,}94$$

Per tant, si una persona fa 5 anys que treballa a l'empresa el seu salari serà de 1 940 €.

Exercicis i problemes (pàgs. 327 a 332)

1 VARIABLE ESTADÍSTICA BIDIMENSIONAL

Pàg. 327

4. a) El pes i l'alçada d'una persona són variables quantitatives i contínues.

b) L'hora del dia i la temperatura en aquesta hora són variables quantitatives i contínues.

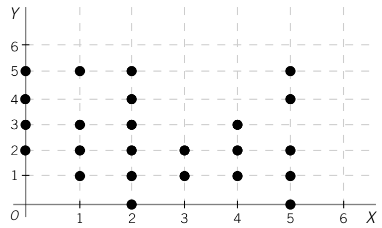
c) Les variables estadístiques província i color de cotxe més venut són variables qualitatives.

d) El nombre d'habitants i de televisors en una casa són variables quantitatives i discretes.

5. a) La taula de doble entrada és la següent:

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

b)



6. a) Calculem les marques de classe:

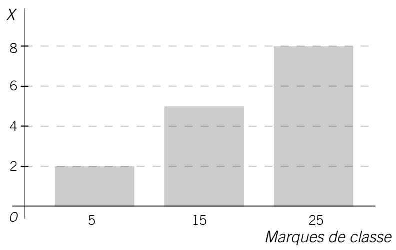
$$x_1 = \frac{0+10}{2} = 5;\ x_2 = \frac{10+20}{2} = 15;$$

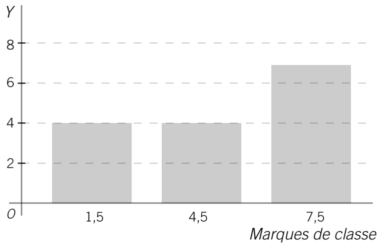
$$x_3 = \frac{20+30}{2} = 25$$

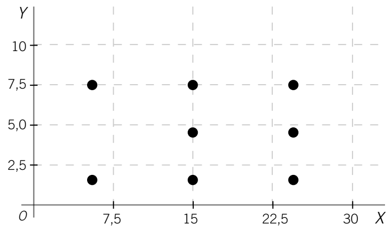
$$y_1 = \frac{0+3}{2} = 1{,}5;\ y_2 = \frac{3+6}{2} = 4{,}5;$$

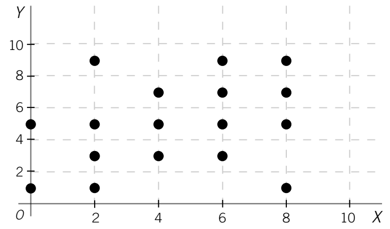
$$y_3 = \frac{6+9}{2} = 7{,}5$$

b)





c) 

7. a) 

b) Marginal de X:

X	0	2	4	6	8
f_i	2	7	9	11	5

Marginal de Y:

Y	1	3	5	7	9
f_j	5	7	8	11	3

c) Mitjana de X:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i \cdot f_i =$$

$$= \frac{1}{34}\cdot(0\cdot 2 + 2\cdot 7 + 4\cdot 9 + 6\cdot 11 + 8\cdot 5) = 4{,}588$$

Mitjana de Y:

$$\overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j \cdot f_j =$$

$$= \frac{1}{34}\cdot(1\cdot 5 + 3\cdot 7 + 5\cdot 8 + 7\cdot 11 + 9\cdot 3) = 5$$

d) Desviació típica de X:

$$\sigma_X = \sqrt{\frac{1}{N}\sum_i f_i \cdot x_i^2 - \overline{X}^2} =$$

$$= \sqrt{\frac{1}{34}\left(2\cdot 0^2 + 7\cdot 2^2 + 9\cdot 4^2 + 11\cdot 6^2 + 5\cdot 8^2\right) - 4{,}588^2} =$$

$$= 2{,}25$$

Desviació típica de Y:

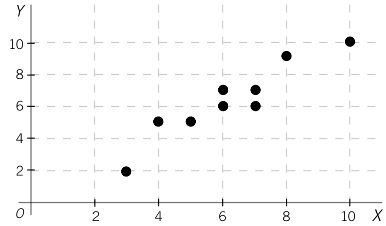
$$\sigma_Y = \sqrt{\frac{1}{M}\sum_j f_j \cdot y_j^2 - \overline{Y}^2} =$$

$$= \sqrt{\frac{1}{34}\left(5\cdot 1^2 + 7\cdot 3^2 + 8\cdot 5^2 + 11\cdot 7^2 + 3\cdot 9^2\right) - 5^2} =$$

$$= 2{,}42$$

8. a) La taula de doble entrada corresponent és:

X \ Y	2	3	4	5	6	7	8	9	10
3	4	0	0	0	0	0	0	0	0
4	0	0	0	6	0	0	0	0	0
5	0	0	0	12	0	0	0	0	0
6	0	0	0	0	4	5	0	0	0
7	0	0	0	0	4	2	0	0	0
8	0	0	0	0	0	0	0	1	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	2

b) 

c) A partir del diagrama de dispersió, observem que les variables són de dependència lineal positiva forta.

2 RELACIÓ ENTRE VARIABLES ESTADÍSTIQUES: CORRELACIÓ Pàgs. 327 i 328

9. a) El pes i l'altura són variables amb dependència estadística.

b) El radi i la longitud d'una circumferència són variables amb dependència funcional.

c) La velocitat d'un cotxe i la despesa de la gasolina són variables amb dependència estadística.

d) La quantitat de pluja i la renda anual són variables independents, ja que no tenen res en comú.

10. Efectuem els càlculs en la taula que necessitem per a les qüestions següents:

X \ Y	1	1,5	2	3	f_i	$X_i \cdot f_i$	$X_i^2 \cdot f_i$
12	12	3	1	0	16	192	2304
15	6	4	4	1	15	225	3375
20	2	2	4	4	12	240	4800
25	0	1	1	5	7	175	4375
f_i	20	10	10	10	50	832	14854
$Y_j \cdot f_j$	20	15	20	30	85		
$Y_j^2 \cdot f_j$	20	22,5	40	90	172,5		

a) Mitjana de X: $\overline{X} = \frac{1}{50}\sum_{i=1}^{n} x_i \cdot f_i = \frac{832}{50} = 16{,}64$

Mitjana de Y: $\overline{Y} = \dfrac{1}{50}\sum_{j=1}^{m} y_j \cdot f_j = \dfrac{85}{50} = 1{,}7$

b) Variància de X:

$$\sigma_X{}^2 = \frac{1}{N}\sum_i f_i \cdot x_i^2 - \overline{X}^2 = \frac{14\,854}{50} - 16{,}64^2 = 20{,}19$$

Variància de Y:

$$\sigma_Y{}^2 = \frac{1}{M}\sum_j f_j \cdot y_j^2 - \overline{Y}^2 = \frac{172{,}5}{50} - 1{,}7^2 = 0{,}56$$

c) La covariància entre les dues variables és la següent:

$$\sigma_{XY} = \frac{1}{N}\sum_i^n\sum_j^m x_i \cdot y_j \cdot f_{ij} - \overline{X}\cdot\overline{Y} =$$

$$= \frac{1}{50}(12\cdot 1\cdot 12 + 12\cdot 1{,}5\cdot 3 + \ldots + 25\cdot 3\cdot 5) -$$

$$- 16{,}64\cdot 1{,}7 = 2{,}302$$

11. Busquem les dades necessaries per a trobar el coeficient de correlació.

X	Y	X^2	Y^2	$X\cdot Y$
0	1	0	1	0
0	2	0	4	0
1	2	1	4	2
1	3	1	9	3
1	3	1	9	3
2	3	4	9	6
2	1	4	1	2
2	2	4	4	4
3	4	9	16	12
3	4	9	16	12
15	25	33	73	44

$$\overline{X} = \frac{15}{10} = 1{,}5;\ \overline{Y} = \frac{25}{10} = 2{,}5$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{33}{10} - 1{,}5^2 = 1{,}05$$

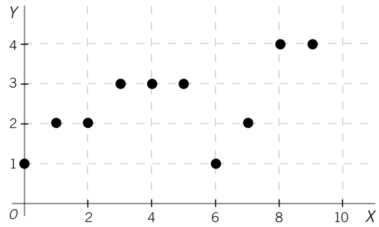
$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{73}{10} - 2{,}5^2 = 1{,}05$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{44}{10} - 1{,}5\cdot 2{,}5 = 0{,}65$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{0{,}65}{\sqrt{1{,}05\cdot 1{,}05}} = 0{,}62$$

— El valor del coeficient de correlació s'aproxima més a 1 que a 0; per tant, les variables tenen una forta dependència lineal positiva.

