Université de Nice Sophia-Antipolis
L3 MASS
Feuille de travaux dirigés

Année Universitaire 2008/2009
Statistique

Feuille de TD n°4

Exercice n°1 :
Nous considérons le modèle linéaire simple permettant d'expliquer une variable aléatoire Y, encore appelée variable dépendante ou expliquée, au moyen d'une variable X, encore appelée variable indépendante ou explicative.
Nous supposons donc avoir n observations notées $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ telles que :

$$Y_i = \alpha + \beta X_i + \epsilon_i$$

avec ϵ_i vérifiant :

$$\forall\ i\ \in\ \{1, \ldots, n\}, \quad \mathbb{E}[\epsilon_i] = 0, \qquad Var[\epsilon_i] = \sigma^2, \qquad cov(\epsilon_i, \epsilon_j) = 0 \text{ pour i différent de j}$$

On souhaite obtenir une estimation de α et β par la méthode des moindres carrés.

1. Rappeler le principe de la méthode des moindres carrés.
2. Démontrer le résultat de cours qui dit :

$$\begin{aligned} \hat{\alpha} &= \bar{Y_n} - \hat{\beta}.\bar{X_n} \\ \hat{\beta} &= \frac{\sum Y_i.X_i - n.\bar{Y_n}.\bar{X_n}}{\sum X_i^2 - \bar{X_n}^2} = \frac{\sum (X_i - \bar{X_n})(Y_i - \bar{Y_n})}{\sum (X_i - \bar{X_n})^2} \end{aligned}$$

3. Démontrer le résultat de cours qui dit :

$$\begin{aligned} V(\hat{\alpha}) &= \frac{\sigma^2}{n} \frac{\sum X_i^2}{\sum (X_i - \bar{X_n})^2} \\ V(\hat{\beta}) &= \frac{\sigma^2}{\sum (X_i - \bar{X_n})^2} \\ cov(\hat{\alpha}, \hat{\beta}) &= \frac{-\bar{X_n}.\sigma^2}{\sum (X_i - \bar{X_n})^2} \end{aligned}$$

Exercice n°2 :
Une étude d'influence de la température sur un processus est menée. Les observations sont :

Temperature x	-5	-4	-3	-2	-1	0	1	2	3	4	5
Processus y	1	5	4	7	10	8	9	13	14	13	18

1. Estimer les paramètres du modèle linéaire simple sur ces données.

2. Représenter la droite de régression sur le nuage de points. Pensez-vous que l'ajustement est correct?
3. A l'aide du coefficient de détermination, valider votre postulat. Calculer le coefficient de corrélation.
4. Tester la nullité de la pente de la droite de régression.
5. Donner un intervalle de confiance pour cette pente.

Exercice n°3 :
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.

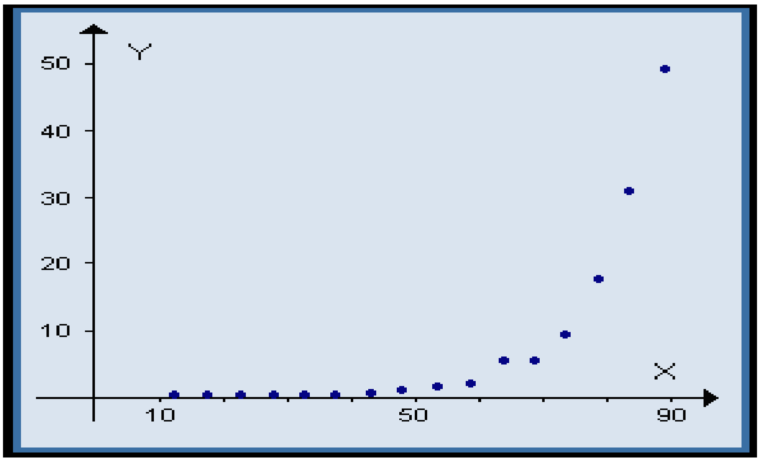
X	Y
5	177
10	349
15	510
20	655
25	781
30	883
35	958
40	1004
45	1019
50	1004
55	958
60	883
65	781
70	655
75	510
80	349
85	177

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 4 :
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 me lien $Y = f(X)$?
2. On pose $Z = log(Y)$. 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 Z et X.
6. Représentez le nuage des résidus.
7. En déduire le lien f entre Y et X.
8. En déduire à 0.01 près le quotient de mortalité pour une femme agée de 50 ans, 55 ans et 60 ans.