Analyse de données
M1 Statistique et économétrie - 2013
V. Monbet

Analyse factorielle des correspondances

A travers ce TD, nous allons apprendre à mettre en oeuvre l'analyse factorielle des correspondances. Le TD commence par une application simple sous SAS. La seconde application est traitée sous R. Il s'agit d'analyser les votes au premier tour des élections présidentielles. Cet exemple permettra d'aider à bien comprendre l'AFC notamment en la comparant à l'ACP.

1 Questions de cours

1. A quel type de questions permet de répondre l'analyse factorielle des correspondances ? Donner un exemple.
2. Peut-on l'utiliser pour analyser des données qui sont des réalisations de variables continues ?
3. Combien de variables peut-on considérer ? Quel est la dimension maximum de l'espace factoriel ?
4. Rappeler la définition du tableau de contingence ainsi que celle des profils moyens.
5. Montrer, en utilisant le cours que chercher les valeurs singulières de la matrice des résidus de Pearson est équivalent à faire l'analyse spectrale de la matrice $A^{-1}NB^{-1}N^T$ (avec les notations du cours).

2 Parfums

On dispose d'un tableau de contingence contenant 12 parfums décrits par 39 mots. Une valeur xij correspond au nombre de fois où le descripteur j a été utilisé pour décrire le parfum i. Nous voulons savoir quels sont les parfums qui ont le même profil de mots, quels sont les mots qui se ressemblent c'est à dire qui sont associés de la même façon aux mêmes parfums.

1. Importer le jeu de données "parfums.tex".
2. Utiliser la fonction `balloonplot` du package `gplots` pour représenter la table de contingence.

3. Utiliser la fonction `mosaicplot` pour visualiser les données et l'équivalent sous l'hypothèse d'indépendance. Que pensez-vous que les deux variables soient dépendantes ?
4. Réaliser un test du chi^2 pour tester l'indépendance des deux variables et confirmer (ou infirmer) votre première impression. Que concluez-vous ?
5. La fonction `chisq.test` permet d'obtenir la résidus de Pearson. Représentez les à l'aide de la fonction `corrplot` disponible sur la page web du cours.
6. Analyser le tableau de données à l'aide d'une AFC. Par exemple,

```
library(FactoMineR)
res.ca = CA(perfume,col.sup=14:39)
plot(res.ca,invisible="row")
plot(res.ca,invisible=c("col","col.sup"))
```

7. Représenter les valeurs propres en utilisant des diagrammes en bâton. Par combien d'axes l'information est-elle représentée de manière satisfaisante ?
8. Interpréter le premier puis le second axe factoriel.
9. Interpréter globalement le premier plan factoriel. Quelles sont les 2 variables qui contribuent le plus ce plan ? Peut-on retrouver ce résultat sur le graphique ? Si vous regardez les données brutes, ceci vous parait-il logique ?
10. Interpréter la proximité entre *J'adore* eau de parfum et *J'adore eau de toilette*.
11. Comment caractériser le parfum Lolita Lempika ? Quel est l'adjectif qui lui correspond le mieux ? Interpréter finement les positions des modalités "Lolita Lempika" de la variable "parfums" et "sugary" et "vanilla" de la variable "descripteurs".

2.1 Comparaison avec l'analyse en composantes principales

Il peut être intéressant de comparer les résultats de l'AFC avec ceux de l'ACP afin de mettre en évidence le rôle joué par la distance du χ^2.

1. Quelles données proposez-vous d'utiliser pour l'ACP ? Pourquoi ?
2. Réaliser l'ACP en considérant les parfums comme individus et en utilisant des données normalisées par le nombre total de vote par candidat.
3. Commenter les graphiques en comparaison avec ceux de l'AFC.