

ESTADÍSTICA BIDIMENSIONAL

- a) $X =$ notes d'un alumne en Matèies
 $Y =$ " en Anglès

X_i	6	7	5	6	...
Y_i	9	4	7	8	...

VARIABLE ESTADÍSTICA BIDIMENSIONAL

$(6,9)$ $(7,4)$ $(5,7)$ $(6,8)$...



recta de regressió

⑥ Enquesta a 20 famílies, X nombre de persones majors d'edat en la família
 Y nombre de cotxes de que disposen

$X \backslash Y$	1	2	3	4	5	n_y
0	1	1	1	0	0	3
1	1	1	2	1	0	5
2	0	2	3	3	2	10
3	0	0	0	1	1	2
n_x	2	4	6	5	3	20

TAULA de DOBLE ENTRADA

N

tamany mostra

PERSONES	COTXES
1	0
1	1
2	0
2	1
2	2
2	2
3	0
3	1
3	1
3	2
3	2
3	2
4	1
4	2
4	2
4	2
4	3
5	2
5	2
5	3

©) Calcula la taula de doble entrada en una enquesta a 16 alumnes d'un curs sobre el nivell d'estudis de la font d'ingressos en una família (home o dona) i el seu salari mitjà..

Nivell 1 : analfabet o sense estudis

Nivell 2 : estudis primaris

Nivell 3 : educació secundària

Nivell 4 : diplomats universitaris o equivalents

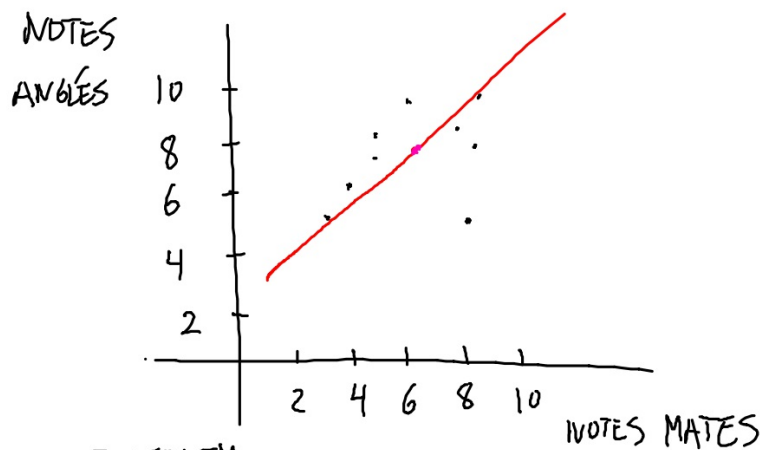
Nivell 5 : llicenciats " " "

Nivell d'estudis	1	2	1	3	4	3	5	5	4	3	5	5	5	2	4	5
Salari Mitjà (milers €)	25	35	45	35	25	55	65	60	45	60	45	50	55	30	50	75

Agrupa els salaris en els intervals següents : $[20, 40)$, $[40, 60)$ i $[60, 80)$

$x \backslash y$	1	2	3	4	5	n_y
$[20, 40)$	1	2	1	1	0	5
$[40, 60)$	1	0	1	2	3	7
$[60, 80)$	0	0	1	0	3	4
n_x	2	2	3	3	6	16

Tornant a l'exemple (a) $\bar{X}$ = notes Mates
 $\bar{Y}$ = " Anglès



OBSERVEM

- * Hi ha relació entre la nota de Mates i d'Anglès
- * La relació es + : quan augmenta la nota de Mates, augmenta la d'anglès
- * Hi ha una relació lineal : és la gràfica (línia recta) que més s'ajusta al núvol de punts

