

Analyse de données – TP n°3 Analyse en Composantes Principales (ACP)

Il s'agit d'analyser le résultat d'une étude de dépenses alimentaires annuelles (en francs) de $n = 12$ ménages français. Les données sont dans le fichier `menages.txt`, et les variables sont :

Menage : l'identificateurs des individus, sous forme de :

- la caractéristique professionnelle du chef de famille :
 - MA travailleur manuel
 - EM employé non manuel
 - CA cadre
- le nombre d'enfants du ménage, entier à valeur dans $\{2, 3, 4, 5\}$.
- Les postes de 7 dépenses alimentaires dont les noms sont explicites : Pain, Légumes, Fruits, Viandes, Volailles, Lait et Vin.

Note : Les sommes sont peu parlantes car les données datent de 1972, mais on peut les considérer comme exprimées en une autre unité que le franc (par exemple l'euro, ou des multiples de l'euro).

**L'analyse de ces données consiste à répondre à la question :
Quels sont les liens entre les habitudes de consommation des
ménages et leur structure ?**

Les questions reprennent le schéma des étapes vues en cours afin d'effectuer l'ACP de ces données sous R avec une fonction du package `ade4`.

- 1) Chargez la table sous R, et construisez à partir de la variable **Menage** deux facteurs qualitatifs **CSP** et **NE**, cf descriptif ci-dessus.
- 2) Etude préliminaire de la table : statut des variables, recherche d'éventuelles valeurs aberrantes etc. Quelles méthodes de statistique descriptive pouvez-vous appliquer ? Quelles conclusions rapides pouvez-vous tirer de ces premières analyses ?
- 3) Choix de la métrique appropriée sur les individus : comment décidez-vous ?
- 4) Regardez `?dudi.pca` du package `ade4`. Appliquez cette fonction aux données ménages. Etudiez la nature de l'objet R que retourne cette fonction, et les liens avec le cours sur l'ACP.
- 5) Regardez l'aide de la fonction `?inertia.dudi`, affichez les pourcentages d'inerties sur les axes principaux, ainsi que le diagramme des valeurs propres. Commentez.
- 6) Calculez avec R les corrélations des composantes principales avec les variables d'origine (voir `?cor`). Affichez le cercle de corrélations à l'aide de la fonction `s.corcircle`, et effectuez-en la synthèse.

- 7) Représentation des plans factoriels principaux (cartes des individus), avec les facteurs qualitatifs pertinents en variables supplémentaires ou par coloration, et avec les projetés des barycentres par classes (cf. TP n°2 et usage de `s.class`).
- 8) Affichez les \cos^2 cumulés sur le plan principal, et représentez-les sur les plans factoriels.
- 9) Listez les 5 individus les moins bien représentés sur le plan principal.
- 10) Interpréter les résultats de l'analyse.