Université de Biskra - FSESNV
Département de Mathématiques
Module: Analyse de Données-Master 1 -2021/2022

TD N°1 : Régression Linéaire Simple

Exercice 1. Soit le modèle de régression simple sans canstante (passe par l'origine) :

$$y_j = a x_j + \varepsilon_j, \quad j = 1, ..., n \text{ et } \varepsilon \sim N\left(0, \sigma^2\right).$$

1) En utilisant la méthode des moindres carrés, donner l'expresion de l'estimateur $\widehat{a}$ de a.
2) Montrer que $\widehat{a}$ est un estimateur sans biais de a. Calculer la variance de $\widehat{a}$.

Exercice 2. Considérons les deux modèles de régression suivants:

$$z_j = \alpha v_j + \beta + \zeta_j \quad et \quad y_j = a x_j + b + \varepsilon_j, \quad \zeta \sim N\left(0, \sigma_1^2\right) \text{ et } \varepsilon \sim N\left(0, \sigma_2^2\right) \quad j = 1, ..., n$$

où $z_j = y_j - \overline{y}$ et $v_j = x_j - \overline{x}$. Montrer que $\widehat{\beta} = 0$ et que $\widehat{\alpha} = \widehat{a}$.

Exercice 3. La distribution suivante montre une relation linéaire simple :

$$Y = aX + b + \varepsilon, \quad j = 1, ..., 10 \text{ et } \varepsilon \sim N\left(0, \sigma^2\right).$$

entre l'évolution du taux de croissance économique Y(en %) et le prix de pétrole X(en \$) pendant 10 ans :

x (\$)	111	120	122	105	101	95	92	90	96	102
y (%)	1.9	4.4	5.6	4.1	3.2	-3	-2	0	-2.9	1.1

1) Donner la représentation graphique des données.

2) Completez le tableau et estimez les paramètres du modèle.

3) Calculer la valeur de $\hat{\sigma}_\varepsilon^2 = s^2$.

4) Calculer les quantités : $\hat{\sigma}_1^2$ *et* $\hat{\sigma}_0^2$ (i.e., estimateurs des variances de a et b resp.).

5) Testez la signification des paramètres du modèle au niveau 95%.

6) Déterminer l'intervalle de confiance à 90% des paramètres a et b.

7) Testez la validité du modèle.

8) Quelle quantité mesure la bonne explication de la variable à expliquer par la variable explicative? Que vaut-elle ici ? Est-ce une valeur satisfaisante ?

9) Donnez le taux de croissance lorsque le prix du pétrole est 130\$?

10) Considérant le cas du prix du pétrole 100\$, estimer la valeur moyenne correspondante du taux de croissance. Et donner une limite supérieure au seuil $\alpha = 10\%$.

11) Supposons que pour un prix du pétrole 100\$, le taux de croissance est 5%. Cette valeur peut-elle être considérée, à 90% d'accord avec l'équation d'estimation trouvée ?

<u>Exercice 4.</u> Considérons la série y_t des indices trimestriels de la production industrielle de 2006 à 2012. On modélise l'indice trimestriel y_t de la production industrielle par :

$$y_t = at + b + \varepsilon_t, \qquad t = 1, ..., 28$$

correspondant à chacun des trimestres entre 2006 et 2012, avec $\varepsilon \sim N(0,1)$. On donne

$$\sum_{t=1}^{28} y_t = 3647, \qquad \sum_{t=1}^{28} y_t^2 = 478765, \qquad \sum_{t=1}^{28} t y_t = 53500.$$

1) Donner une estimation de l'équation de la tendance (droite de régression).

2) Calculez le coefficient de détermination D de la régression.

3) Quelles sont les valeurs prédites de l'indice pour les années 2016 et 2017.

<u>Exercice 5.</u> La loi de *Boyle* relie la pression P d'un gaz au volume V selon l'équation :

$$PV^\alpha = \beta, \qquad \alpha \text{ et } \beta \text{ sont deux constantes.}$$

1) Donner une estimation des paramètres α et β du modèle à l'aide des données :

P	0.6	1	1.5	2	2.5	3.5
V	0.2	1	0.8	1	1.6	0.4

2) Calculer une prévision du pression pour un volume $V = 0.5$.

<u>Exercice 6.</u> On considère le modèle linéaire : $y_j = ax_j + b + \varepsilon_j, \quad j = 1, .., n$, avec $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma_\varepsilon^2$ et $E(\varepsilon_i \varepsilon_j) = 0 \ (i \neq j)$. Notons y_p la valeur prédite de y quand $x = x_p$. Montrer que l'erreur de prévision $\hat{\varepsilon}_p = y_p - \hat{y}_p$ verifiée la propriété suivante :

$$Var(\hat{\varepsilon}_p) = \sigma_\varepsilon^2 \left(1 + \frac{1}{n} + \frac{(x_p - \overline{x})^2}{\sum (x_j - \overline{x})^2} \right).$$