CENTRE de GRAVITAT

És el punt que té com a coordenades les mitjanes de les variables, és a dir,

$$\left(\bar{X}, \bar{Y} \right)$$

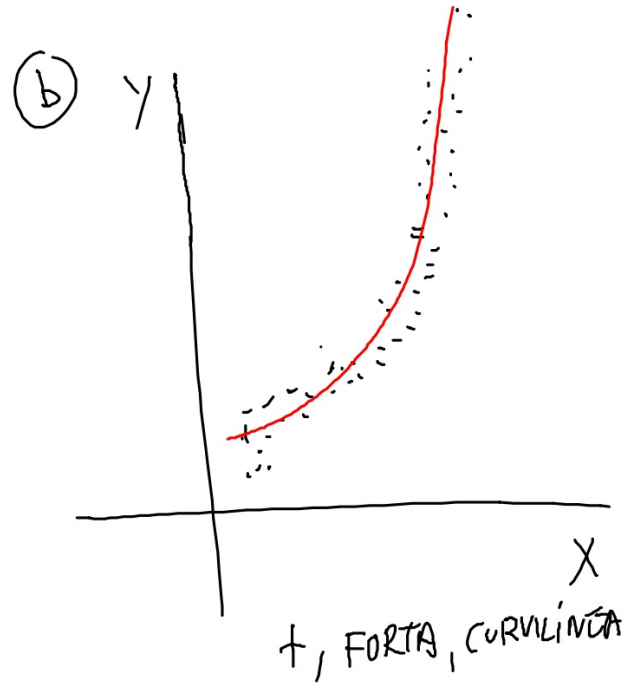
CENTRE de GRAVITAT = $\left(6, 7.3 \right)$

CORRELACIÓ

Correlació pot ser: + FORTA, LÍNEAL
- DÈBIL, CURVILÍNEA



- , FORTA, LINEAL

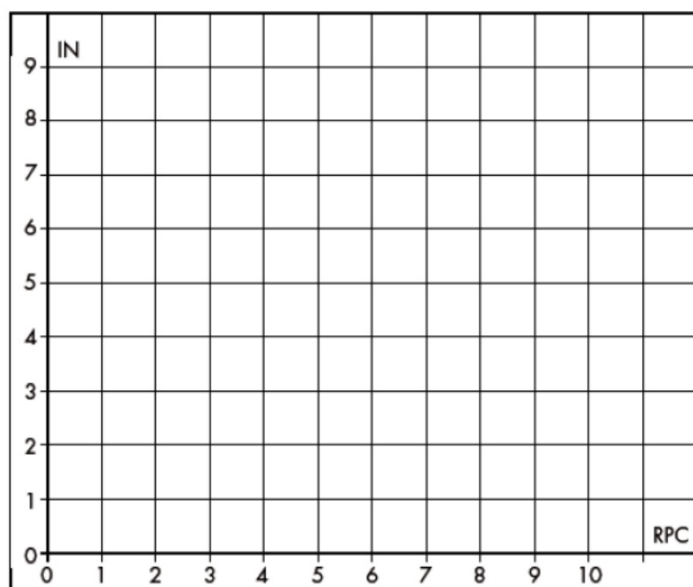


+ , FORTA, CURVILÍNEA

6. La taula següent mostra la renda per capita (RPC) i l'índex de natalitat (IN) de dotze països:
[2 punts: 1 punt pel núvol de punts i 0,5 punts per la recta de regressió, a l'apartat a, i 0,5 punts per l'apartat b]

PAÏSOS	A	B	C	D	E	F	G	H	I	J	K	L
RPC	3	3	4	5	6	5	7	7	8	9	9	10
IN	8	7	7	6	6	5	6	4	5	4	3	3

- a) Representeu els resultats mitjançant un núvol de punts i traceu-ne aproximadament una recta de regressió.

