

TP 3

Analyse de données - Master 1 Analyse des Correspondances Multiples (ACM)

Objectif et Données

Nous analysons le jeu de données `tea` (consommation de thé) avec le package `FactoMineR`. L'objectif est de créer des profils types de consommateurs en utilisant `factoextra` pour des visualisations avancées.

Partie 1 — Préparation des données

1. **Chargement** : Chargez les packages `FactoMineR`, `factoextra` et chargez les données `tea`.
2. **Nettoyage des étiquettes (Pré-traitement)** : Par défaut, les graphiques afficheront uniquement le nom de la modalité (ex : "other", "lemon"), ce qui crée des confusions.
Écrivez un script (une boucle `for` ou des commandes manuelles) pour modifier les niveaux des facteurs : ajoutez le nom de la variable comme préfixe devant chaque modalité.
Objectif : Transformer l'étiquette "lemon" en "how_lemon" ou "green" en "Tea_green".
Indication : Utilisez les fonctions `levels()` et `paste()`.
3. **Sélection des variables** : Pour une analyse plus pertinente, nous allons travailler uniquement les colonnes suivantes (mélange de variables de comportement et de profil) :
`Tea, How, how, sugar, price, where, sex, age_Q, SPC, tea.time`
Indication : `new_tea <- tea[, c("Tea", ...)]`

Partie 2 — Calcul et Visualisation

4. **Lancement de l'ACM** : Réalisez l'ACM sur votre tableau réduit `ma_selection`.
Indication : similaire à l'ACP avec la fonction `MCA`
5. **Variance expliquée** : Affichez l'éboulis des valeurs propres pour juger de l'information capturée par les axes.
Indication : `fviz_eig` ou `barplot`
6. **Nuage des modalités** : Affichez le graphique des variables. Grâce à l'étape 2, les étiquettes devraient être explicites (ex : `Tea_green`).
Indication : `fviz_mca_var`
7. **Biplot optimisé** : Nous allons superposer les individus et les variables, mais en rendant les individus discrets pour garder le graphique lisible. Utilisez la fonction `fviz_mca_biplot` avec les paramètres suivants :
 - Points individus (`geom.ind`) en "point" et couleur gris clair (`col.ind = "gray90"`).

- Variables en bleu (`col.var = "blue"`).
- Afficher uniquement le texte des variables (`label = "var"`).

Partie 3 — Interprétation

8. **Analyse de l'Axe 1 :** Observez le nuage des modalités. Quelles habitudes de consommation s'opposent sur l'axe horizontal ?
9. **Analyse des individus (Biplot) :** Sur le Biplot généré à la question 7 :
10. Observez et analysez les différents groupes d'individus.