

DISTRIBUCIONS BIDIMENSIONALS

PROBLEMA RESOLT

La taula següent mostra les notes de matemàtiques, x , i de física, y , de sis estudiants.

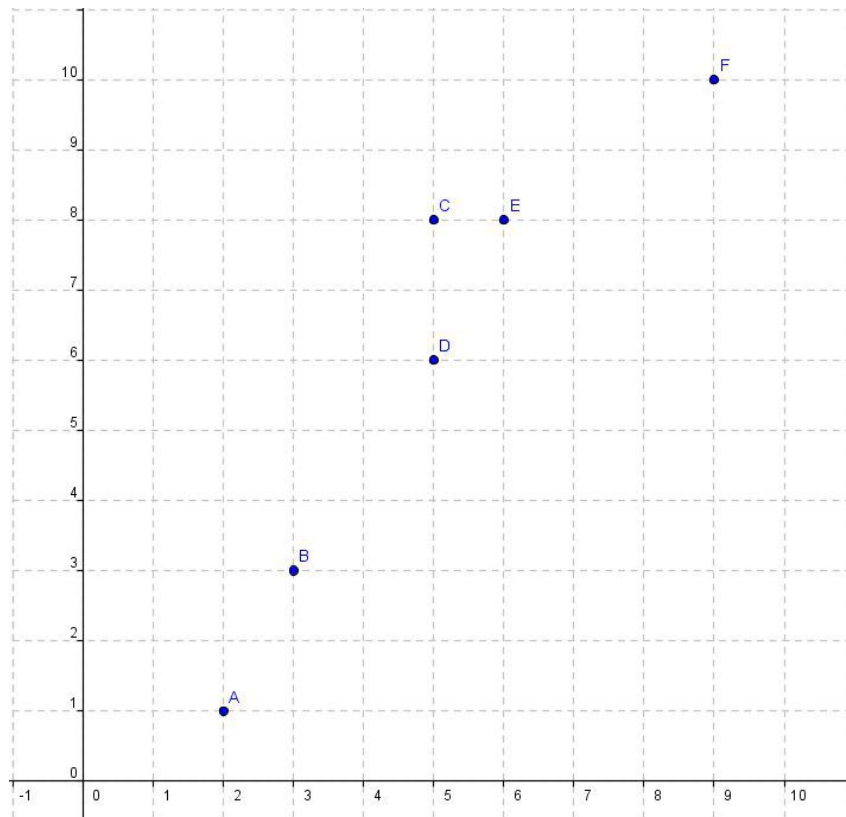
x	2	3	5	5	6	9
y	1	3	8	6	8	10

- Representeu gràficament els valors de la taula com a núvol de punts.
- Trobeu el coeficient de correlació. Creieu que les variables estan fortament relacionades?
- Trobeu la recta de regressió que ens expliqui la nota de física a partir de la de matemàtiques, i representeu-la gràficament en el sistema de representació de l'apartat a).
- Si un estudiant té un 7 de matemàtiques, quina predicció faríeu per a la nota de física?

Expliqueu detalladament tots els càlculs que feu.

RESOLUCIÓ

a) Tal i com ens indiquen, representarem aquests parells de valors en uns eixos cartesianes a fi d'obtenir el diagrama de dispersió o **núvol de punts**.



Observant aquest gràfic veiem que els punts tendeixen a alinear-se sobre una recta de pendent positiu, és a dir, podem dir que la correlació és lineal i forta (tots els punts estan gairebé alineats) i directa o positiva (el pendent d'aquesta possible recta que intuïm, anomenada recta de regressió és positiva).

A aquesta mateixa conclusió arribarem mitjançant el càlcul del coeficient de correlació lineal (apartat b).

b) L'expressió per calcular el coeficient de correlació lineal r és:

$$\text{coeficient de correlació lineal} \rightarrow r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Per calcular-lo necessitem saber el valor de la covariància i el de les desviacions típiques de les variables X i Y .

