

DISTRIBUCIONS BIDIMENSIONALS

Coeficient de correlació

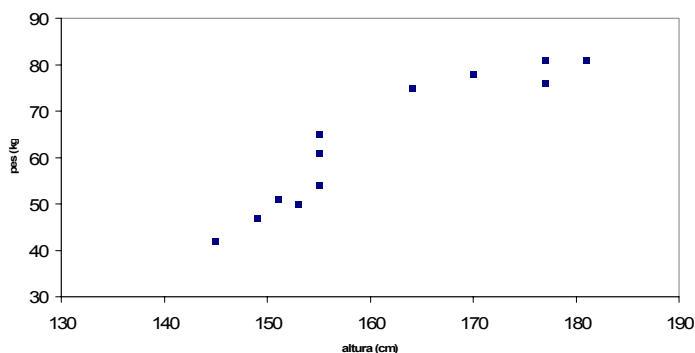
Covariància

Regressions lineals (y sobre x i x sobre y)

Quan es realitza un estudi estadístic de dues variables estadístiques x i y , es pot analitzar el grau de relació o dependència que hi ha entre les dues variables implicades. Fixeu-vos que si representem els valors de x i y en un diagrama x - y podem obtenir diversos tipus de dependències.

EXEMPLE 1

x_i (altura en cm)	y_i (pes en kg)	f
145	42	1
149	47	1
151	51	3
153	50	2
155	65	3
155	61	4
155	54	2
164	75	7
170	78	4
177	81	1
177	76	1
181	81	5
		34



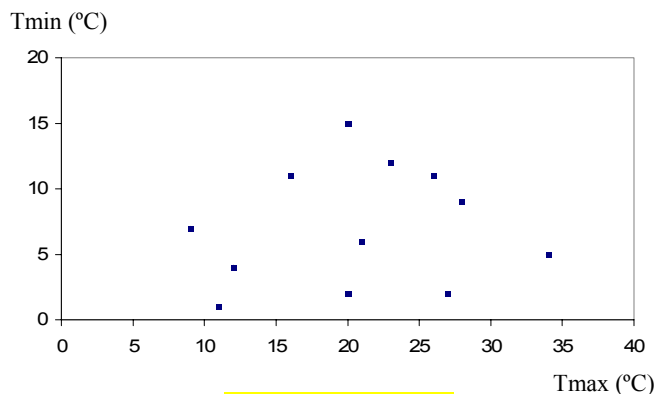
Correlació forta



Quan es realitza un estudi estadístic de dues variables estadístiques x i y , es pot analitzar el grau de relació o dependència que hi ha entre les dues variables implicades. Fixeu-vos que si representem els valors de x i y en un diagrama x - y podem obtenir diversos tipus de dependències.

EXEMPLE 2

T max (°C)	T min (°C)	f
23	12	1
26	11	1
28	9	1
27	2	1
20	2	1
11	1	1
34	5	1
12	4	1
21	6	1
20	15	1
9	7	1
16	11	1
		12



Correlació feble

TIPUS DE DEPENDÈNCIES ENTRE LES VARIABLES X-Y.

Correlació lineal positiva	Correlació lineal negativa	
		Correlació forta
		Correlació feble

COEFICIENT DE CORRELACIÓ DE PEARSON.

Per avaluar el grau de dependència entre les variables x i y usarem *el coeficient de correlació de Pearson*, definit com:

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

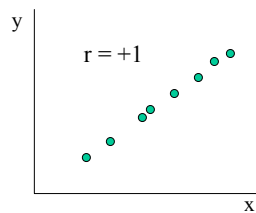
On S_{xy} és la covariància de les dues variables:

$$S_{xy} = \frac{\sum x \cdot y \cdot F}{N} - \bar{x} \cdot \bar{y}$$

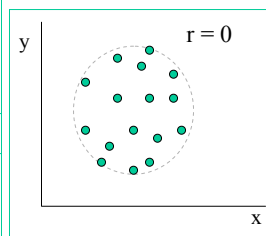
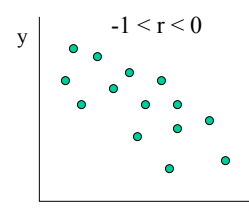
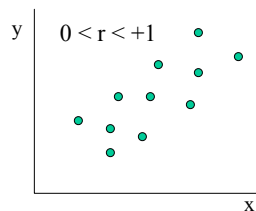
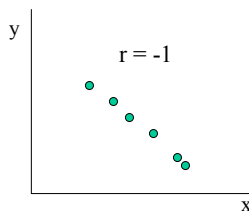
I on S_x i S_y són les desviacions tipus de les variables x i y respectivament:

$$\begin{cases} S_x = \sqrt{S_x^2} = \sqrt{\frac{\sum x^2 F}{N} - \bar{x}^2} \\ S_y = \sqrt{S_y^2} = \sqrt{\frac{\sum y^2 F}{N} - \bar{y}^2} \end{cases}$$

