TD-N° 02 : Statistique bidimensionnelle-Régression linéaire

Exercice 1 *On veut étudier la relation entre le QI du père* (X) *et le QI du fils* (Y)*. Sur un échantillon aléatoire de 12 couples (père, fils), on a relevé le QI du père et le QI du fils. On dispose de 2 échantillons appariés de mesures* $(x_i; y_i)$ *:*

couple (père,fils) Num i	1	2	3	4	5	6	7	8	9	10	11	12
QI du père x_i	123	144	105	110	98	138	131	90	119	109	125	100
QI du fils y_i	102	138	126	133	95	146	115	100	142	105	130	120

- *Placer dans un repère orthogonal les points* (x_i, y_i) *correspondant à ces mesures.*
- *Calculer le coefficient de corrélation linéaire de* X *et* Y*.*
- *Un ajustement linéaire semble-t-il justifié ? Si oui, calculer les paramètres de la droite de régression de* Y *en* X*.*
- *Calculer les coordonnées du point moyen* G *du nuage de points correspondant.*
- *Exploiter cette droite d'ajustement pour déterminer le QI d'un fils dont le QI de père est 120.*

On donne les quantités suivantes : $\sum x_i = 1392$ *;* $\sum x_i^2 = 164566$ *;* $\sum y_i = 1452$ *;* $\sum y_i^2 = 179068$ *;* $\sum x_i y_i = 170394$.

Exercice 2 *On étudier l'influence d'un antibiotique sur une culture bactérienne.*

Partie A *: On répartit dans 10 tubes des volumes égaux de culture additionnées d'une quantité X d'antibiotique, et on mesure, après incubation, la densité optique D. La densité optique permet de déterminer la concentration en bactérie du milieu de culture.*

Antibiotique X	0.2	0.2	0.4	0.4	0.6	0.6	0.8	0.8	1	1
densité optique D	19	21	35	38	64	66	115	130	200	210

1) *Construire le nuage des points représentant la densité optique en fonction de la concentration d'antibiotique.*
2) *Calculer le coefficient de corrélation linéaire entre* X *et* D*.*
3) *Un ajustement linéaire semble-t-il justifié ?*

Partie B : *On reprend l'analyse en posant* $Z = ln(D)$*, où ln représente le logarithme népérien.*

Z=ln(D)	2.94	3.045	3.56	3.64	4.16	4.19	4.74	4.87	5.3	5.35

1) *Reprendre les questions 1, 2 et 3 de la partie A.*
b) *Calculer les moyennes des observations* x_i *et* z_i *ainsi que les écarts-types (S) de ces observations.*
2) *Déterminer la droite de régression de* Z *en* X *et calculer le coefficient de détermination.*
3) *En déduire une expression de D en fonction de X. Justifier le modèle utilisé.*

NB *:* $\sum x_i = 6$ *;* $\sum x_i^2 = 4.4$ *;* $\sum d_i = 898$ *;* $\sum d_i^2 = 126148$ *;* $\sum x_i d_i = 721.2$ *;* $\sum z_i = 41.79$ *;* $\sum z_i^2 = 181.57$ *;* $\sum x_i z_i = 27.42$.

$\mathscr{B}.\ \mathscr{T}.$

Rappels

- *La variance de* X *est* **estimé** *par (dans le cas d'un échantillon) :* $S_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \sum_{i=1}^{n}(x_i)^2 - (\overline{X})^2$
- *La covaiance de* X *et* Y *est par :* $Cov(X,Y) = S_{X,Y} = \frac{1}{n}\sum_{i=1}^{n}(x_i y_i) - (\frac{1}{n}\sum_{i=1}^{n} x_i)(\frac{1}{n}\sum_{i=1}^{n} y_i)$
- *La pente* β *est donnée par la formule suivante :*

$$b = \frac{S_{X,Y}}{S_X^2}$$

- $a = m_Y - b m_X = \overline{Y} - b\,\overline{X}$ *où* $m_Y = \overline{Y} = \frac{\sum y_i}{n}$ *et* $m_X = \overline{X} = \frac{\sum x_i}{n}$
- *Le cœfficient de corrélation entre* X *et* Y *:* $r_{X,Y} = corr(X,Y) = \frac{S_{X,Y}}{S_X\,S_Y}$.