

TP 3 - Analyse des données avec R

1. Définitions

1. Définir le principe de l'Analyse Factorielle des Correspondances (AFC)
2. Définir le principe de l'Analyse des Correspondances Multiples (ACM)
3. Expliquer les principales différences avec l'ACP (type de données, métrique utilisée,...)
4. Définir la(les) matrice(s) en entrée de l'ACM
5. Définir les outils d'aide à l'interprétation des résultats en sortie de l'ACM

2. Installation des packages R

Afficher dans RStudio le descriptif des packages suivants :

- *ggplot2*
- *FactoMineR*

Expliquer en quelques mots le contenu de chaque package puis les installer via la fonction *install.packages()*

3. Application de l'AFC

L'objectif est d'appliquer une analyse factorielle des correspondances au jeu de données *children* disponible dans la librairie *FactominerR*

1. Décrire le jeu de données
2. Analyser les statistiques descriptives
3. Représenter par des diagrammes en bâtons le profil ligne moyen et le profil colonne moyen à l'aide de la librairie **ggplot2**. Interpréter les résultats
4. Réaliser l'analyse factorielle des correspondances de ce jeu de données en considérant les modalités de la variable "*Education*" comme actives et les autres supplémentaires.
5. Représenter le Scree Plot des valeurs propres (en pourcentage cumulé et non)
6. Combien d'axes factoriels vous semblent intéressants à interpréter ?

7. Calculer l'inertie totale
8. Interpréter la qualité de représentation des résultats
9. Représenter séparément les lignes et les colonnes sur le premier plan factoriel.
10. Quels sont les catégories de personnes qui ont contribué le plus à l'axe 1 ?
11. Quelles sont les lignes (motifs) qui ont contribué le plus à l'axe 1 ?
12. Construire et interpréter le plan factoriel (1,2).

3. Application de l'ACM

Les données décrivent 66 clients ayant souscrit un crédit à la consommation dans un organisme de crédit.

1. Importer le jeu de données *credit.txt*
2. Analyser les statistiques descriptives du jeu de données
3. Afficher le tableau de Burt et le tableau disjonctif complet du jeu de données.
4. Représenter par des diagrammes en bâtons les modalités des variables qualitatives en utilisant la librairie *ggplot2*. Analyser les résultats (en particulier identifier les modalités sous ou sur-représentées)
5. Regrouper la modalité Side-car avec la modalité Moto.
6. Effectuer l'analyse des correspondances multiples en sélectionnant comme variables actives : Marche, Apport, Impayé, Assurance, Endettement. Les autres variables sont supplémentaires.
7. Construire le Scree Plot des valeurs propres. Conclure sur le nombre d'axes à conserver pour l'analyse.
8. Calculer l'inertie totale et retrouver la formule qui conduit à cette valeur.
9. Construire le plan (1,2) des individus, celui des variables actives, puis enfin celui des variables supplémentaires.
10. Interpréter les résultats (à l'aide des outils d'aide à l'interprétation)

4. Utilisation de Shiny

1. Reprendre le jeu de données Décathlon et relancer l'ACP à l'aide du package *FactoShiny*

2. Interpréter les résultats
3. Relancer les résultats de l'AFD et l'ACM en utilisant ***FactoShiny***