



Modèles Linéaires

Devoir Maison Licence 3 MIASHS (2023 - 2024)

Guillaume Metzler
Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

Résumé

Le présent travail est à rendre, sur feuille, pour le jeudi 7 mars 2024.
Il vous permettra de vous préparer pour votre premier contrôle du 14 mars 2024.

| | | | | | | |
|-------|----|---|----|-----|-----|-----|
| y | 1 | 4 | 2 | 4.5 | 2.2 | 4.2 |
| x_1 | -2 | 3 | -1 | 1 | -4 | 3 |
| x_2 | 0 | 2 | 0 | 0 | 0 | -2 |

TABLE 1 – Tableau des données : y représente les valeurs de la variable dépendante, x_1 et x_2 les valeurs des covariables indépendantes.

Etude d'un problème de régression


On considère le jeu de données présenté en Table 1.

On considère le modèle de régression multiple suivant

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

1. Rappeler les hypothèses du modèle linéaire gaussien.
2. Dans le cas présent, définir les objets \mathbf{y} , \mathbf{X} , $\boldsymbol{\beta}$ et $\boldsymbol{\varepsilon}$ ainsi que leurs dimensions.
3. On considère le problème d'optimisation

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2.$$

- (a) Donner l'expression littérale de $\hat{\boldsymbol{\beta}}$, l'estimateur de $\boldsymbol{\beta}$.
 - (b) Calculer sa valeur numérique, à la main. On donnera un résultat précis au centième près, soit deux chiffres après la virgule.
4. A l'aide du logiciel .
 - (a) Tester la significativité des paramètres du modèle. On prendra le soin de définir le test employé, la définition et la valeur de la statistique de test ainsi que la distribution associée à ce test.
 - (b) Comment tester si le modèle est globalement significatif? Définir le test, calculer sa valeur et conclure quant à la significativité du modèle.
 5. On s'intéresse maintenant à la qualité du modèle.

- (a) Calculer le coefficient de détermination R^2 du modèle de régression construit.
 - (b) En déduire la valeur du coefficient de détermination ajusté R_{aj}^2 .
 - (c) Calculer le BIC du modèle.
6. Retour sur un modèle linéaire simple. On considère le modèle suivant, qui n'utilise que la covariable \mathbf{x}_1 , *i.e.*,

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \varepsilon.$$

- (a) Effectuer la régression linéaire simple, à la main, et déterminer le coefficient de détermination ajusté de modèle.
- (b) Comparer le résultat obtenu au modèle multiple et dire qu'elle est le meilleur modèle.