12. a)



b) Per a calcular la covariància, necessitem uns càlculs previs:

	X	Y	$X\cdot Y$
	0	1	0
	1	2	2
	2	2	4
	3	3	9
	4	3	12
	5	3	15
	6	1	6
	7	2	14
	8	4	32
	9	4	36
Sumes	45	25	130

$$\overline{X} = \frac{45}{10} = 4{,}5;\ \overline{Y} = \frac{25}{10} = 2{,}5$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{130}{10} - 4{,}5\cdot 2{,}5 = 1{,}75$$

13. a) Marginal de X:

X	0	1	2	3	4	5	6
f_i	58	64	41	29	22	11	5

Marginal de Y:

Y	0	1	2	3	4	5	6
f_j	58	66	57	24	12	9	4

b) Calculem el coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i\cdot f_i = \frac{406}{230} = 1{,}77;$$

$$\overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j\cdot f_j = \frac{369}{230} = 1{,}6$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{91}{230} - 1{,}77^2 = -2{,}74$$

$$\sigma_Y^2 = \frac{1}{M}\sum_i Y_i^2 - \overline{Y}^2 = \frac{91}{230} - 1{,}6^2 = -2{,}16$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n\sum_j^m x_i\cdot y_j\cdot f_{ij} - \overline{X}\cdot\overline{Y} = \frac{971}{230} - 1{,}77\cdot 1{,}6 = 1{,}4$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{1{,}4}{\sqrt{(-2{,}74)\cdot(-2{,}16)}} = 0{,}6$$

c) La relació que hi ha entre aquestes dues variables és de dependència lineal forta i positiva, ja que el valor del coeficient de correlació s'aproxima més a 1 que a 0.

14. Calculem la covariància entre les variables:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{67{,}63}{8} = 8{,}453\,75$$

$$\overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j = \frac{-73{,}2}{8} = -9{,}15$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i \cdot Y_i - \overline{X} \cdot \overline{Y} =$$

$$= \frac{-12\,077{,}19}{8} - 8{,}453\,75 \cdot (-9{,}15) = -1\,432{,}29$$

15. a) Construïm la taula de doble entrada:

X \ Y	0	1	2
0	7	6	0
1	7	5	1
2	0	2	1
3	0	0	1

b) Marginal de la variable germans, X:

X	0	1	2	3
f_i	13	13	3	1

Marginal de la variable germanes, Y:

Y	0	1	2
f_j	14	13	3

c) Trobem el valor de la covariància entre les variables:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i \cdot f_i = \frac{1}{30}(0 \cdot 13 + 1 \cdot 13 + 2 \cdot 3 + 3 \cdot 1) =$$

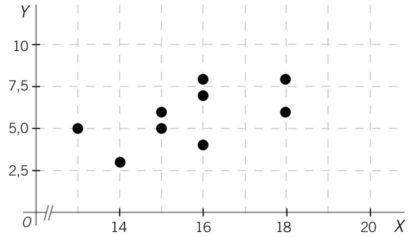
$$= \frac{22}{30} = 0{,}73$$

$$\overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j \cdot f_j = \frac{1}{30}(0 \cdot 14 + 1 \cdot 13 + 2 \cdot 3) = \frac{19}{30} = 0{,}63$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n}\sum_{j}^{m} x_i \cdot y_j \cdot f_{ij} - \overline{X} \cdot \overline{Y} =$$

$$= \frac{1}{30}(0 \cdot 0 \cdot 7 + 0 \cdot 1 \cdot 6 + \ldots + 3 \cdot 2 \cdot 1) - 0{,}73 \cdot 0{,}63 = 0{,}24$$

16. El núvol de punts corresponent és el següent:



Ara, calculem la covariància entre les dues variables:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{155}{10} = 15{,}5; \ \overline{Y} = \frac{1}{N}\sum_{j=1}^{m} y_j = \frac{55}{10} = 5{,}5$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{870}{10} - 15{,}5 \cdot 5{,}5 = 1{,}75$$

17. Utilitzant les funcions estadístiques de la calculadora, resulten les mateixes solucions que en l'exercici anterior.

18. La taula de doble entrada és la següent:

X \ Y	5	6	7	8	9	f_i	$X_i \cdot f_i$
5	0	1	5	0	3	9	45
6	1	2	2	3	0	8	48
7	2	1	2	1	0	6	42
8	1	2	2	1	0	6	48
9	1	3	2	1	0	7	63
f_i	5	9	13	6	3	36	246
$Y_i \cdot f_i$	25	54	91	48	27	245	

Marginal de X:

X	5	6	7	8	9
f_i	9	8	6	6	7

Marginal de Y:

Y	5	6	7	8	9
f_i	5	9	13	6	3

Covariància entre les variables:

$$\overline{X} = \frac{246}{36} = 6{,}83; \ \overline{Y} = \frac{245}{36} = 6{,}81$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n}\sum_{j}^{m} x_i \cdot y_j \cdot f_{ij} - \overline{X} \cdot \overline{Y} =$$

$$= \frac{1}{36}(5 \cdot 5 \cdot 0 + 5 \cdot 6 \cdot 1 + \ldots + 9 \cdot 9 \cdot 0) - 6{,}83 \cdot 6{,}81 = -0{,}59$$

19. a) Si observem el gràfic, veiem que les variables tenen dependència lineal positiva feble, de manera que el coeficient de correlació és més gran que zero però no gaire proper a 1.

b) En aquest cas, les variables tenen dependència lineal negativa; per tant, el coeficient de correlació és negatiu i intermedi.

c) En aquest últim gràfic, la dependència entre les variables és lineal negativa i forta. Així que el coeficient de correlació és proper a –1.

20. Calculem les dades necessàries per a trobar el coeficient de correlació:

X \ Y	1	2	3	4	5	f_j	$X_i \cdot f_i$	$X_i^2 \cdot f_i$
1	0	2	5	7	11	25	25	25
2	5	1	0	2	3	11	22	44
3	2	4	1	8	0	15	45	135
4	3	0	2	0	1	6	24	96
5	1	5	6	2	7	21	105	525
f_i	11	12	14	19	22	78	221	825
$Y_i \cdot f_i$	11	24	42	76	110	263		
$Y_i^2 \cdot f_i$	11	48	126	304	550	1 039		

$$\overline{X} = \frac{221}{78} = 2{,}8\widehat{3};\ \overline{Y} = \frac{263}{78} = 3{,}37179$$

$$\sigma_X{}^2 = \frac{1}{N}\sum_i f_i \cdot x_i^2 - \overline{X}^2 = \frac{825}{78} - 2{,}8\widehat{3}^2 = 2{,}549$$

$$\sigma_Y{}^2 = \frac{1}{M}\sum_j f_j \cdot y_j^2 - \overline{Y}^2 = \frac{1039}{78} - 3{,}37179^2 = 1{,}951$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n \sum_j^m x_i \cdot y_j \cdot f_{ij} - \overline{X} \cdot \overline{Y} =$$

$$= \frac{1}{78}(1 \cdot 1 \cdot 0 + 1 \cdot 2 \cdot 2 + \ldots + 5 \cdot 5 \cdot 7) - 2{,}8\widehat{3} \cdot$$

$$\cdot\, 3{,}37179 = -0{,}4$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{-0{,}4}{\sqrt{2{,}549 \cdot 1{,}951}} = -0{,}18$$

— El valor del coeficient de correlació és negatiu i baix, de manera que la dependència és feble i de signe negatiu (quan una de les variables augmenta, l'altra disminueix).

3 REGRESSIÓ LINEAL Pàgs. 329 a 331

21. a) Les rectes de regressió són decreixents; per tant, el signe de la correlació és negatiu. D'altra banda, no es tracta de variables independents ni fortament correlacionades, sinó d'un cas intermedi.

b) Aquestes dues rectes no representen rectes de regressió, ja que són perpendiculars entre elles, però no són paral·leles als eixos de coordenades.

c) Les rectes de regressió són perpendiculars entre elles i paral·leles als eixos de coordenades, de manera que les variables són independents.

d) Les rectes de regressió són bastant a prop; per tant, el coeficient de correlació és proper a 1, la qual cosa implica que la dependència entre les variables és forta. Com que les dues rectes són creixents, el signe de la correlació és positiu.