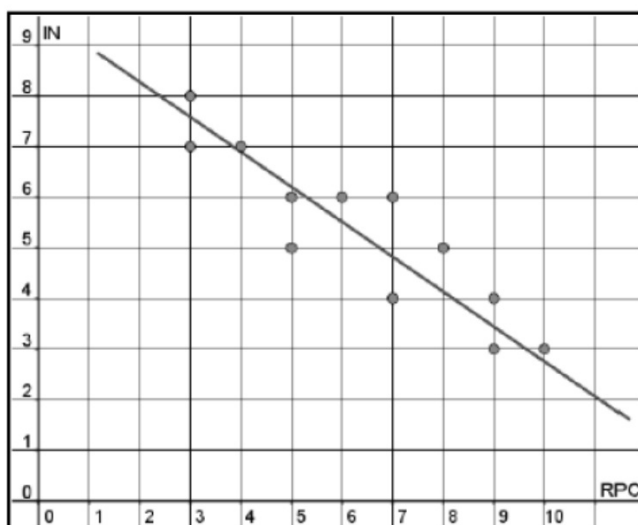


- b) Digau com és la correlació entre les dues variables (lineal o curvilínia, positiva o negativa, forta o dèbil).

6. La taula següent mostra la renda per capita (RPC) i l'índex de natalitat (IN) de dotze països:
[2 punts: 1 punt pel núvol de punts i 0,5 punts per la recta de regressió, a l'apartat a, i 0,5 punts per l'apartat b]

PAÏSOS	A	B	C	D	E	F	G	H	I	J	K	L
RPC	3	3	4	5	6	5	7	7	8	9	9	10
IN	8	7	7	6	6	5	6	4	5	4	3	3

- a) Representeu els resultats mitjançant un núvol de punts i traceu-ne aproximadament una recta de regressió.



La correlació que hi ha entre les dues variables és lineal, negativa i tirant a forta.

x_i	f_i
0	1
1	5
2	6
3	7
4	4
5	4
6	3

Calcula mitjana, variància, desviació típica i coef. de variació de Pearson

x_i	f_i	$f_i x_i$	$f_i x_i^2$
0	1	0	0
1	5	5	5
2	6	12	24
3	7	21	63
4	4	16	64
5	4	20	100
6	3	18	108
	30	92	364

MEDIA: $\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{92}{30} = 3,06$

DESVIACIÓN TÍPICA: $\sigma = \sqrt{2,52} = 1,59$

VAR.: $\frac{\sum f_i x_i^2}{\sum f_i} - \bar{x}^2 = \frac{364}{30} - 3,1^2 = 2,52$

COEFICIENTE DE VARIACIÓN: C.V. = $\frac{\sigma}{\bar{x}} = 0,5129 \rightarrow 51,29\%$

El càlcul del **coeficient de correlació lineal de Pearson** de dues variables s'estableix mitjançant el paràmetre covariància:

$\sigma_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \cdot \bar{y}$ i $r = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$. És a dir, és el quocient entre la covariància i el producte de les desviacions típiques de la x i de la y.

Calcula el coeficient de correlació d'aquesta distribució bidimensional:

x	1	3	6	8	11	12	14	15
y	12	17	16	15	20	21	18	21

Elaborem una taula de distribució per fer tots els càlculs (l'última fila és la suma de cada columna).

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
1	12	1	144	12
3	17	9	289	51
6	16	36	256	96
8	15	64	225	120
11	20	121	400	220
12	21	144	441	252
14	18	196	324	252
15	21	225	441	315
70	140	796	2.520	1.318

$$\text{Per tant: } \bar{x} = \frac{\sum x_i}{N} = \frac{70}{8} = 8,75$$

$$\bar{y} = \frac{\sum y_i}{N} = \frac{140}{8} = 17,5$$

$$\sigma_x = \sqrt{\frac{\sum x_i^2}{N} - \bar{x}^2} = \sqrt{\frac{796}{8} - 8,75^2} = \sqrt{22,9375} = 4,79$$