Correlació lineal positiva

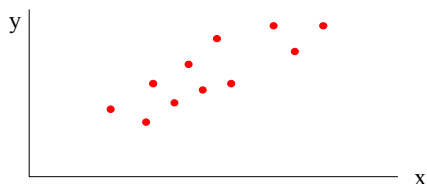


Correlació lineal negativa

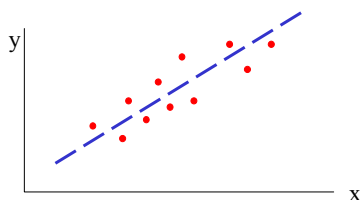


RECTA DE REGRESSIÓ LINEAL Y SOBRE X.

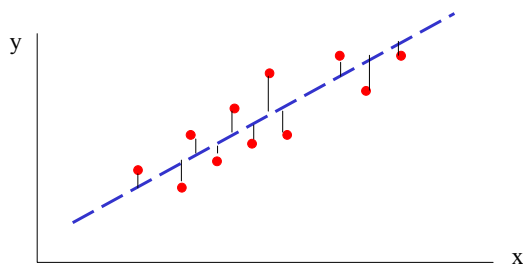
Si la correlació entre les variables x - y és forta, els punts dibuixats en el diagrama semblen estar bastant alineats.



Si la variable y depèn dels valors de x , la recta que millor ajusta la correlació x - y s'anomena recta y sobre x .



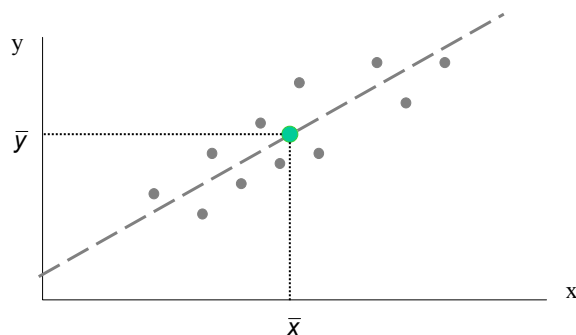
Aquesta recta compleix que el valor de la suma dels quadrats de les distàncies verticals de tots els punts a la recta és el més petit possible.



Tenint en compte aquesta idea, es determina l'equació de la recta anterior (mètode de mínims quadrats), obtenint en aquest cas la relació:

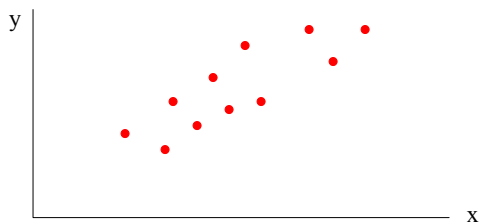
$$y - \bar{y} = \frac{S_{xy}}{S_x^2} (x - \bar{x})$$

S'observa fàcilment que la recta de regressió y sobre x passa sempre pel punt definit per les mitjanes aritmètiques.

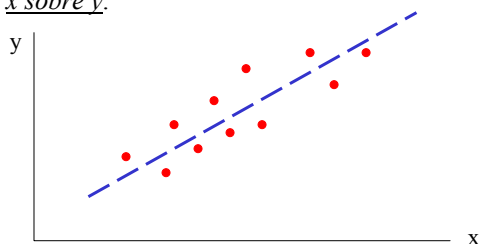


RECTA DE REGRESSIÓ LINEAL X SOBRE Y

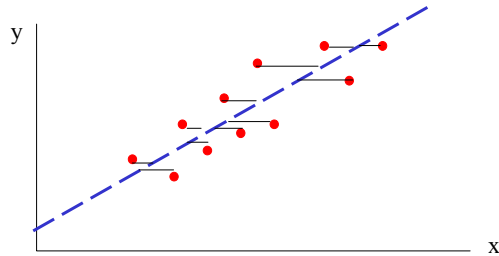
Si la correlació entre les variables x - y és forta, els punts dibuixats en el diagrama es situen al voltant d'una recta.



Si la variable x depèn dels valors de y , la recta que millor ajusta la correlació x - y s'anomena recta x sobre y.