22. a) La fórmula de la recta de regressió de Y sobre X és $y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right)$. Calculem les dades que necessitem per a aquesta recta:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^n x_i = \frac{60}{5} = 12;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^m y_j = \frac{6{,}1}{5} = 1{,}22$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{80{,}2}{5} - 12 \cdot 1{,}22 = 1{,}4$$

$$\sigma_X^2 = \frac{1}{M}\sum_{i=1}^m x_i{}^2 - \overline{X}^2 = \frac{760}{5} - 12^2 = 8$$

$$y - 1{,}22 = \frac{1{,}4}{8} \cdot (x - 12) \Rightarrow y - 1{,}22 = 0{,}175 \cdot (x - 12) \Rightarrow$$

$$\Rightarrow y = -0{,}88 + 0{,}175x$$

b) El coeficient de regressió és el pendent de la recta anterior; és a dir, $b = \frac{\sigma_{XY}}{\sigma_X^2} = 0{,}175$.

c) Calculem el coeficient de correlació:

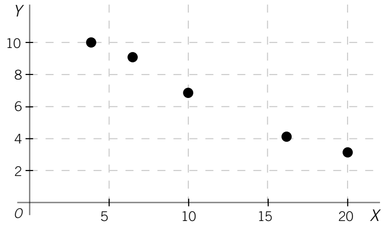
$$\sigma_Y^2 = \frac{1}{M}\sum_{j=1}^m y_j{}^2 - \overline{Y}^2 = \frac{8{,}83}{5} - 1{,}22^2 = 0{,}2776$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{1{,}4}{\sqrt{8 \cdot 0{,}2776}} = 0{,}94$$

d) El coeficient de determinació és el següent:

$$r^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \cdot \sigma_Y^2} = \frac{1{,}96}{8 \cdot 0{,}2776} = 0{,}88$$

23. a)



b) Recta de regressió de X sobre Y:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^n x_i = \frac{57}{5} = 11{,}4;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^m y_j = \frac{33}{5} = 6{,}6$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{297}{5} - 11{,}4 \cdot 6{,}6 = -15{,}84$$

$$\sigma_Y^2 = \frac{1}{M}\sum_{j=1}^m y_j{}^2 - \overline{Y}^2 = \frac{255}{5} - 6{,}6^2 = 7{,}44$$

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \Rightarrow x - 11{,}4 = \frac{-15{,}84}{7{,}44} \cdot (y - 6{,}6) \Rightarrow$$

$$\Rightarrow x - 11{,}4 = -2{,}13 \cdot (y - 6{,}6) \Rightarrow$$

$$\Rightarrow x = 25{,}45 - 2{,}13y$$

c) Recta de regressió de Y sobre X:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^n x_i = \frac{57}{5} = 11{,}4;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^m y_j = \frac{33}{5} = 6{,}6$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{297}{5} - 11{,}4 \cdot 6{,}6 = -15{,}84$$

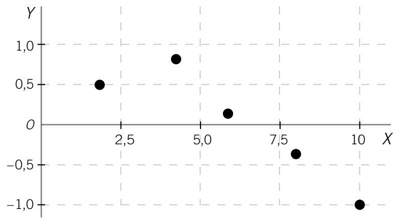
$$\sigma_X^2 = \frac{1}{M}\sum_{i=1}^m x_i{}^2 - \overline{X}^2 = \frac{821}{5} - 11{,}4^2 = 34{,}24$$

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 6{,}6 = \frac{-15{,}84}{34{,}24} \cdot (x - 11{,}4) \Rightarrow$$

$$\Rightarrow y - 6{,}6 = -0{,}46 \cdot (x - 11{,}4) \Rightarrow y = 11{,}87 - 0{,}46x$$

24. a)



b) Recta de regressió de Y sobre X:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{30}{5} = 6;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j = \frac{0}{5} = 0$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i \cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{-8{,}4}{5} - 6\cdot 0 = -1{,}68$$

$$\sigma_X^2 = \frac{1}{M}\sum_{i=1}^{m} x_i^2 - \overline{X}^2 = \frac{220}{5} - 6^2 = 8$$

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow y - 0 = \frac{-1{,}68}{8}\cdot(x-6) \Rightarrow$$

$$\Rightarrow y = -0{,}21\cdot(x-6) \Rightarrow y = 1{,}26 - 0{,}21x$$

c) Recta de regressió de X sobre Y:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{30}{5} = 6;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j = \frac{0}{5} = 0$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i \cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{-8{,}4}{5} - 6\cdot 0 = -1{,}68$$

$$\sigma_Y^2 = \frac{1}{M}\sum_{j=1}^{m} y_j^2 - \overline{Y}^2 = \frac{2{,}06}{5} - 0^2 = 0{,}412$$

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2}\cdot\left(y - \overline{Y}\right) \Rightarrow x - 6 = \frac{-1{,}68}{0{,}412}\cdot(y-0) \Rightarrow$$

$$\Rightarrow x - 6 = -2{,}13y \Rightarrow x = 6 - 4{,}078y$$

25. a) Si observem el gràfic, veiem que les variables tenen una dependència lineal positiva forta, la qual cosa implica que les rectes de regressió són molt a prop entre elles.

b) Les variables d'aquest gràfic tenen una dependència lineal negativa forta, la qual cosa implica que les rectes de regressió són molt properes entre si.

c) Aquestes dues variables estan relacionades entre elles en forma de corba. Aquesta corba és clarament decreixent; si calculéssim l'índex de correlació, obtindríem un valor negatiu i no gaire proper a –1, és a dir, tindríem una relació negativa feble. Les rectes de regressió estarien allunyades.

d) Aquí veiem que les variables són independents; per tant, les rectes de regressió són perpendiculars i paral·leles als eixos de coordenades.

e) Aquestes dues variables estan relacionades entre elles en forma de corba. Aquesta corba és clarament creixent; si calculéssim l'índex de correlació, obtindríem un valor negatiu i proper a 1; és a dir, tindríem una relació positiva forta. Les rectes de regressió serien a prop.

f) Les variables tenen una relació de dependència funcional i a més és lineal; les rectes de regressió coincideixen.

26. a) Trobem el coeficient de correlació, en què X = anys i Y = edat.

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{98}{6} = 16{,}\hat{3};\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{457}{6} = 76{,}1\hat{6}$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{2\,634}{6} - 16{,}\hat{3}^2 = 172{,}\hat{2}$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{34\,987}{6} - 76{,}1\hat{6}^2 = 29{,}80\hat{5}$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{7\,070}{6} - 16{,}\hat{3}\cdot 76{,}1\hat{6} = -65{,}7\hat{2}$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{-65{,}7\hat{2}}{\sqrt{172{,}\hat{2}\cdot 29{,}80\hat{5}}} = -0{,}917$$

b) Busquem la recta de regressió de l'edat en funció dels anys fumant:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 76{,}1\hat{6} = \frac{-65{,}7\hat{2}}{172{,}\hat{2}}\cdot\left(x - 16{,}\hat{3}\right) \Rightarrow$$

$$\Rightarrow y - 76{,}1\hat{6} = 0{,}38\cdot\left(x - 16{,}\hat{3}\right) \Rightarrow y = 82{,}4 - 0{,}38x$$

27. a) Calculem el coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{164}{10} = 16{,}4;\ \overline{Y} = \frac{1}{N}\sum_i Y_i = \frac{149}{10} = 14{,}9$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{2\,848}{10} - 16{,}4^2 = 15{,}84$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{2\,265}{10} - 14{,}9^2 = 4{,}49$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{2\,364}{10} - 16{,}4\cdot 14{,}9 = -7{,}96$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{-7{,}96}{\sqrt{15{,}84\cdot 4{,}49}} = -0{,}94$$

Com que aquest valor és proper a –1, podem concloure que les variables tenen una forta dependència lineal.

b) Trobem la recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 14{,}9 = \frac{-7{,}96}{15{,}84}\cdot(x - 16{,}4) \Rightarrow$$

$$\Rightarrow y - 14{,}9 = -0{,}5\cdot(x - 16{,}4) \Rightarrow y = 23{,}1 - 0{,}5x$$

28. Per a saber quina d'aquestes distribucions té una relació més forta, hem de calcular els coeficients de correlació de cada cas.

a) $r_a = \dfrac{\sigma_{XY}}{\sigma_X\cdot\sigma_Y} = \dfrac{0{,}8}{0{,}95\cdot 0{,}89} = 0{,}946$

b) $y = 0{,}81x + 0{,}5 \Rightarrow \dfrac{\sigma_{XY}}{\sigma_x^2} = 0{,}81 \Rightarrow \sigma_x^2 = \dfrac{\sigma_{XY}}{0{,}81}$

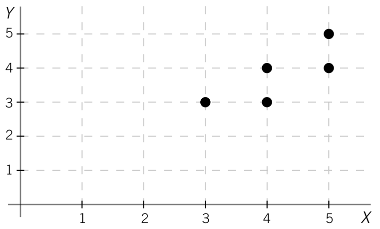
$$x = 1{,}12y - 0{,}7 \Rightarrow \frac{\sigma_{XY}}{\sigma_Y^2} = 1{,}12 \Rightarrow \sigma_Y^2 = \frac{\sigma_{XY}}{1{,}12}$$

$$r_b = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{\sigma_{XY}}{\sqrt{\dfrac{\sigma_{XY}}{0{,}81}\cdot\dfrac{\sigma_{XY}}{1{,}12}}} =$$

$$= \frac{\sigma_{XY}}{\dfrac{\sigma_{XY}}{0{,}952}} = 0{,}952$$

La relació entre les variables serà més forta quan el valor del coeficient de correlació sigui més a prop d'1; en aquest cas, el coeficient de correlació de l'apartat b) és més gran que el de l'apartat a); per tant, la relació entre les variables és més forta en b).

