

Série 3

Exercice 1

Une grosse société veut prédire la satisfaction au travail qu'éprouvent ses employés. Elle a recueilli des données chez 15 employés qui ont chacun donné des informations sur leur satisfaction au travail (Y), le degré de responsabilité (X_1), le nombre de personnes supervisées (X_2), l'évaluation de l'environnement de travail (X_3) et les années de service (X_4). Les données sont les suivantes :

Satisfaction	2	2	3	3	5	5	6	6	6	7	8	8	8	9	9
Responsabilité	4	2	3	6	2	8	4	5	8	8	9	6	3	7	9
Personnes supervisées	5	3	4	7	4	8	6	5	9	8	9	3	6	9	9
Environnement	1	1	7	3	5	8	5	5	6	4	7	2	8	7	9
Années de service	5	7	5	3	3	6	3	2	7	3	5	5	8	8	1

- Proposez un modèle linéaire pour décrire la satisfaction des travailleurs à l'aide des quatre prédicteurs.
- Estimez les paramètres de votre modèle.
- Calculez l'erreur standard de régression.
- Calculez la valeur du coefficient de corrélation multiple.
- Calculez les résidus e_i et les scores ajustés \hat{y}_i .
- Montrez que la corrélation entre les résidus et les scores ajustés est égale à 0.
- Montrez que les résidus ne sont pas du tout corrélés aux prédicteurs.
- Calculez la corrélation entre les scores prédits \hat{y}_i et les scores observés y_i . Comparez ce coefficient de corrélation au coefficient de corrélation multiple.
- La liaison globale entre Y et les variables explicatives X_1 , X_2 , X_3 et X_4 est-elle significative au seuil de 5% ?

Exercice 2

Dans le fichier `Angell.csv`, 43 villes des États-Unis sont caractérisées selon 3 variables¹. Ces variables sont l'intégration morale (*i.e.* intégration sociale) (INT), l'hétérogénéité ethnique (HET) et la mobilité géographique (MOB).

L'intégration morale est un indice qui combine des informations sur la criminalité, d'une part, et sur l'assistance sociale, d'autre part. L'hétérogénéité ethnique est un indice calculé à partir de la proportion des habitants de couleur et de celle des habitants blancs nés à l'étranger. La mobilité géographique est une variable construite à partir du nombre d'habitants entrant et sortant quotidiennement de la ville.

- Représentez toutes les relations binaires entre les variables INT, HET et MOB.
- Estimez les paramètres β_0 , β_1 et β_2 de l'équation :

$$\text{INT} = \beta_0 + \beta_1 \cdot \text{HET} + \beta_2 \cdot \text{MOB}.$$

- Interprétez les résultats de vos calculs.
- Les données sont-elles bien décrites par le modèle ? Pour répondre à cette question calculez l'erreur standard de régression.
- Calculez le coefficient de corrélation multiple.
- Calculez également le coefficient de détermination ajusté.
- La liaison globale entre INT et les variables explicatives HET et MOB est-elle significative ? Pour répondre à cette question réalisez un test au seuil de 5%.

1. Angell, R. C. (1951). The moral integration of American Cities, *American Journal of Sociology*, 57(2), 1-140.

Exercice 3

Dans le cadre du projet d'immersion linguistique de l'Université Catholique de Louvain (Belgique), des chercheurs se sont intéressés à la performance en français (résultats en français) selon certaines caractéristiques individuelles chez les enfants de 14 ans. Pour mener à bien ce projet, plusieurs informations concernant 360 enfants ont été recueillies, parmi lesquelles, le quotient intellectuel verbal (*QIV*), le nombre d'amis à 14 ans (*Amis*), le nombre d'heures d'étude moyen par semaine durant la période des examens (*Heures*) et les résultats en français à 14 ans (*Scores*). Ces informations sont contenues dans le fichier `Immersion.csv`.

- a) Calculez l'ensemble des corrélations simples de l'ensemble des variables contenus dans le fichier `Immersion.csv`.
- b) Etablissez l'équation de régression multiple sous-jacente à la situation du projet puis construisez le modèle complet.
- c) Que valent les paramètres de ce modèle ?
- d) Que vaut l'erreur standard de régression ?
- e) Quelle est la valeur du coefficient de corrélation multiple au carré ?
- f) Le modèle de régression multiple est-il globalement significatif au seuil de 5% ?
- g) Quelles sont les variables significatives au seuil de 5% dans le modèle ? Justifiez vos réponses ?