

Analyse en composantes principales

Benjamin Arras

On s'intéresse dans ce TP à des données issues de l'observatoire de l'OCDE (Organisation de Coopération et de Développement Economiques). Pour chaque pays membre, on a relevé un certain nombre de mesures :

- NATA : taux brut de natalité,
- CHOM : taux de chômage,
- APRI : pourcentage d'actifs dans le secteur primaire
- ASEC : pourcentage d'actifs dans le secteur secondaire,
- PIB : produit intérieur brut (par habitant),
- FBCF : formation brute de capital fixe (par habitant),
- INFL : hausse des prix,
- RECC : recettes courantes (par habitant),
- MINF : mortalité infantile,
- PROT : consommation de protéines animales (par habitant),
- NRJ : consommation d'énergie (par habitant).

1. Importer la base de données `ocde.txt` et effectuer une analyse descriptive de la base de données.
2. Charger les packages **FactoMineR** et **factoextra**, et faire une analyse en composantes principales de la base de données en utilisant la fonction `PCA` du package **FactoMineR**, dans un premier temps sans réduire les données. Commenter.
3. Quelle est la part de variance expliquée par le premier axe principal ? En étudiant le graphe des variables selon les deux premières composantes, comment peut-on interpréter le premier axe et expliquer le résultat ?
4. 'A la vue des résultats précédents, on se propose de relancer l'ACP, cette fois en réduisant les données. Comparer les résultats obtenus à l'aide des fonctions `PCA` (package **FactoMineR**, `prcomp` (package **stats**) et `dudi.pca` (package **ade4**). Pour chaque fonction, indiquer les commandes permettant d'afficher :
 - les valeurs propres
 - les vecteurs propres
 - la part d'inertie expliquée par chaque axe
 - les coordonnées des individus dans la nouvelle base
 - la contribution des individus et des variables à chaque axe
5. Proposer un nombre d'axes à retenir, en justifiant ce choix.
6. Étudier la qualité de représentation et les contributions aux axes, des variables et des individus.