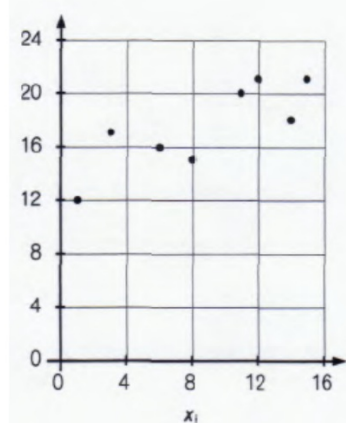
$$\sigma_y = \sqrt{\frac{\sum y_i^2}{N} - \bar{y}^2} = \sqrt{\frac{2.520}{8} - 17,5^2} = \sqrt{8,75} = 2,96$$

$$\sigma_{xy} = \frac{\sum x_i y_i}{N} - \bar{x} \bar{y} = \frac{1.318}{8} - 8,75 \cdot 17,5 = 11,225$$

$$i \ r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{11,225}{4,79 \cdot 2,96} = 0,79, \text{ que ens dona una correlació positiva forta, tal com també es veu al núvol de punts del marge}$$

$$\sigma_x = \sqrt{\frac{\sum n_i \cdot x_i}{N} - \bar{x}^2}$$

$$\sigma_{xy} = \frac{\sum x_i \cdot y_i}{N} - \bar{x} \cdot \bar{y}$$



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

7. Indiqueu quina és la correlació corresponent a cadascun dels núvols de punts.
Justifiqueu la resposta.

a) $r=0,9$

b) $r=-1$

c) $r=0$

d) $r=-0,6$

el seu
gràfic

4), ja que

3), " "

1), " "

2), " "

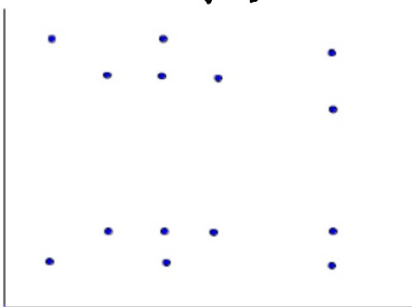
hi ha molta correlació i és positiva

la correlació és negativa i màxima (funcional)

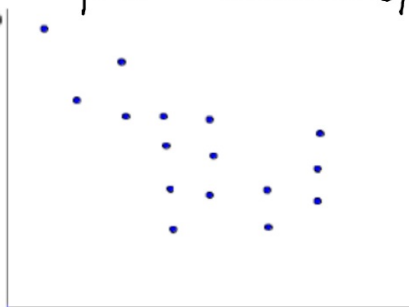
els punts estan molt dispersos i no hi ha correlació

els punts es distribueixen cap avall i molt
dèbilment

1)



2)



3)



4)



Solució Correcció:

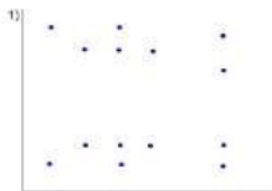
7. Indiqueu quina és la correlació corresponent a cadascun dels núvols de punts. Justifiqueu la resposta.

a) $r=0,9$

b) $r=-1$

c) $r=0$

d) $r=-0,6$



- a) $r=0,9$ ↔ 4) ja que es tracta d'una corr. lineal positiva forta.
b) $r=-1$ ↔ 3) ja que es tracta d'una corr. lineal negativa funcional.
c) $r=0$ ↔ 1) ja que no hi ha correlació.
d) $r=-0,6$ ↔ 2) ja que es tracta d'una corr. lineal negativa dèbil.