CÀLCULS PREVIS:

$$\text{Mitjana de la variable } X \rightarrow \bar{x} = \frac{2+3+5+5+6+9}{6} = 5$$

$$\text{Mitjana de la variable } Y \rightarrow \bar{y} = \frac{1+3+8+6+8+10}{6} = 6$$

Desviació típica de la variable X

$$\rightarrow \sigma_x = \sqrt{\frac{2^2+3^2+5^2+5^2+6^2+9^2}{6} - 5^2} = \sqrt{\frac{180}{6} - 25} = \sqrt{5} = 2,24$$

Desviació típica de la variable Y

$$\rightarrow \sigma_y = \sqrt{\frac{1^2+3^2+8^2+6^2+8^2+10^2}{6} - 6^2} = \sqrt{\frac{274}{6} - 36} = \sqrt{9,67} = 3,11$$

$$\text{Covariància} \rightarrow \sigma_{xy} = \frac{2 \cdot 1 + 3 \cdot 3 + 5 \cdot 8 + 5 \cdot 6 + 6 \cdot 8 + 9 \cdot 10}{6} - 5 \cdot 6 = \frac{219}{6} - 30 = 6,5$$

Ara, ja estem en condicions de calcular r :

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{6,5}{2,24 \cdot 3,11} = 0,933$$

El coeficient de correlació r val **0,933**, és a dir, un valor molt proper a 1, això ens indica que la correlació és *lineal i forta*.

I, donat que aquest valor és positiu, *la correlació és directa (positiva)*, això significa que quan augmenten (disminueixen) els valors de la variable X , també augmenten (disminueixen) els valors de la variable Y .

En el cas pràctic que ens ocupa, significa que els alumnes que tenen notes altes de matemàtiques també tenen notes altes a física i els que tenen notes baixes a matemàtiques també les tenen a física.

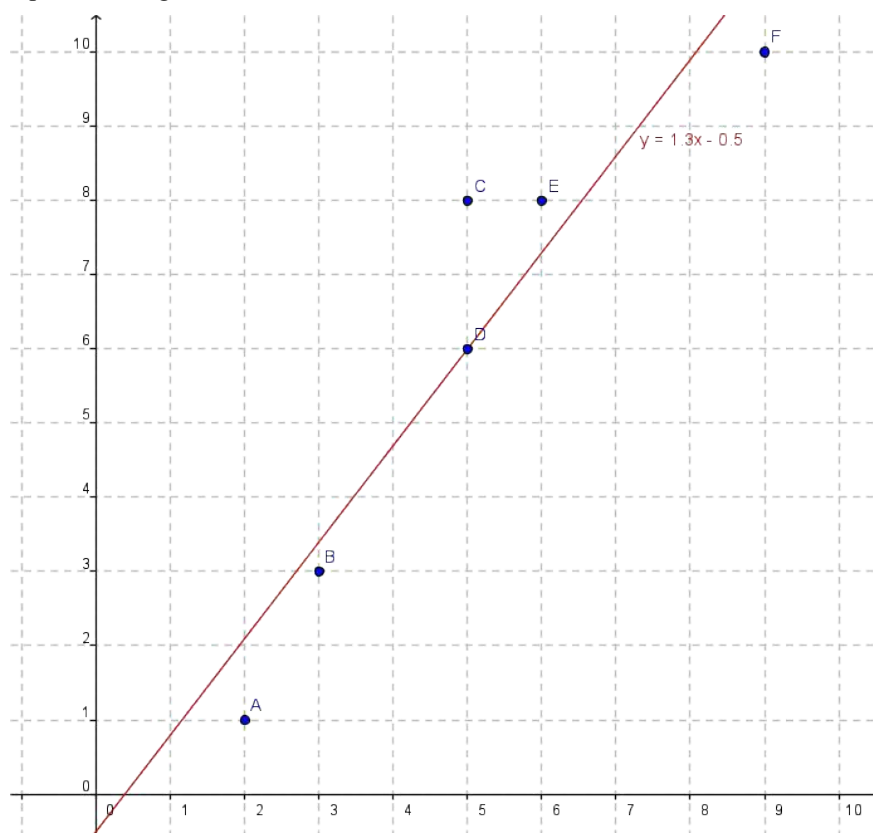
c) En aquest apartat, ens demanen trobar la recta de regressió que ens permeti trobar (estimar) la nota de física (y) a partir de la de matemàtiques (x), per tant hem de trobar la recta de regressió de Y sobre X.

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

$$y - 6 = \frac{6,5}{5}(x - 5)$$

$$y = 1,3x - 0,5$$

I, la seva representació gràfica és:



d) Ara, el que ens demanen és estimar la nota de física que traurà un alumne que ha tret un 7 a matemàtiques.(els valors estimats es representen com a \hat{y}). Aquest valor de \hat{y} el trobarem a partir de la recta de regressió:

$$\hat{y} = 1,3x - 0,5$$

$$\hat{y} = 1,3 \cdot 7 - 0,5$$

$$\hat{y} = 8,6$$

Atès que la correlació és forta, el risc d'equivocar-nos quan estimem la nota de física és baix. Això significa que un alumne que ha tret un 7 de matemàtiques obtindrà una nota de física de 8,6 o, si més no, un valor molt proper.