

Statistiques avec SAS : Sujet du TD N°7

Jean-Sébastien Roy (js@jeannot.org)

1 Classification hiérarchique

La table `commlimi` contient des informations issues du recensement des communes limitrophes de la ville de Paris. Normaliser la table `commlimi` en divisant pour chaque commune toutes les variables par la population. Appeler `commlimi_norm` la table ainsi créée.

Effectuer une classification ascendante hiérarchique avec la distance de Ward sur les variables `dip12-dip19`, réduites.

Tracer le dendrogramme et la variation de la distance entre les classes jointes, au cours de la classification.

Garder trois classes et ajouter cette information de classification à la table initiale `commlimi_norm`.

Placer les communes sur un plan `cs5*cs4`, et les colorier en fonction de la classe. Que remarque-t-on ?

2 Analyse en composantes principales

Effectuer une ACP sur la table `commlimi_norm`, pour les variables `cs2-cs7 cs81 cs84 cs85 cs86 cs23 jeunes adultes ages franc euro maghreb afriq world iber`.

Tracer la part d'inertie expliquée par chaque axe en valeur, en proportion, et cumulée.

Calculer les contributions des communes aux axes et remarquer que la commune 92051 perturbe l'analyse du deuxième axe. La supprimer et recommencer l'ACP. De même avec la commune 92064.

Garder trois axes, et placer les communes, coloriées à l'aide de la classification, sur les plans `axe1*axe2` et `axe2*axe3`.

3 Analyse des correspondances multiples

Créer une table à partir du fichier `chiens.txt` (du à Tenenhaus). Créer des formats adaptés au tableau.

Effectuer une ACM sur le tableau avec `fon` comme variable supplémentaire.

Placer les chiens et les variables sur le plan formé des deux premiers axes. Faire de même en colorant les chiens à l'aide de leur fonction.