TD-N° 03 : Régression linéaire

Exercice 1 *Le tableau ci-dessous contient les tailles en centimètres (échantillon x) et les poids en kilogrammes (échantillon y) de 10 enfants de 6 ans :*

x	121	123	108	118	111	109	114	103	110	115
y	25	22	19	24	19	18	20	15	20	21

- *Dessiner le nuage des points de ces observations.*
- *Sur le même repère, dessiner la droite de régression $D_{y/x}$ (de Y sur X) sachant que $\overline{x} = 113.2$, $\overline{y} = 20.3$, $s_x'^2 = 38.62$, $s_y'^2 = 8.46$, $\widehat{cov}(X,Y) = s'_{xy} = 16.27$.*
- *Calculer le coefficient de corrélation linéaire (estimer) de X et Y.*
- *En utilisant le résultat de la question précédente, donner une prédiction de la taille d'un enfant de 6 ans qui pèse 17 kg.*

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

B. T.

Rappels

- *La variance de X est* **estimé** *par (dans le cas d'un échantillon) :* $S_X'^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i)^2 - n(\overline{X})^2}{n-1}$
- *La covaiance de X et Y est* **estimé** *par :* $\widehat{Cov}(X,Y) = S'_{X,Y} = \frac{\sum_{i=1}^{n}(x_i y_i) - \frac{1}{n}(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n-1}$
- *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* $m_X = \overline{X} = \frac{\sum x_i}{n}$
- *Estimation du cœfficient de corrélation entre X et Y :* $\hat{r} = \hat{cor}(X,Y) = \frac{S'_{X,Y}}{S'_X\, S'_Y}$.