

Série 4

Exercice 1

Les psychométriciens ont souvent voulu mettre en lien la personnalité et les aptitudes. C'est ainsi, par exemple, qu'un concept comme l'intelligence émotionnelle est apparu. Certains auteurs postulent que le facteur *Ouverture* du BIG-5 est lié à l'intelligence. Cette hypothèse a été corroborée par Ashton, Lee et Vernon (2000). Ces derniers montrent plus précisément, dans leur étude, que l'*Ouverture* est liée à l'intelligence cristallisée (à l'exception de l'arithmétique), mais pas à l'intelligence fluide. Nous avons observé quelques variables sur un échantillon de 75 personnes. Les données enregistrées dans le fichier `Ouverture.csv` (les séparateurs de champ sont des “,”) contiennent six variables :

Variable	Sous-échelle	Concept mesuré
OUV	Ouverture	Trait de personnalité
ARI	Arithmétique	Facette de l'intelligence cristallisée
INF	Information	Facette de l'intelligence cristallisée
VOC	Vocabulaire	Facette de l'intelligence cristallisée
COM	Compréhension	Facette de l'intelligence cristallisée
SPA	Spatial	Facette de l'intelligence fluide

- Calculez l'ensemble des corrélations simples entre les six variables du fichier `Ouverture.csv`.
- Dérivez l'équation de régression qui prédit l'*Ouverture* sur la base des cinq autres variables d'intelligence.
- Que valent les paramètres du modèle ?
- Quelle est la valeur du coefficient de corrélation multiple au carré (coefficient de détermination) ?
- La régression multiple est-elle significative au seuil de 5% ?
- Quelles sont les variables qui jouent un rôle significatif au seuil de 5% dans notre modèle ?
- Comparez le modèle complet estimé au point c) à un deuxième modèle où vous supprimez la variable explicative **ARI** et à un troisième modèle où vous supprimez **INF**. Pour chaque test (il y en a deux), prenez comme seuil de signification $\alpha = 5\%$.
- Comparez les résultats obtenus en f) à ceux obtenus en g).
- Dans leur étude, Ashton, Lee et Vernon nous disent que l'*Ouverture* corrèle uniquement avec les facteurs d'intelligence cristallisée (à l'exception de l'arithmétique). Construisez donc le modèle proposé par les auteurs :

$$\text{OUV} \sim \text{INF} + \text{VOC} + \text{COM}$$
- Est-ce que ce nouveau modèle est emboîté dans le modèle estimé en c) ?
- Comparez les deux modèles à l'aide d'un test, que concluez-vous ?

Exercice 2

Utilisez l'ensemble de données `Mireault.csv` pour examiner la relation entre le niveau actuel de dépression et les autres variables. Un modèle plausible pourrait suggérer que la dépression (**Depres**) est fonction des éléments suivants : (1) la perception actuelle par l'individu de sa propre vulnérabilité vis-à-vis de toute perte additionnelle (**VulPer**), (2) le niveau de soutien social dont bénéficie la personne (**SupTot**) et (3) l'âge auquel la personne a perdu un de ses parents durant l'enfance (**AgePer**).

- Évaluez le modèle esquissé ci-dessus.

- b) Les données utilisées pour construire ce modèle satisfont-elles les conditions nécessaires à la mise en application d'une régression multiple ?
- Vérifiez que la moyenne des résidus est égale à 0 conditionnellement à $\widehat{\text{Depres}}$.
 - La variance des erreurs est-elle constante ?
 - Les résidus se distribuent-ils normalement ? Pour répondre à cette question, utilisez deux approches différentes.

Exercice 3

Lorsque les prédicteurs sont incommensurables, il est courant de les standardiser avant de construire un modèle de régression. La comparaison de l'impact des prédicteurs sur la variable dépendante est alors plus aisée.

- a) À partir des données du fichier `PrestigeCA.csv`, construisez le modèle qui permette de prédire le prestige d'une profession à partir du niveau d'instruction des personnes qui l'exercent, de leur revenu moyen et de la proportion de femmes qui pratiquent le métier.
- b) Construisez un modèle similaire au précédent mais en standardisant au préalable toutes les variables (prédicteurs et critère).
- c) Comparez les deux modèles.
- d) Le modèle qui permet d'exprimer la variable Y en fonction des variables non standardisées X_1 , X_2 et X_3 s'écrit de la manière suivante :

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \varepsilon$$

Celui qui permet d'exprimer la même réalité mais à l'aide des variables standardisées s'écrit :

$$Z_y = \beta_0 + \beta_1Z_1 + \beta_2Z_2 + \beta_3Z_3 + \varepsilon$$

Avec :

$$Z_y = \frac{Y - \bar{Y}}{s_y} \quad \text{et} \quad Z_j = \frac{X_j - \bar{X}_j}{s_j} \quad \text{pour } j = 1, 2, 3.$$

Comment calculer les coefficients β_0 , β_1 , β_2 et β_3 , à partir des coefficients B_0 , B_1 , B_2 et B_3 ? Vérifiez l'exactitude de vos formules sur les modèles construits aux deux premiers points de cet exercice.