29. a)



b) Recta de regressió de X sobre Y:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{25}{6} = 4{,}1\hat{6};\ \overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j = \frac{23}{6} = 3{,}8\hat{3}$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i \cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{98}{6} - 4{,}1\hat{6}\cdot 3{,}8\hat{3} = 0{,}365$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{107}{6} - 4{,}1\hat{6}^2 = 0{,}477$$

$$\sigma_Y^2 = \frac{1}{M}\sum_{j=1}^{m} y_j^2 - \overline{Y}^2 = \frac{91}{6} - 3{,}8\hat{3}^2 = 0{,}477$$

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2}\cdot\left(y - \overline{Y}\right) \Rightarrow$$

$$\Rightarrow x - 4{,}1\hat{6} = \frac{0{,}365}{0{,}477}\cdot\left(y - 3{,}8\hat{3}\right) \Rightarrow$$

$$\Rightarrow x - 4{,}1\hat{6} = 0{,}76\cdot\left(y - 3{,}8\hat{3}\right) \Rightarrow x = 1{,}24 + 0{,}76y$$

Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow y - 3{,}8\hat{3} = \frac{0{,}365}{0{,}477}\cdot\left(x - 4{,}1\hat{6}\right) \Rightarrow$$

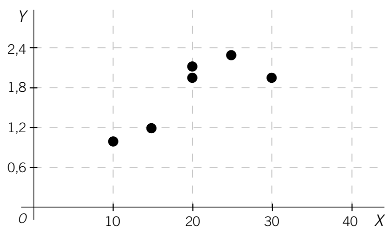
$$\Rightarrow y - 3{,}8\hat{3} = 0{,}76\cdot\left(x - 4{,}1\hat{6}\right) \Rightarrow y = 0{,}65 + 0{,}76x$$

c) $r = \dfrac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \dfrac{0{,}365}{\sqrt{0{,}477\cdot 0{,}477}} = 0{,}76$

d) El coeficient de correlació és més proper a 1 que a 0; per tant, podem dir que l'ajust realitzat és bastant bo.

30. Si utilitzem les eines de la pàgina web indicada, surten els mateixos resultats que en l'apartat anterior, encara que agafant més decimals dels resultats.

31. a)



b) Recta de regressió de Y sobre X:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{120}{6} = 20;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j = \frac{10{,}5}{6} = 1{,}75$$

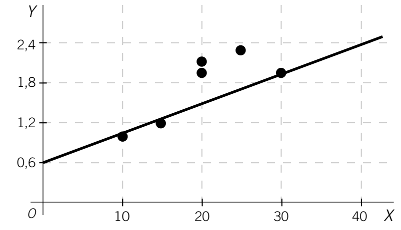
$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{225}{6} - 20\cdot 1{,}75 = 2{,}5$$

$$\sigma_X^2 = \frac{1}{M}\sum_{i=1}^{m} x_i^2 - \overline{X}^2 = \frac{2\,650}{6} - 20^2 = 41{,}67$$

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 1{,}75 = \frac{2{,}5}{41{,}67}\cdot\left(x - 20\right) \Rightarrow$$

$$\Rightarrow y - 1{,}75 = 0{,}06\cdot\left(x - 20\right) \Rightarrow y = 0{,}55 + 0{,}06x$$



32. a) Calculem el coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{1\,776}{10} = 177{,}6;$$

$$\overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{1\,761}{10} = 176{,}1$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{315\,970}{10} - 177{,}6^2 = 55{,}24$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{310\,445}{10} - 176{,}1^2 = 33{,}29$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} =$$

$$= \frac{313\,145}{10} - 177{,}6\cdot 176{,}1 = 39{,}14$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{39{,}14}{\sqrt{55{,}24\cdot 33{,}29}} = 0{,}91$$

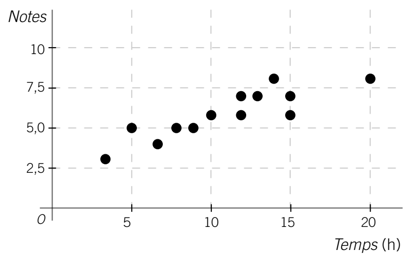
b) Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 176{,}1 = \frac{39{,}14}{55{,}24}\cdot\left(x - 177{,}6\right) \Rightarrow$$

$$\Rightarrow y - 176{,}1 = 0{,}7\cdot\left(x - 177{,}6\right) \Rightarrow y = 50{,}26 + 0{,}7x$$

33. a)



b) A partir del diagrama de dispersió, comprovem que la relació entre les hores estudiades i la mitjana de les notes s'aproxima bastant a una recta, de manera que sí, és apropiat.

c) Busquem la recta de regressió de Y sobre X, en què X = hores i Y = notes.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{165}{15} = 11;\ \overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j = \frac{89}{15} = 5{,}93$$

$$\sigma_{XY} = \frac{1}{N}\sum_{i}^{n} X_i \cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{1053}{15} - 11\cdot 5{,}93 = 4{,}97$$

$$\sigma_X^2 = \frac{1}{M}\sum_{i=1}^{m} x_i^2 - \overline{X}^2 = \frac{2075}{15} - 11^2 = 17{,}33$$

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow y - 5{,}93 = \frac{4{,}97}{17{,}33}\cdot(x - 11) \Rightarrow$$

$$\Rightarrow y - 5{,}93 = 0{,}28\cdot(x - 11) \Rightarrow y = 2{,}8 + 0{,}28x$$

d) Un alumne estudia un parell d'hores a la setmana; és a dir, $x = 2$. Substituïm aquest valor en l'equació anterior i tindrem la nota mitjana d'aquest alumne.

$$y = 2{,}8 + 0{,}28\cdot 2 \rightarrow y = 3{,}36$$

34. a) Coeficient de correlació entre les variables:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{205}{6} = 34{,}1\widehat{6};\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{4\,290}{6} = 715$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{7163}{6} - 34{,}1\widehat{6}^2 = 26{,}47$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{3\,077\,900}{6} - 715^2 = 1\,758{,}33$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{147\,790}{6} - 34{,}1\widehat{6}\cdot 715 = 202{,}5$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{202{,}5}{\sqrt{26{,}47\cdot 1\,758{,}33}} = 0{,}939$$

b) Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 715 = \frac{202{,}5}{26{,}47}\cdot\left(x - 34{,}1\widehat{6}\right) \Rightarrow$$

$$\Rightarrow y - 715 = 7{,}65\cdot\left(x - 34{,}1\widehat{6}\right) \Rightarrow y = 453{,}6 + 7{,}65x$$

35. Sigui X = inversió i Y = rendibilitat.

a) $\overline{X} = \frac{1}{N}\sum_i X_i = \frac{80}{6} = 13{,}\widehat{3};\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{27}{6} = 4{,}5$

$$\sigma_X = \sqrt{\frac{1}{N}\sum_i X_i^2 - \overline{X}^2} = \sqrt{\frac{1\,086}{6} - 13{,}\widehat{3}^2} = 1{,}8$$

$$\sigma_Y = \sqrt{\frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2} = \sqrt{\frac{123}{6} - 4{,}5^2} = 0{,}5$$

b) Coeficient de correlació:

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{364}{6} - 13{,}\widehat{3}\cdot 4{,}5 = 0{,}\widehat{6}$$

$$r = \frac{\sigma_{XY}}{\sigma_X\cdot\sigma_Y} = \frac{0{,}\widehat{6}}{1{,}8\cdot 0{,}5} = 0{,}74$$

c) Recta de regressió de X sobre Y:

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2}\cdot\left(y - \overline{Y}\right) \Rightarrow x - 13{,}\widehat{3} = \frac{0{,}\widehat{6}}{0{,}25}\cdot(y - 4{,}5) \Rightarrow$$

$$\Rightarrow x - 13{,}\widehat{3} = 0{,}76\cdot(y - 4{,}5) \Rightarrow x = 1{,}33 + 2{,}67y$$

Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow y - 4{,}5 = \frac{0{,}\widehat{6}}{3{,}24}\cdot\left(x - 13{,}\widehat{3}\right) \Rightarrow$$

$$\Rightarrow y - 4{,}5 = 0{,}2\cdot\left(x - 13{,}\widehat{3}\right) \Rightarrow y = 1{,}74 + 0{,}2x$$

d) Tenim que $x = 13{,}5$; així que substituïm aquest valor en la recta de regressió de Y sobre X:

$$y = 1{,}74 + 0{,}2x \rightarrow y = 1{,}74 + 0{,}2\cdot 13{,}5 \rightarrow y = 4{,}44$$

e) En aquest cas, tenim que $y = 5{,}5$, així que substituïm aquest valor en la recta de regressió de X sobre Y.

$$x = 1{,}33 + 2{,}67y \rightarrow x = 1{,}33 + 2{,}67\cdot 5{,}5 \rightarrow x = 16$$