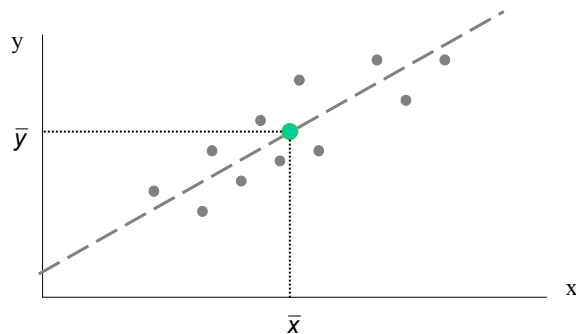
Aquesta recta compleix que el valor de la suma dels quadrats de les distàncies horitzontals de tots els punts a la recta és el més petit possible.



Tenint en compte aquesta idea, es determina l'equació de la recta anterior (mètode de mínims quadrats), obtenint en aquest cas la relació:

$$x - \bar{x} = \frac{S_{xy}}{S_y^2} (y - \bar{y})$$

S'observa fàcilment que la recta de regressió **x sobre y** passa sempre pel punt definit per les mitjanes aritmètiques.

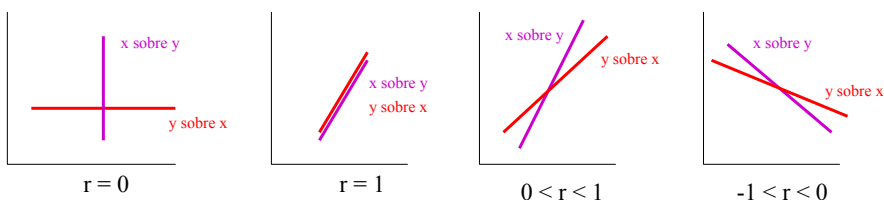


En resum:

- Si $r \sim 1$ (correlació forta), podem usar les rectes de regressió per fer estimacions a partir de valors que no estan a taula de les dades inicials
- Si $r \sim 0$ (correlació feble), no té cap sentit calcular la recta de regressió ja que les estimacions no tindran cap mena de validesa.

Les dues rectes x sobre y i y sobre x ...

- ...són perpendiculars si $r = 0$
- ...són coincidents si $r = 1$
- ...l'angle que formen les dues rectes està relacionat directament amb el valor coeficient de correlació lineal r



EXAMPLE

Anotem l'altura i el pes de 34 alumnes d'una classe de batxillerat. Un cop recollides les dades les ordenem per parelles i construïm la taula de freqüències corresponent.

- Calculeu el coeficient de correlació de x i y .
- Trobeu l'equació de la recta de regressió lineal y sobre x i la x sobre y .
- Segons la recta de regressió lineal quin és el pes que li correspondrà a un noi que medeix 167 cm d'altura? Si el nen pesés 60 kg quina altura li correspondria?

x_i	y_i	f	xf	yf	x^2f	y^2f	xyf
145	42	1	145	42	21025	1764	6090
149	47	1	149	47	22201	2209	7003
151	51	3	453	153	68403	7803	23103
153	50	2	306	100	46818	5000	15300
155	65	3	465	195	72075	12675	30225
155	61	4	620	244	96100	14884	37820
155	54	2	310	108	48050	5832	16740
164	75	7	1148	525	188272	39375	86100
170	78	4	680	312	115600	24336	53040
177	81	1	177	81	31329	6561	14337
177	76	1	177	76	31329	5776	13452
181	81	5	905	405	163805	32805	73305
		34	5535	2288	905007	159020	376515

a) Calculeu el coeficient de correlació de x i y .

xi	yi	f	xf	yf	x ² f	y ² f	xyf
145	42	1	145	42	21025	1764	6090
149	47	1	149	47	22201	2209	7003
151	51	3	453	153	68403	7803	23103
153	50	2	306	100	46818	5000	15300
155	65	3	465	195	72075	12675	30225
155	61	4	620	244	96100	14884	37820
155	54	2	310	108	48050	5832	16740
164	75	7	1148	525	188272	39375	86100
170	78	4	680	312	115600	24336	53040
177	81	1	177	81	31329	6561	14337
177	76	1	177	76	31329	5776	13452
181	81	5	905	405	163805	32805	73305
		34	5535	2288	905007	159020	376515