Compteu 0,5 punts per cada apartat

Recta de Regressió de Y sobre X

$$y - \bar{y} = \frac{\sigma_{xy}}{\sigma_x^2} \cdot (x - \bar{x})$$

EXEMPLE

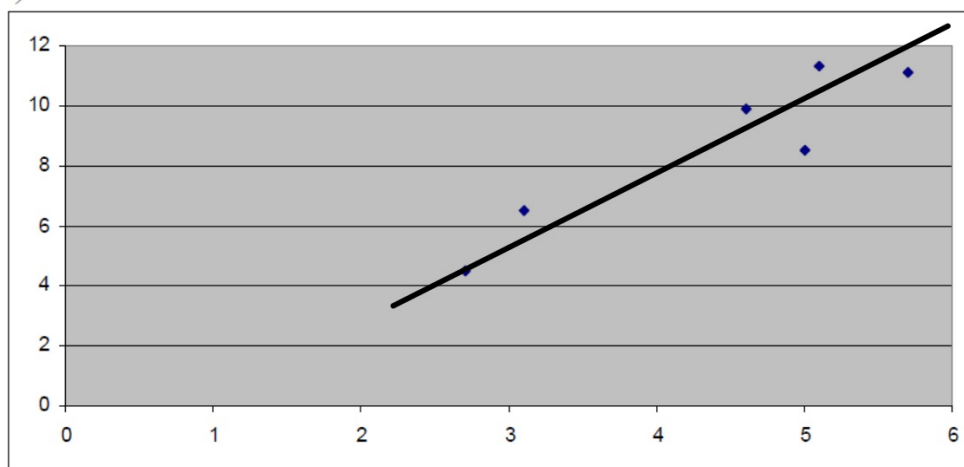
El consum d'energia per càpita en milers de KW/h i la renda per càpita en milers de € en sis països de la UE són els següents:

	Consum	Renda
<i>Alemanya</i>	5'7	11'1
<i>Bèlgica</i>	5	8'5
<i>Dinamarca</i>	5'1	11'3
<i>Espanya</i>	2'7	4'5
<i>França</i>	4'6	9'9
<i>Itàlia</i>	3'1	6'5

Es demana: a) Nívol de punts de la distribució. b) Coeficient de correlació i rectes de regressió. c) Quina predicció podem fer sobre el consum d'energia de Grècia si se sap que la renda per càpita és de 4'4 milers de €?

RAONAMENT

a)



ENERGIA x_i	RENDA y_i	$x_i y_i$	x_i^2	y_i^2
5'7	11'1	57'63	32'49	123'21
5	8'5	42'5	25	72'25
5'1	11'3	57'63	26'01	127'69
2'7	4'5	12'15	7'29	20'25
4'6	9'9	45'54	21'16	98'01
3'1	6'5	20'15	9'61	42'25
26'2	51'8	235'7	121'97	483'66

241,24

$$\bar{x} = \frac{26'2}{6} = 4'366$$

$$\sigma_x^2 = \frac{121'97}{6} - 4'366^2 = 1'2663$$

$$\sigma_y^2 = \frac{483'66}{6} - 8'633^2 = 6'081$$

$$\sigma_{xy} = \frac{\cancel{241'24}^{2,51}}{6} - 4'366 \cdot 8'633 = \frac{\cancel{1'49}^{2,51}}{6}$$

$$\bar{y} = \frac{51'8}{6} = 8'633$$

$$\sigma_x = \sqrt{1'2663} = 1'125$$

$$\sigma_y = \sqrt{6'081} = 2'466$$

$$r = \frac{\cancel{1'49}^{2,51}}{1'125 \cdot 2'466} = \cancel{0'537}^{2,51} = 0,904$$

Recta de regressió de (y/x)

$$y - 8'633 = \left(\frac{\cancel{1'49}^{2,51}}{1'2663} \right) (x - 4'366) \Leftrightarrow y = \cancel{1'176}^{1,98} x + 3'495$$

Recta de regressió de (x/y)

$$x - 4'366 = \left(\frac{1'49}{6'081} \right) (y - 8'633) \Leftrightarrow y = 4'08x - 9'185$$

c) recta de (x/y) $x = 4'366 + \left(\frac{1'49}{6'081} \right) (4'4 - 8'633) = 3'328$

és una estimació ~~poc~~ fiable ja que $r = 0,904$

Variància

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2 \cdot n_i}{n} = \frac{\sum x_i^2 \cdot n_i}{n} - \bar{x}^2$$

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2 \cdot n_i}{n} = \frac{\sum y_i^2 \cdot n_i}{n} - \bar{y}^2$$