36. $y = 0{,}95x + 1 \Rightarrow \frac{\sigma_{XY}}{\sigma_X^2} = 0{,}95 \Rightarrow \sigma_X^2 = \frac{\sigma_{XY}}{0{,}95}$

$$x = 1{,}01y + a \Rightarrow \frac{\sigma_{XY}}{\sigma_Y^2} = 1{,}01 \Rightarrow \sigma_Y^2 = \frac{\sigma_{XY}}{1{,}01}$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{\sigma_{XY}}{\sqrt{\frac{\sigma_{XY}}{0{,}95}\cdot\frac{\sigma_{XY}}{1{,}01}}} =$$

$$= \frac{\sigma_{XY}}{\frac{\sigma_{XY}}{\sqrt{0{,}959\,5}}} = 0{,}979\,5$$

37. Sabem que $r = 0{,}99$; $\sigma_X^2 = 0{,}8$; $\sigma_Y^2 = 1{,}1$; per tant, si substituïm en la fórmula del coeficient de correlació, tindrem:

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} \Rightarrow 0{,}99 = \frac{\sigma_{XY}}{\sqrt{0{,}8\cdot 1{,}1}} \Rightarrow \sigma_{XY} = 0{,}928\,7$$

38. a) Per a saber si les variables estan relacionades, hem de buscar el coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{39}{8} = 4{,}875;\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{56}{8} = 7$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{225}{8} - 4{,}875^2 = 4{,}359\,38$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{404}{8} - 7^2 = 1{,}5$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i\cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{283}{8} - 4{,}875\cdot 7 = 1{,}25$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}} = \frac{1{,}25}{\sqrt{4{,}359\,38\cdot 1{,}5}} = 0{,}49$$

A partir d'aquest resultat, podem afirmar que les variables estan relacionades, però no fortament, ja que el valor del coeficient de correlació no és gaire proper a 1.

b) Per resoldre aquest exercici, calculem la recta de regressió de Y sobre X:

$$y-\overline{Y}=\frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x-\overline{X}\right)\Rightarrow y-7=\frac{1{,}25}{4{,}359\,38}\cdot\left(x-4{,}875\right)\Rightarrow$$

$$\Rightarrow y-7=0{,}29\cdot\left(x-4{,}875\right)\Rightarrow y=5{,}602+0{,}286x$$

Ara, com que $x = 7{,}5$, substituïm aquest valor en l'equació anterior: $y = 5{,}602 + 0{,}286 \cdot 7{,}5 = 7{,}75$.

Per tant, la nota que s'espera que obtingui aquest alumne és de 7,75.

39. a) Aquestes rectes poden formar perfectament dues rectes de regressió d'alguna variable bidimensional, ja que si resolem el seu sistema resultant obtenim una solució.

$$\left.\begin{aligned}3x-y&=-0{,}8\\x-1{,}1y&=1\end{aligned}\right\}\Rightarrow x=-0{,}82;\ y=-1{,}65$$

b) Siguin r: $3x - y = -0{,}8$ i s: $x - 1{,}1y = 1$.

Suposem que la recta r és la regressió de Y sobre X i que s és la regressió de X sobre Y. Si això és cert, tindrem que:

$$\frac{\sigma_{XY}}{{\sigma_X}^2}=3;\quad\frac{\sigma_{XY}}{{\sigma_Y}^2}=1{,}1$$

El coeficient de Pearson és:

$$r=\frac{\sigma_{XY}}{\sigma_X\cdot\sigma_Y}\Rightarrow r^2=\frac{{\sigma_{XY}}^2}{{\sigma_X}^2\cdot{\sigma_Y}^2}\Rightarrow r=\sqrt{\frac{\sigma_{XY}}{{\sigma_X}^2}\cdot\frac{\sigma_{XY}}{{\sigma_Y}^2}}$$

Substituïm les dades que hem obtingut:

$$r=\sqrt{\frac{\sigma_{XY}}{{\sigma_X}^2}\cdot\frac{\sigma_{XY}}{{\sigma_Y}^2}}=\sqrt{1{,}1\cdot3}=1{,}81>1$$

Com que aquest valor és més gran que 1, arribem a una contradicció, ja que aquest valor no pot ser més gran que 1. És a dir, la recta r és la regressió de X sobre Y i la recta s és la regressió de Y sobre X.

40. Sigui X = alçada i Y = pes.

a) $$\overline{X}=\frac{1}{N}\sum_i X_i=\frac{13{,}53}{8}=1{,}69;$$

$$\overline{Y}=\frac{1}{M}\sum_i Y_i=\frac{557}{8}=69{,}63$$

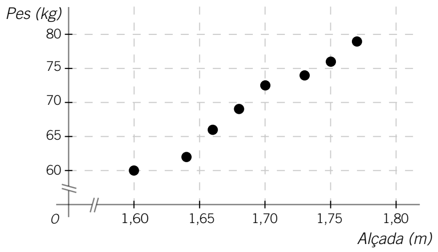
$$\sigma_X^2=\frac{1}{N}\sum_i X_i^2-\overline{X}^2=\frac{22{,}905\,9}{8}-1{,}69^2=0{,}003$$

$$\sigma_Y^2=\frac{1}{N}\sum_i Y_i^2-\overline{Y}^2=\frac{39\,081}{8}-69{,}63^2=37{,}5$$

b) $$\sigma_{XY}=\frac{1}{N}\sum_i^n X_i\cdot Y_i-\overline{X}\cdot\overline{Y}=\frac{944{,}64}{8}-1{,}69\cdot69{,}63=0{,}327$$

$$r=\frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}}=\frac{0{,}327}{\sqrt{0{,}003\cdot37{,}5}}=0{,}975$$

c)



d) La recta de regressió del pes sobre l'alçada és la recta de regressió de Y sobre X:

$$y-\overline{Y}=\frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x-\overline{X}\right)\Rightarrow$$

$$\Rightarrow y-69{,}63=\frac{0{,}327}{0{,}003}\cdot\left(x-1{,}69\right)\Rightarrow$$

$$\Rightarrow y-69{,}63=\cdot\left(x-1{,}69\right)\Rightarrow y=-114{,}58+109x$$

e) Tenim que $x = 1{,}69$, per tant:

$$y=-114{,}58+109\cdot1{,}69\rightarrow y=69{,}63$$

f) La predicció és bona, ja que per a 1,69, que està entre els valors 1,68 i 1,7, tenim que els valors del pes estaran entre 69 i 72. Com que la nostra predicció per al pes és de 69,63, que està entre els valors 69 i 72, la predicció és bona. A més a més, el valor de l'índex de correlació oposat, 0,975, és molt alt.

41. Utilitzant les eines de la pàgina web indicada, surten els mateixos resultats que en l'apartat anterior, encara que agafant més decimals dels resultats.

42. a) $$\overline{X}=\frac{1}{N}\sum_i X_i=\frac{5{,}5}{10}=0{,}55;$$

$$\overline{Y}=\frac{1}{M}\sum_i Y_i=\frac{17{,}62}{10}=1{,}762$$

$$\sigma_{XY}=\frac{1}{N}\sum_i^n X_i\cdot Y_i-\overline{X}\cdot\overline{Y}=\frac{12{,}417}{10}-0{,}55\cdot1{,}762=0{,}27$$

b) $$\sigma_X^2=\frac{1}{N}\sum_i X_i^2-\overline{X}^2=\frac{3{,}85}{10}-0{,}55^2=0{,}082\,5$$

$$\sigma_Y^2=\frac{1}{M}\sum_{j=1}^m {y_j}^2-\overline{Y}^2=\frac{40{,}607\,8}{10}-1{,}762^2=0{,}956\,136$$

$$r=\frac{\sigma_{XY}}{\sqrt{\sigma_X^2\cdot\sigma_Y^2}}=\frac{0{,}27}{\sqrt{0{,}082\,5\cdot0{,}956\,136}}=0{,}97$$

Les variables tenen una dependència lineal positiva forta, ja que el valor del coeficient de correlació és molt proper a 1.

43. a) Sigui $m = 1$ kg, aleshores:

$$\left.\begin{aligned}Y&=mg\frac{t^2}{2}\\X&=\frac{t^2}{2}\end{aligned}\right\}\Rightarrow Y=mgX\Rightarrow Y=gX$$

Aquesta equació relaciona les dues variables X i Y amb la gravetat.

b) En aquest cas, no es tracta d'una recta de regressió, ja que en l'expressió no hi ha terme independent. El valor de g és el quocient entre les mitjanes de Y i X.

$$\overline{X}=\frac{1}{N}\sum_i X_i=\frac{1}{4}\left(10+12+15+8\right)=11{,}25$$

$$\overline{Y}=\frac{1}{M}\sum_i Y_i=\frac{1}{4}\left(100+118+150+80\right)=112$$

$$Y=gX\Rightarrow112=g\cdot11{,}25\Rightarrow g=\frac{112}{11{,}25}\Rightarrow g=9{,}96$$

Per tant, el valor de la constant g per a aquests valors és: $g = 9{,}96$.