El coeficient de correlació ve donat per:

$$r = \frac{S_{xy}}{S_x \cdot S_y}$$

Substituint S_{xy} , S_x i S_y a la fórmula anterior, ens quedarà:

$$r = \frac{S_{xy}}{S_x \cdot S_y} = \frac{118,88}{10,77 \cdot 12,19} = 0,906$$

Correlació molt forta

Hem de calcular S_{xy} , S_x , S_y i les mitjanes aritmètiques de x i y :

$$S_{xy} = \frac{\sum x \cdot y \cdot f}{N} - \bar{x} \cdot \bar{y} \Rightarrow S_{xy} = \frac{376.515}{34} - 162,79 \cdot 67,29 = 118,88$$

$$S_x^2 = \frac{\sum x^2 \cdot f}{N} - \bar{x}^2 = \frac{905.007}{34} - 162,79^2 = 115,99 \Rightarrow S_x = \sqrt{115,99} = 10,77$$

$$S_y^2 = \frac{\sum y^2 \cdot f}{N} - \bar{y}^2 = \frac{159.020}{34} - 67,29^2 = 168,74 \Rightarrow S_y = \sqrt{168,74} = 12,19$$

$$\bar{x} = \frac{\sum x \cdot f}{N} = \frac{5.535}{34} = 162,79$$

$$\bar{y} = \frac{\sum y \cdot f}{N} = \frac{2.288}{34} = 67,29$$

b) Trobeu l'equació de la recta de regressió lineal **y sobre x** i la **x sobre y**.

Per calcular la recta de regressió **y sobre x** aplicarem l'equació

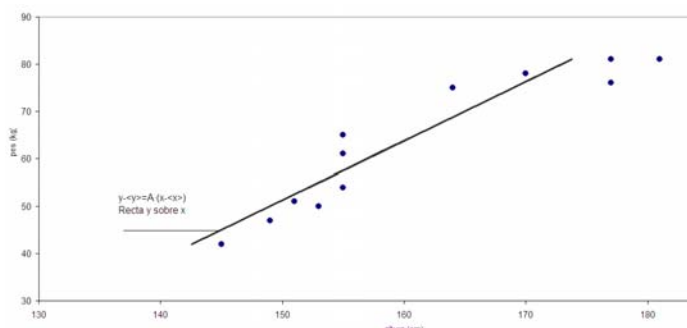
$$y - \bar{y} = \frac{S_{xy}}{S_x^2} (x - \bar{x})$$

Substituirem els valors de S_{xy} , S_x , $\bar{y}$ i $\bar{x}$

$$y - 67,29 = \frac{118,88}{115,99} (x - 162,79)$$

I reordenarem els termes

$$y = 1,02 \cdot x - 99,59$$



Per calcular la recta de regressió x sobre y aplicarem l'equació

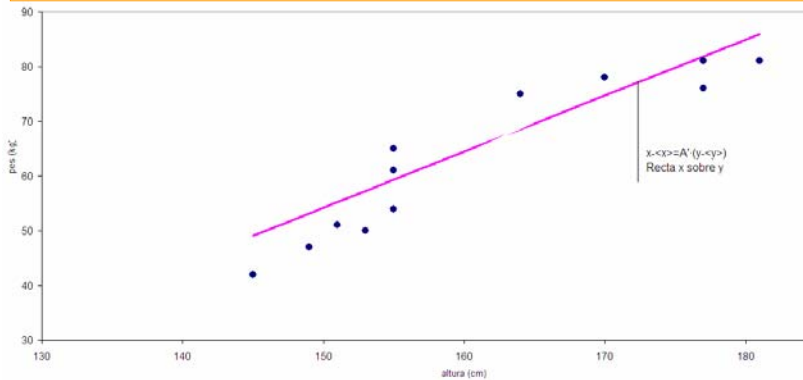
$$x - \bar{x} = \frac{S_{xy}}{S_y^2} (y - \bar{y})$$

Substituïrem els valors de S_{xy} , S_y , $\bar{y}$ i $\bar{x}$

$$x - 162,79 = \frac{118,88}{168,74} (y - 67,29)$$

I reordenarem els termes

$$x = 0,80 \cdot y + 108,94$$



c) Segons la recta de regressió lineal quin és el pes que li correspondrà a un noi que medeix 167 cm d'altura?

Prenem la recta de regressió y sobre x de l'apartat anterior.

$$y = 1,02 \cdot x - 99,59$$

Substituïm x per 167 cm i calculem el pes que li correspon a aquest noi.

$$y = 1,02 \cdot 167 - 99,59 = 70,75 \text{ kg}$$

Si el nen pesés 60 kg quina altura li correspondria?

*Prenem la recta de regressió **x sobre y** de l'apartat anterior.*

$$x = 0,80 \cdot y + 108,94$$

Substituïm y per 60 kg i calculem l'altura que li correspon a aquest noi.

$$y = 0,80 \cdot 60 + 108,64 = 156,64 \text{ cm}$$

Problemes del llibre:

Pg. 189 – 12, 13, 14, 15

Pg. 190 – 16, 21, 22