Desviació típica o tipus

$$\sigma_x = \sqrt{\sigma_x^2} \quad \sigma_y = \sqrt{\sigma_y^2}$$

Covariància

$$\sigma_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})n_i}{n} = \frac{\sum x_i y_i n_i}{n} - \bar{x} \bar{y}$$

$\sum$ Σ sigma
p s

Coefficient de correlació

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

Recta de regressió de y sobre x (y/x)

$$y - \bar{y} = \left(\frac{\sigma_{xy}}{\sigma_x^2} \right) (x - \bar{x}) \quad \Leftrightarrow \quad y = \left(\frac{\sigma_{xy}}{\sigma_x^2} \right) x + \left(\bar{y} - \frac{\sigma_{xy}}{\sigma_x^2} \bar{x} \right)$$

Recta de regressió de x sobre y (x/y)

$$x - \bar{x} = \left(\frac{\sigma_{xy}}{\sigma_y^2} \right) (y - \bar{y}) \quad \Leftrightarrow \quad y = \left(\frac{\sigma_y^2}{\sigma_{xy}} \right) x + \left(\bar{y} - \frac{\sigma_y^2}{\sigma_{xy}} \bar{x} \right)$$

S'han analitzat sis models d'impressores amb color i en blanc i negre donant com a resultat el que figura a la taula següent on els valors representen el cost per pàgina en cèntims d'euro.

Blanc i Negre(x_i)	8	11	17	21	14	10
Color(y_i)	33	49	95	106	58	53

Trobeu a) La recta de regressió de y sobre x (y/x). b) Quant costaria imprimir una pàgina en color d'una impressora on una pàgina en blanc i negre costa 12 cèntims d'euro.

x_i	y_i	x_i^2	y_i^2	$x_i y_i$
8	33	64	1089	264
11	49	121	2401	539
17	95	289	9025	1615
21	106	441	11236	2226
14	58	196	3364	812
10	53	100	2809	530
81	394	1211	29924	5986

$$\bar{x} = 13'5 \quad \bar{y} = 65'67 \quad \sigma_x^2 = 19'58 \quad \sigma_y^2 = 674'78 \quad \sigma_{xy} = 111'12$$

$$\sigma_x = 4'425 \quad \sigma_y = 25'97 \quad r = 0'97$$

La recta de regressió de y sobre x :

$$y - 65'67 = \left(\frac{111'12}{19'58} \right) (x - 13'5) \quad \Leftrightarrow \quad y = 5'68x - 11'01$$

S'ha fet un estudi a nivell de primer de Batxillerat sobre la nota mitjana de Matemàtiques i Anglès en sis centres de secundària, com reflecteix la següent taula:

<i>Matemàtiques x</i>	6'5	5'2	6	6'5	7	6
<i>Anglès y</i>	7	5	5	6	7'5	5

Es demana: a) Recta de regressió de y sobre x b) Si un altre centre té per nota mitjana de Matemàtiques 5'5, que cal esperar de la nota d'Anglès?

Matemàt.(x_i)	Anglès(y_i)	x_i^2	y_i^2	$x_i y_i$
6'5	7	42'25	49	45'5
5'2	5	27'04	25	26
6	5	36	30 25	30
6'5	6	42'25	39 36	39
7	7'5	49	52'5 56'25	52'5
6	5	36	30 25	30
37'2	35'5	232'54	225'5	223

$$\bar{x} = 6'2 \quad \bar{y} = 5'916 \quad \sigma_x^2 = 0'31 \quad \sigma_y^2 = \frac{216,25}{6} - 5,916^2 \approx 1,040 \quad \sigma_{xy} = 0'4874 \quad r = 0'5475$$

Recta de regressió de (y/x) $y - 5'916 = \left(\frac{0'4874}{0'31} \right) (x - 6'2) \quad y = 1'57x - 3'832$

b)

$y(5'5) = 4'8$ és la nota esperada en Anglès però donat que la dependència de les variables és dolenta aleshores la nota no és fiable.

