Feuille de TD n°6
Régression
Durée : 1 semaine

Exercice 1 :
On considère un nuage de points $(x_i, y_i)_{1\leq i\leq n}$. On se place dans le cadre de la régression linéaire simple et on cherche donc $\hat{a} \in \mathbb{R}$ et $\hat{b} \in \mathbb{R}$ tels que

$$\sum_{i=1}^{n} \left(y_i - \left(\hat{a}x_i + \hat{b}\right)\right)^2$$

soit le plus petit possible.

1. Démontrer par deux méthodes que

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n(\bar{x})^2} \quad \text{et} \quad \hat{b} = \bar{y} - \hat{a}\bar{x}.$$

2. Montrer que $\hat{a}$ vérifie également

$$\hat{a} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

3. Montrer que si l'on note r le coefficient de corrélation linéaire et r^2 celui de détermination, on a bien $(r)^2 = (r^2)$.
4. Montrer la relation suivante:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2.$$

Exercice 2 :
On considère un nuage de points $(x_i, y_i)_{1\leq i\leq n}$. Déterminer $\hat{a}$ et $\hat{b}$ tels que l'équation de régression s'écrit

$$y = \hat{a} + \hat{b}x.$$

Exercice 3:
Une étude sur le développement et la croissance de plants de canola provenant d'un nouveau croisement est en cours. Une équipe choisit au hasard des plants cultivés dans un terreau identique et mesure le poids des feuilles et des racines (en g). Les résultats sont résumés dans le tableau ci-dessous :

Poids des feuilles	Poids des racines
106,5	85,2
155,2	90,1
155,6	88,4
122,8	85,3
125,9	80,5
130,8	81
136,2	87,8
145,7	86,3
117,7	78,8
135,4	88,7
141,5	89,5
112	75,2
139,5	86,6
114,8	82,4
120,4	87,1

1. Dessiner le nuage de point associé à ce tableau.
2. On souhaite expliquer le poids des racines à l'aide du poids des feuilles en utilisant une régression linéaire simple. Donner l'équation de régression.
3. Calculer (à l'aide d'une calculatrice) le coefficient de corrélation ainsi que celui de détermination. Les interpréter.
4. Que peut-on prévoir comme poids des racines pour un plant dont le poids des feuilles est 95g?, 110g?, 150g?

Exercice 4 :
Des artilleurs procèdent à des tirs de canon. Pour un certain canon, la portée Y est fonction de l'angle de tir X. Voici les résultats.

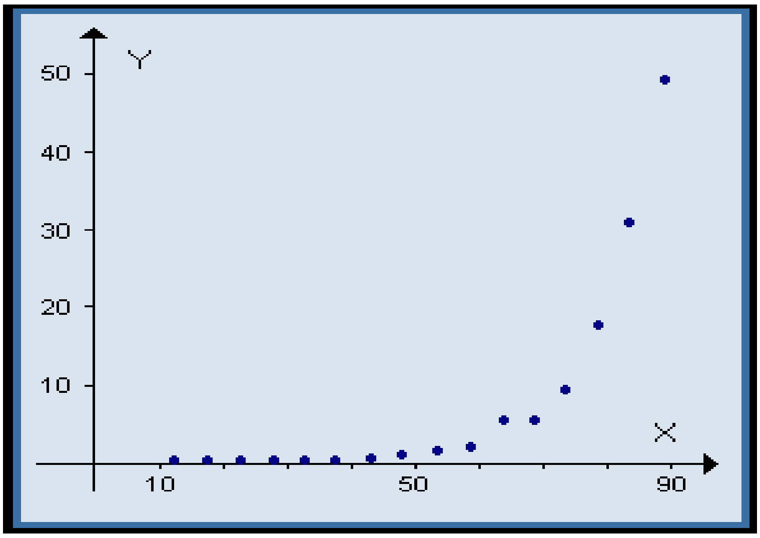
X	Y
5	177
10	349
15	510
20	655
25	781
30	883
35	958
40	1004
45	1019
50	1004
55	958
60	883
65	781
70	655
75	510
80	349
85	177

1. Calculer le coefficient de corrélation entre X et Y.
2. Pouvez-vous en conclure qu'il n'existe pas de lien entre X et Y?
3. Effectuer le nuage de points (X,Y).
4. Que pouvez-vous en conclure?

Exercice 5 :
Le quotient de mortalité est la probabilité de décéder avant l'âge x+5 années pour les individus ayant atteint l'âge x.
Voici pour le Limousin, les données relatives au sexe féminin et concernant la période 1981-82-83.
L'âge a été réparti en classes d'amplitude de 5 années.

Age	Y=Quotient en %
]10-15]	0.15
]15-20]	0.16
]20-25]	0.21
]25-30]	0.26
]30-35]	0.35
]35-40]	0.57
]40-45]	0.57
]45-50]	1.21
]50-55]	1.69
]55-60]	2.33
]60-65]	3.51
]65-70]	5.72
]70-75]	9.36
]75-80]	17.68
]80-85]	30.85
]85-90]	49.09

On définit par X le centre des classes et Y le quotient en pourcentage.
Voici le nuage de points associé à ces données :



1. A partir de cette représentation graphique, vous semble-t'il possible d'envisager une régression linéaire entre X et Y? Si non, quelle fonction f envisageriez-vous pour le lien $Y = f(X)$?
2. On pose $Z = \ln Y$. Calculer les valeurs de X et Z.
3. Réalisez le nuage de points (X, Z).
4. Que pouvez-vous en dire? Comment pouvez-vous valider votre hypothèse? Le faire.
5. Calculez les coefficients de la droite de régression entre X et Z et en déduire le lien entre X et Y.
6. En déduire à 0.01 près le quotient de mortalité pour une femme agée de 50 ans, 55 ans et 60 ans.