SÍNTESI

Pàgs. 331 i 332

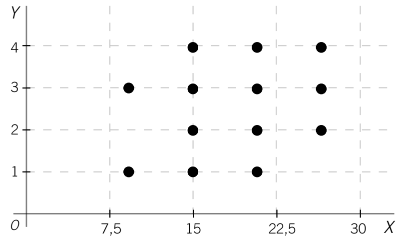
44. a) Marginal de X:

X	10	15	20	25
f_i	3	12	13	9

Marginal de Y:

Y	1	2	3	4
f_i	8	9	11	9

b)



c) $\overline{X} = \frac{1}{N}\sum_{i=1}^{n} x_i \cdot f_i = \frac{1}{37}(10 \cdot 3 + 15 \cdot 12 + 20 \cdot 13 + 25 \cdot 9) =$
$= 18{,}78$

$\overline{Y} = \frac{1}{M}\sum_{j=1}^{m} y_j \cdot f_j = \frac{1}{37}(1 \cdot 8 + 2 \cdot 9 + 3 \cdot 11 + 4 \cdot 9) =$
$= 2{,}57$

d) $\sigma_X = \sqrt{\frac{1}{N}\sum_i f_i \cdot x_i^2 - \overline{X}^2} = \sqrt{\frac{13\,825}{37} - 18{,}78^2} = 4{,}56$

$\sigma_Y = \sqrt{\frac{1}{M}\sum_j f_j \cdot y_j^2 - \overline{Y}^2} = \sqrt{\frac{287}{37} - 2{,}57^2} = 1{,}08$

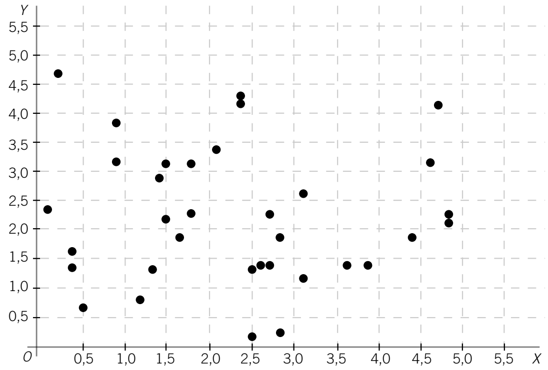
e) $\sigma_{XY} = \frac{1}{N}\sum_i^n \sum_j^m x_i \cdot y_j \cdot f_{ij} - \overline{X} \cdot \overline{Y} =$

$= \frac{1}{37}(10 \cdot 1 \cdot 2 + 10 \cdot 2 \cdot 0 + ... + 25 \cdot 4 \cdot 5) - 18{,}78 \cdot 2{,}57 =$
$= 2{,}68$

45. a) Construïm la taula de doble entrada, i hi indiquem les marques de classe:

	Y_j	0,5	1,5	2,5	3,5	4,5
X_i	X \ Y	0-1	1-2	2-3	3-4	4-5
0,5	0-1	1	2	1	2	1
1,5	1-2	1	2	3	2	0
2,5	2-3	2	4	1	1	2
3,5	3-4	0	3	1	0	0
4,5	4-5	0	1	3	1	1

b)



46. Calculem el valor del coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{233}{10} = 23{,}3;\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{19}{10} = 1{,}9$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{6\,067}{10} - 23{,}3^2 = 63{,}81$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{49}{10} - 1{,}9^2 = 1{,}29$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{490}{10} - 23{,}3 \cdot 1{,}9 = 4{,}73$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{4{,}73}{\sqrt{63{,}81 \cdot 1{,}29}} = 0{,}52$$

La recta de regressió de X sobre Y és la següent:

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2} \cdot (y - \overline{Y}) \Rightarrow x - 23{,}3 = \frac{4{,}73}{1{,}29} \cdot (y - 1{,}9) \Rightarrow$$

$$\Rightarrow x - 23{,}3 = 3{,}67 \cdot (y - 1{,}9) \Rightarrow x = 16{,}33 + 3{,}67y$$

El valor del coeficient de correlació és més proper a 1 que a 0; per tant, podem dir que la relació entre les variables és forta.

47. a) Coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{27{,}7}{5} = 5{,}54;\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{126}{5} = 25{,}2$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{176{,}09}{5} - 5{,}54^2 = 4{,}526\,4$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{3\,456}{5} - 25{,}2^2 = 56{,}16$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{775{,}1}{5} - 5{,}54 \cdot 25{,}2 = 15{,}412$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{15{,}412}{\sqrt{4{,}5264 \cdot 56{,}16}} = 0{,}967$$

Recta de regressió de X sobre Y:

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2} \cdot (y - \overline{Y}) \Rightarrow$$

$$\Rightarrow x - 5{,}54 = \frac{15{,}412}{56{,}12} \cdot (y - 25{,}2) \Rightarrow$$

$$\Rightarrow x - 5{,}54 = 0{,}27 \cdot (y - 25{,}2) \Rightarrow x = -1{,}38 + 0{,}27y$$

Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 25{,}2 = \frac{15{,}412}{4{,}5264} \cdot \left(x - 5{,}54\right) \Rightarrow$$

$$\Rightarrow y - 25{,}2 = 3{,}4 \cdot \left(x - 5{,}54\right) \Rightarrow y = 6{,}34 + 3{,}4x$$

b) Sigui $x = 5$; substituïm aquest valor en la recta de regressió de Y sobre X:

$$y = 6{,}34 + 3{,}4 \cdot 5 \rightarrow y = 23{,}34 \text{ kg}$$

Sigui $y = 36$; substituïm aquest valor en la recta de regressió de X sobre Y:

$$x = -1{,}38 + 0{,}27 \cdot 36 \rightarrow x = 8{,}34 \text{ anys}$$

48. a) Trobem les rectes de regressió lineal:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{179}{6} = 29{,}8\hat{3};\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{239}{6} = 39{,}8\hat{3}$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{5\,783}{6} - 29{,}8\hat{3}^2 = 73{,}80\hat{5}$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{11107}{6} - 39{,}8\hat{3}^2 = 264{,}47\hat{2}$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{7\,287}{6} - 29{,}8\hat{3} \cdot 39{,}8\hat{3} =$$

$$= 26{,}13\hat{8}$$

Recta de regressió de X sobre Y:

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \Rightarrow$$

$$\Rightarrow x - 29{,}8\hat{3} = \frac{26{,}13\hat{8}}{264{,}47\hat{2}} \cdot \left(y - 39{,}8\hat{3}\right) \Rightarrow$$

$$\Rightarrow x - 29{,}8\hat{3} = 0{,}1 \cdot \left(y - 39{,}8\hat{3}\right) \Rightarrow x = 25{,}9 + 0{,}1y$$

Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 39{,}8\hat{3} = \frac{26{,}13\hat{8}}{73{,}80\hat{5}} \cdot \left(x - 29{,}8\hat{3}\right) \Rightarrow$$

$$\Rightarrow y - 39{,}8\hat{3} = 0{,}35 \cdot \left(x - 29{,}8\hat{3}\right) \Rightarrow y = 29{,}268 + 0{,}35x$$

b) $r = \dfrac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \dfrac{26{,}13\hat{8}}{\sqrt{73{,}80\hat{5} \cdot 264{,}47\hat{2}}} = 0{,}187$

Aquest resultat és lluny d'1, de manera que la dependència entre les dues variables és feble.

49. a) Trobem la recta de regressió de X sobre Y, on X és la variable que determina les cendres sulfuroses contingudes en l'aire i Y és el nombre de persones hospitalitzades més de 7 dies per problemes respiratoris.

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{84}{7} = 12;\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{96}{7} = 13{,}714\,3$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{1142}{7} - 12^2 = 19{,}143$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{1578}{7} - 13{,}714\,3^2 = 37{,}347$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{1\,326}{7} - 12 \cdot 13{,}714\,3 =$$

$$= 24{,}857$$

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2} \cdot \left(y - \overline{Y}\right) \Rightarrow$$

$$\Rightarrow x - 12 = \frac{24{,}857}{37{,}347} \cdot \left(y - 13{,}714\,3\right) \Rightarrow$$

$$\Rightarrow x - 12 = 0{,}67 \cdot \left(y - 13{,}714\,3\right) \Rightarrow x = 2{,}87 + 0{,}67y$$

b) Calculem la recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 13{,}714\,3 = \frac{24{,}857}{19{,}143} \cdot \left(x - 12\right) \Rightarrow$$

$$\Rightarrow y - 13{,}714\,3 = 1{,}298 \cdot \left(x - 12\right) \Rightarrow$$

$$\Rightarrow y = -1{,}868 + 1{,}298x$$

Sigui $x = 20$; substituïm aquest valor en la recta de regressió que acabem de calcular:

$$y = -1{,}868 + 1{,}298 \cdot 20 \rightarrow y = 24{,}1 \text{ persones}$$

50. a) Hem de buscar la recta de regressió de Y sobre X:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{10}{4} = 2{,}5;$$

$$\overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{3\,462}{4} = 865{,}5$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{30}{4} - 2{,}5^2 = 1{,}25$$

$$\sigma_Y^2 = \frac{1}{M}\sum_i Y_i^2 - \overline{Y}^2 = \frac{9\,163\,074}{4} - 865{,}5^2 =$$

$$= 1\,541\,678{,}25$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X} \cdot \overline{Y} = \frac{13\,317}{4} - 2{,}5 \cdot 865{,}5 =$$

$$= 1\,165{,}5$$

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2} \cdot \left(x - \overline{X}\right) \Rightarrow$$

$$\Rightarrow y - 865{,}5 = \frac{1165{,}5}{1{,}25} \cdot \left(x - 2{,}5\right) \Rightarrow$$

$$\Rightarrow y - 865{,}5 = 932{,}4 \cdot \left(x - 2{,}5\right) \Rightarrow$$

$$\Rightarrow y = -1\,465{,}5 + 932{,}4x$$

Com que volem saber el nombre de bacteris en l'hora següent, tenim que $x = 5$; Substituïm aquest valor en l'equació anterior:

$$y = -1\,465{,}5 + 932{,}4 \cdot 5 = 3\,196{,}5$$

b) Per a saber si aquest ajust és un bon ajust, hem de calcular el coeficient de correlació:

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{1165{,}5}{\sqrt{1{,}25 \cdot 1\,541\,678{,}25}} = 0{,}84$$

Malgrat que l'ajust lineal és bo, observem que el creixement de la funció augmenta amb el pas del temps; així, entre la primera i la segona hores, l'augment és de 48 bacteris, entre la segona i la tercera, de 345, i entre la tercera i la quarta, de 2600. Probablement, una funció exponencial s'ajustaria millor al comportament del cultiu.

51. a) Sabem que $b = \dfrac{\sigma_{XY}}{\sigma_X^2}$ o $b = \dfrac{\sigma_{XY}}{\sigma_Y^2}$ on el numerador sempre és positiu perquè és un quadrat. Com que σ_{XY} és positiu, aleshores el valor de *b* no pot ser negatiu.

b) Sabem que $b = \dfrac{\sigma_{XY}}{\sigma_X^2}$ o $b = \dfrac{\sigma_{XY}}{\sigma_Y^2}$ i $r = \dfrac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$ i que els signes de les variàncies i les desviacions típiques són positius. Aleshores, $\sigma_X > 0$, $\sigma_Y > 0$, $\sigma_X{}^2 > 0$, $\sigma_Y{}^2 > 0$.

Com que el valor del coeficient de correlació és positiu, aleshores el valor de la covariància també és positiu. Això contradiu el fet que *b* sigui negatiu.

c) Sabem que $r = \dfrac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \dfrac{120}{\sqrt{9 \cdot 15}} = 10{,}33 > 1.$

Això no pot ser, ja que el valor del coeficient de correlació ha d'estar entre els valors de –1 i 1.

d) Sabem que $r = \dfrac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \dfrac{15}{\sqrt{30 \cdot 15}} = 0{,}71 < 1.$

Aquest resultat sí que és possible, ja que el valor del coeficient de correlació està entre els valors de –1 i 1.

52. a) Aquestes rectes poden formar perfectament dues rectes de regressió d'alguna variable bidimensional, ja que, si resolem el seu sistema resultant, obtenim una solució.

$$\left.\begin{array}{r} 3x + 2y = 7 \\ x + 4y = 5 \end{array}\right\} \Rightarrow x = \frac{9}{5};\ y = \frac{4}{5}$$

b) Suposem que la recta *r* és la regressió de *Y* sobre *X* i que *s* és la regressió de *X* sobre *Y*. Si això és cert, tindrem que:

r: $3x + 2y = 7 \Rightarrow y = -3x/2 + 7/2$

s: $x + 4y = 5 \Rightarrow x = 5 - 4y$

$$\frac{\sigma_{XY}}{\sigma_X{}^2} = -\frac{3}{2};\quad \frac{\sigma_{XY}}{\sigma_Y{}^2} = -4$$

El coeficient de Pearson és:

$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \Rightarrow r^2 = \frac{\sigma_{XY}{}^2}{\sigma_X{}^2 \cdot \sigma_Y{}^2} \Rightarrow r = \sqrt{\frac{\sigma_{XY}}{\sigma_X{}^2} \cdot \frac{\sigma_{XY}}{\sigma_Y{}^2}}$$

Substituïm les dades que hem obtingut:

$$r = \sqrt{\frac{\sigma_{XY}}{\sigma_X{}^2} \cdot \frac{\sigma_{XY}}{\sigma_Y{}^2}} = \sqrt{\left(-\frac{3}{2}\right) \cdot (-4)} = 2{,}45 > 1$$

Com que aquest valor és més gran que 1, arribem a una contradicció, de manera que queda falsejada la hipòtesi inicial. És a dir, la recta *r* és la regressió de *X* sobre *Y* i la recta *s* és la regressió de *Y* sobre *X*.

Avaluació (pàg. 334)

(pàg. 334)

1. a) El color de cabells i el color d'ulls són variables qualitatives.

b) El nombre de germans i el nombre d'ordinadors a casa són variables quantitatives i discretes.

2. Construïm una taula de doble entrada per facilitar els càlculs dels diferents apartats:

Y_j	1	3	5	7	9	11	
X \ Y	0-2	2-4	4-6	6-8	8-10	10-12	f_i
0	3	5	8	7	5	3	31
1	1	2	4	7	3	1	18
2	1	1	1	1	1	0	5
3	0	1	1	0	0	1	3
4	0	0	1	0	1	0	2
5	0	0	0	0	0	1	1
f_i	5	9	15	15	10	6	60

a) $\overline{X} = \dfrac{1}{N}\displaystyle\sum_{i=1}^{n} x_i \cdot f_i = \dfrac{50}{60} = 0{,}83;$

$\overline{Y} = \dfrac{1}{M}\displaystyle\sum_{j=1}^{m} y_j \cdot f_j = \dfrac{368}{60} = 6{,}13$

$\sigma_X{}^2 = \dfrac{1}{N}\displaystyle\sum_i f_i \cdot x_i^2 - \overline{X}^2 = \dfrac{122}{60} - 0{,}83^2 = 1{,}34$

$\sigma_Y{}^2 = \dfrac{1}{M}\displaystyle\sum_j f_j \cdot y_j^2 - \overline{Y}^2 = \dfrac{2\,732}{60} - 6{,}13^2 = 7{,}92$

b) $\sigma_{XY} = \dfrac{1}{N}\displaystyle\sum_i^n\sum_j^m x_i \cdot y_j \cdot f_{ij} - \overline{X} \cdot \overline{Y} =$

$= \dfrac{1}{60}\left(0 \cdot 1 \cdot 3 + 0 \cdot 3 \cdot 5 + \ldots + 5 \cdot 11 \cdot 1\right) - 0{,}83 \cdot 6{,}13 =$
$= 0{,}42$

c) Per a saber la relació que hi ha entre les variables, hem de calcular el valor del coeficient de correlació:

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{0{,}42}{\sqrt{1{,}34 \cdot 7{,}92}} = 0{,}129$$

Com que aquest valor és molt proper a 0, la dependència lineal entre les variables és molt feble.

3. Les rectes de regressió es tallen en el centre de gravetat d'una distribució, és a dir, en el punt $\left(\overline{X}, \overline{Y}\right)$. Plantegem el sistema següent i el resolem:

$$\left.\begin{array}{l} y = 2{,}6x - 2{,}55 \\ x = 0{,}36y + 1 \end{array}\right\} \Rightarrow x = 1{,}28;\ y = 0{,}78$$

Per tant, la solució del sistema anterior coincideix amb les mitjanes de *X* i de *Y*. És a dir, $\overline{X} = 1{,}28$; $\overline{Y} = 0{,}78$.

