Exercices : Séries statistiques à deux variables. ***Révisions autonomes***.

2) Soit la série statistique à deux variables X et Y suivante :

X	2	3	4	5	6
Y	3	4	7	10	10

a) Représenter graphiquement cette série statistique.

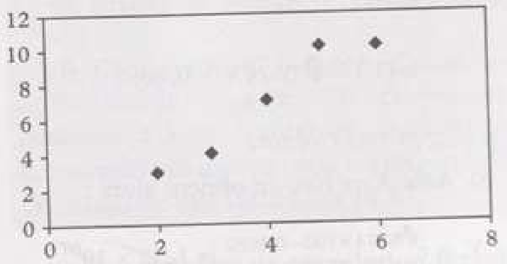
b) Les points obtenus sont-ils grossièrement alignés ?

c) Calculer le coefficient de **corrélation linéaire** de cette série. Ce résultat corrobore-t-il la réponse de la question précédente ?

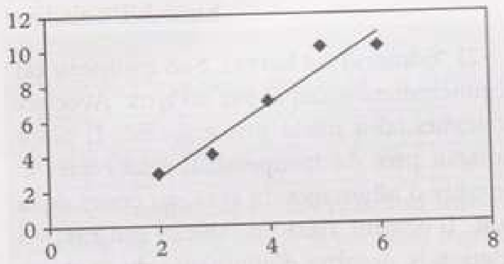
d) Calculer les coefficients a et b de **régression linéaire** par la méthode des **moindres carrés.**

Correction (obtention des différents coeefficients à l'aide de la calculatrice graphique).

a) Voici le graphe :



b) Les points obtenus ne sont pas alignés. Cependant, ils donnent grossièrement une direction, indiquée sur le graphe ci-dessous :



Ils sont donc grossièrement alignés.

c) Pour calculer le coefficient de **corrélation linéaire**, il faut un tableau plus complet :

x_i	2	3	4	5	6	20
y_i	3	4	7	10	10	34
x_i^2	4	9	16	25	36	90
y_i^2	9	16	49	100	100	274
x_iy_i	6	12	28	50	60	156

Le coefficient de **corrélation linéaire** est alors :

$$r = \frac{\mathrm{Cov}(X,\,Y)}{\sigma(X)\times\sigma(Y)} = \frac{\frac{156}{5}-\frac{20}{5}\times\frac{34}{5}}{\sqrt{\frac{90}{5}-\left(\frac{20}{5}\right)^2}\sqrt{\frac{274}{5}-\left(\frac{34}{5}\right)^2}} \approx 0,9667$$

Ce résultat est très proche de 1. On peut considérer que la **régression linéaire** est réussie. Cela corrobore la réponse de la question précédente.

d) Les calculs des coefficients a et b sont :

$$a = \frac{\mathrm{Cov}(X,\,Y)}{\mathrm{Var}\,(X)} = \frac{\frac{156}{5}-\frac{20}{5}\times\frac{34}{5}}{\frac{90}{5}-\left(\frac{20}{5}\right)^2} = 2$$

$$\text{et } b = \overline{Y} - a\overline{X} = \frac{34}{5} - 2\times\frac{20}{5} = -1,2$$

Remarque :

Ces coefficients (r, a et b) se calculent très facilement à l'aide de calculatrices scientifiques. Aussi, dans les exercices suivants, seules seront inscrites les valeurs des coefficients.

Enoncé :

2) Edmond est berger. Son troupeau est régulièrement attaqué par un lynx. Avec ses collègues, il a mené une enquête. Il place un spot près du troupeau. Il peut régler le nombre d'allumages du spot, au cours de la nuit. Il obtient ainsi un tableau qui met en relation le nombre d'allumages du spot, au cours d'une nuit, et le pourcentage d'attaque par le lynx, au cours de cette même nuit :

Nb d'allumages	16	27	45	74	250
Chances d'attaque (%)	16,75	15,5	14,25	13	10,5

Soit X la variable correspondant au nombre d'allumages, et Y la variable correspondant au pourcentage d'attaque.

a) Calculer le coefficient de **corrélation linéaire** de cette série.

b) Edmond veut établir une relation entre ces deux données, et veut que cette relation soit sûre à plus de 99 %. Est-ce que la relation linéaire a cette propriété ?

c) Il décide alors de poser la variable $Z = \ln(X)$. Construire le tableau liant les variables Y et Z.

d) Calculer le coefficient de **corrélation linéaire** de cette série. Est-il meilleur que le précédent ? Répond-il aux espérances d'Edmond ?

e) Établir la meilleure relation entre X et Y.

f) Déterminer le nombre d'allumages qu'il faudrait pour que le pourcentage descende à 5 %. À combien de flashs par heure, puis par minute, cela correspond-il si on considère que la nuit fait 12 h ?

Correction : (obtention des différents coeefficients à l'aide de la calculatrice graphique).

a) Le coefficient de **corrélation linéaire** est $r \approx -0,9220$.

b) Ce coefficient est inférieur, en valeur absolue, à 0,9587, donc il ne respecte pas le souhait d'Edmond.

c) Voici le tableau demandé :

Z	2,8	3,3	3,8	4,3	5,5
Y	16,75	15,5	14,25	13	10,5

d) Le coefficient de **corrélation linéaire** est alors : $r \approx -0,9987$. Il est bien meilleur que le précédent et répond aux exigences d'Edmond.

e) Avec cette série, on a alors :

$$a \approx -2,32 \qquad \text{et} \qquad b \approx 23,13$$

Soit

$$Y = -2,32Z + 23,13$$

On peut alors écrire :

$$Y = -2,32\ln(X) + 23,13$$

f) Si on veut que $Y = 5$, il faut que : $-2,32\ln(X) + 23,13 = 5$

La solution de cette équation est :

$$X \approx 2476,6$$

Cela correspond à 206,4 flashs par heure, soit 3,4 par minute.

Enoncés et réponses partielles :

Calculer le coefficient de **corrélation linéaire** des séries statistiques à deux variables suivantes :

6) Série n° 1

x_i	2	4	6	8	10	12
y_i	1	1	2	2	4	4
n_i	1	2	1	3	3	2

RF : $r \approx 0,923$

7) Série n° 2

x_i	10	1	9	2	12	12
y_i	1	9	8	-1	2	3
n_i	2	2	4	1	1	2

RF : $r \approx -0,302$

Calculer les coefficients a et b de **régression linéaire** des séries statistiques à deux variables suivantes :

8) Série n° 1

x_i	2	4	6	8	10	12
y_i	1	1	2	2	4	4
n_i	1	2	1	3	3	2

RF : $a \approx 0,373 ; b \approx -0,34$

9) Série n° 2

x_i	3	1	0	-2	2	5
y_i	5	1	-1	-5	3	9
n_i	5	4	3	2	3	4

RF : $a = 2 ; b = -1$

10) Série n° 3

x_i	10	1	9	2	12	12
y_i	1	9	8	-1	2	3
n_i	2	2	4	1	1	2

RF : $a \approx 0,268 ;\ b \approx 7,06$