

TP 2

Analyse de données - Master 1
Analyse en Composantes Principales (ACP)

Packages et données

Ce TP s'appuie principalement sur le package `FactoMineR` pour l'ACP. Assurez-vous que les packages nécessaires sont installés et chargés.

Indication R : `library(FactoMineR)`

Partie 1 — Lancement d'une ACP avec FactoMineR

1. **Chargement des données** : Chargez le jeu de données `decathlon` et affichez ses premières lignes.

Indication R : `data(decathlon); head(decathlon)`

2. **Exécution de l'ACP** : Réalisez une ACP sur les 10 premières colonnes (performances) et stockez le résultat dans l'objet `res.pca`.

Indication R : `res.pca <- PCA(decathlon[,1:10], graph = FALSE)`

3. **Valeurs propres** : Affichez le tableau des valeurs propres de votre ACP.

Indication R : `print(res.pca$eig)`

Partie 2 — Interprétation d'une ACP sur les données decathlon

On continue avec l'objet `res.pca` de la partie 1.

4. **Cercle des corrélations** : Visualisez le cercle des corrélations pour les axes 1 et 2. Pour cela, on utilise la fonction `plot` appliquée à notre objet ACP, en spécifiant le graphique des variables.

Indication R : `plot(res.pca, choix = "var")`

5. **Analyse de l'axe 1** : D'après le graphique, quelles variables sont fortement corrélées (positivement ou négativement) à l'axe 1 ?

6. **Analyse de l'axe 2** : Identifiez maintenant les variables qui sont le plus fortement corrélées à l'axe 2.

7. **Interprétation des axes** : Proposez une interprétation simple pour l'axe 1 et l'axe 2.

8. **Graphe des individus** : Affichez le graphique des athlètes sur les deux premiers axes en utilisant la même fonction `plot`, mais en spécifiant le graphique des individus.

Indication R : `plot(res.pca, choix = "ind")`

9. **Analyse des individus** : Quels sont les athlètes qui se distinguent le plus sur l'axe 1 ? Et sur l'axe 2 ? Que pouvez-vous en déduire sur leurs performances ?

Partie 3 — Cas pratique : ACP sur des données financières

L'objectif est d'analyser la structure des rendements d'un portefeuille d'actions du CAC40 pour identifier les facteurs de risque communs.

12. **Téléchargement des données** : Chargez le package `quantmod` et téléchargez les données de 10 actions du CAC40 pour 2022-2023.

Indication R : `library(quantmod); tickers <- c("GLE.PA", ..., "MC.PA"); getSymbols(tickers, from="2022-01-01", to="2023-12-31")`

13. **Préparation et calcul des rendements** : En utilisant les prix ajustés, créez un unique dataframe `returns_df` contenant les rendements journaliers de toutes les actions.

Indications R : Utilisez `Ad()` pour extraire les prix, `do.call(merge, lapply(...))` ou une boucle `for` pour les assembler. Pensez à transformer les prix en log.

14. **Lancement de l'ACP** : Réalisez une ACP centrée-réduite sur votre tableau de rendements.

Indication R : `res.pca.fin <- PCA(returns_df, scale.unit = TRUE, graph = FALSE)`

15. **Analyse de la variance expliquée** : Visualisez l'éboulis des valeurs propres. Quel pourcentage de variance est expliqué par la CP1 ? Est-ce suffisant pour une analyse sur 2 axes ?

Indications R : L'information se trouve dans `res.pca.fin$eig`. Pensez à `barplot()`.

16. **Interprétation de l'axe 1 (Le Facteur Marché)** : Affichez le cercle des corrélations. Analysez la structure de l'axe 1 et donnez son interprétation financière.

Indications R : `plot(..., choix = "var")`

17. **Interprétation de l'axe 2 (Les Oppositions Sectorielles)** : Analysez l'axe 2. Identifie-t-il des oppositions entre groupes d'actions ? Quel est l'intérêt pour un investisseur ?

18. **Bonus : Visualisation du Facteur Marché** : Tracez l'évolution temporelle de la première composante principale.

Indications R : Les coordonnées sont dans `res.pca.finindcoord`. Utilisez `plot(..., type = 'l')`.