Calculem el coeficient de correlació:

$$y = 2{,}6x - 2{,}55 \Rightarrow \frac{\sigma_{XY}}{\sigma_X^2} = 2{,}6 \Rightarrow \sigma_X^2 = \frac{\sigma_{XY}}{2{,}6}$$

$$x = 0{,}36y + 1 \Rightarrow \frac{\sigma_{XY}}{\sigma_Y^2} = 0{,}36 \Rightarrow \sigma_Y^2 = \frac{\sigma_{XY}}{0{,}36}$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{\sigma_{XY}}{\sqrt{\dfrac{\sigma_{XY}}{2{,}6} \cdot \dfrac{\sigma_{XY}}{0{,}36}}} = \frac{\sigma_{XY}}{\dfrac{\sigma_{XY}}{0{,}967}} = 0{,}967$$

4. Busquem el coeficient de correlació:

$$y = 3x - 1 \Rightarrow \frac{\sigma_{XY}}{\sigma_x^2} = 3 \Rightarrow \sigma_x^2 = \frac{\sigma_{XY}}{3}$$

$$\sigma_{XY} = 0{,}6 \Rightarrow \sigma_x^2 = \frac{0{,}6}{3} \Rightarrow \sigma_x^2 = 0{,}2$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{0{,}6}{\sqrt{0{,}2 \cdot 2^2}} = \frac{0{,}6}{\sqrt{0{,}2 \cdot 4}} = 0{,}67$$

5. a) Primerament, calculem el coeficient de correlació:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{712}{5} = 142{,}4;\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{79}{5} = 15{,}8$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{102\,576}{5} - 142{,}4^2 = 37{,}44$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{1285}{5} - 15{,}8^2 = 7{,}36$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{11\,320}{5} - 142{,}4 \cdot 15{,}8 = 14{,}08$$

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{14{,}08}{\sqrt{37{,}44 \cdot 7{,}36}} = 0{,}848$$

La recta de regressió de Y sobre X és la següent:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow$$

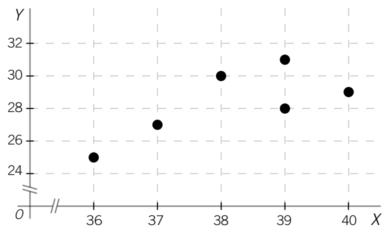
$$\Rightarrow y - 15{,}8 = \frac{14{,}08}{37{,}44}\cdot\left(x - 142{,}4\right) \Rightarrow$$

$$\Rightarrow y - 15{,}8 = 0{,}376\cdot\left(x - 142{,}4\right) \Rightarrow y = -37{,}75 + 0{,}376x$$

b) Sigui $x = 135$; substituïm aquest valor en la recta de regressió de Y sobre X:

$y = -37{,}75 + 0{,}376 \cdot 135 = 13{,}01$ milions de tones.

6. a)



b) Calculem les dades necessàries per a trobar les rectes de regressió:

$$\overline{X} = \frac{1}{N}\sum_i X_i = \frac{229}{6} = 38{,}1\hat{6};\ \overline{Y} = \frac{1}{M}\sum_i Y_i = \frac{170}{6} = 28{,}\hat{3}$$

$$\sigma_X^2 = \frac{1}{N}\sum_i X_i^2 - \overline{X}^2 = \frac{8\,751}{6} - 38{,}1\hat{6}^2 = 1{,}80\hat{5}$$

$$\sigma_Y^2 = \frac{1}{N}\sum_i Y_i^2 - \overline{Y}^2 = \frac{4\,840}{6} - 28{,}\hat{3}^2 = 3{,}\hat{8}$$

$$\sigma_{XY} = \frac{1}{N}\sum_i^n X_i \cdot Y_i - \overline{X}\cdot\overline{Y} = \frac{6\,500}{6} - 38{,}1\hat{6} \cdot 28{,}\hat{3} = 1{,}9\hat{4}$$

Recta de regressió de X sobre Y:

$$x - \overline{X} = \frac{\sigma_{XY}}{\sigma_Y^2}\cdot\left(y - \overline{Y}\right) \Rightarrow x - 38{,}1\hat{6} = \frac{1{,}9\hat{4}}{3{,}\hat{8}}\cdot\left(y - 28{,}\hat{3}\right) \Rightarrow$$

$$\Rightarrow x - 38{,}1\hat{6} = 0{,}5\cdot\left(y - 28{,}\hat{3}\right) \Rightarrow x = 24 + 0{,}5y$$

Recta de regressió de Y sobre X:

$$y - \overline{Y} = \frac{\sigma_{XY}}{\sigma_X^2}\cdot\left(x - \overline{X}\right) \Rightarrow y - 28{,}\hat{3} = \frac{1{,}9\hat{4}}{1{,}80\hat{5}}\cdot\left(x - 38{,}1\hat{6}\right) \Rightarrow$$

$$\Rightarrow y - 28{,}\hat{3} = 1{,}077\cdot\left(x - 38{,}1\hat{6}\right) \Rightarrow y = -12{,}77 + 1{,}077x$$

c) El coeficient de correlació és el següent:

$$r = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \cdot \sigma_Y^2}} = \frac{1{,}9\hat{4}}{\sqrt{1{,}80\hat{5} \cdot 3{,}\hat{8}}} = 0{,}734$$

Aquest valor és molt proper a 1, la qual cosa significa que la dependència entre les dues variables és forta.

7. a) En aquest gràfic observem que les variables tenen una dependència lineal positiva feble, ja que els punts estan bastant dispersos.

b) En aquest cas, les variables tenen una dependència lineal negativa forta, ja que els punts es podrien aproximar fàcilment per una recta. Aquí, la correlació és negativa i propera a −1.

c) Aquest gràfic presenta una dependència lineal positiva forta entre les variables, pel mateix motiu que en el gràfic anterior. La correlació és positiva i propera a 1.

d) Finalment, les variables representades en aquest gràfic són independents, ja que els punts no formen cap recta. En aquest cas, la correlació és propera a 0.

8. a) El signe de la covariància i el del coeficient de correlació han de ser els mateixos, ja que les variàncies i les desviacions típiques són sempre positives. Per tant, aquest cas no es pot donar, ja que la covariància i el coeficient de correlació tenen signes diferents.

b) Aquest cas tampoc no es pot donar, ja que el valor del coeficient de correlació no pot ser més gran que 1.

c) Aquesta opció sí que és vàlida, perquè els signes de la covariància i del coeficient de correlació són iguals i el valor de r està entre −1 i 1.

d) Aquest resultat també és possible per la mateixa raó anterior. Els signes són iguals i el valor del coeficient de correlació està entre −1 i 1.

9. a) Aquestes rectes poden ser perfectament dues rectes de regressió d'alguna distribució bidimensional, ja que, si resolem el seu sistema resultant, obtenim una solució:

$$\left.\begin{aligned} 7x + 3y &= 2 \\ 2x + 6y &= 9 \end{aligned}\right\} \Rightarrow x = -\frac{5}{12};\ y = \frac{59}{36}$$

b) Siguin r: $7x + 3y = 2$ i s: $2x + 6y = 9$.

Suposem que la recta r és la regressió de Y sobre X i que s és la regressió de X sobre Y. Si això és cert, tindrem que:

$$\left.\begin{aligned} 7x + 3y = 2 &\Rightarrow y = -7x/3 + 2/3 \\ 2x + 6y = 9 &\Rightarrow x = 4{,}5 - 3y \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \frac{\sigma_{XY}}{{\sigma_X}^2} = -\frac{7}{3};\ \frac{\sigma_{XY}}{{\sigma_Y}^2} = -3$$

El coeficient de Pearson és:

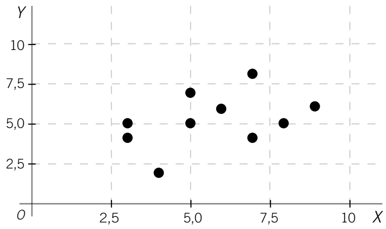
$$r = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} \Rightarrow r^2 = \frac{{\sigma_{XY}}^2}{{\sigma_X}^2 \cdot {\sigma_Y}^2} \Rightarrow r = \sqrt{\frac{\sigma_{XY}}{{\sigma_X}^2} \cdot \frac{\sigma_{XY}}{{\sigma_Y}^2}}$$

Substituïm les dades que hem obtingut:

$$r = \sqrt{\frac{\sigma_{XY}}{{\sigma_X}^2} \cdot \frac{\sigma_{XY}}{{\sigma_Y}^2}} = \sqrt{\left(-\frac{7}{3}\right) \cdot (-3)} = 2{,}65 > 1$$

Però sabem que r no pot ser més gran que 1. És a dir, la recta r és la regressió de X sobre Y i la recta s és la regressió de Y sobre X.

10.



Atès que la distribució de punts és creixent, el valor de r ha de ser positiu; per tant, només poden ser els valors a) i d). D'altra banda, els punts estan bastant dispersos, per la qual cosa el valor del coeficient de correlació ha de ser més a prop de 0 que d'1. Aleshores, l'opció més adequada és l'opció d), 0,4.

Zona + (pàg. 335)

— **El naixement de l'estadística aplicada**

El significat del coeficient de Pearson és determinat de la manera següent:

- Si $r = 0$, no hi ha correlació.
- Si $r = 1$, la correlació és positiva perfecta, i podem dir que hi ha una relació funcional lineal entre totes dues variables.
- Per a $0 < r < 1$, la correlació és positiva.
- Si $r = -1$, la correlació és negativa perfecta, i podem dir que hi ha una relació funcional lineal entre totes dues variables.
- Per a $-1 < r < 0$, la correlació és negativa.

— **Correlacions que confonen**

La paraula *espuri* ve del llatí *spurius*, i accepta dues definicions: 1) 'bastard' (degenerat del seu origen) i 2) 'enganyós, fals'.

La segona accepció té el seu millor exemple en estadística, en què es parla de, *correlació espúria o il·lusòria*, que es produeix quan tenim una situació en la qual les mesures de dues o més variables estan estadísticament relacionades, però en realitat no hi ha cap relació de causalitat entre elles.