

1. AFC sous R (package « ade4 »)

Tutoriels de référence :

<http://tutoriels-data-mining.blogspot.fr/2008/03/afc-association-medias-et-professions.html>

<http://tutoriels-data-mining.blogspot.fr/2009/05/analyse-factorielle-des-correspondances.html>

Sans oublier les sites indiqués dans le document principal.

Questions (à rendre un fichier Markdown exporté en .docx éditable incluant : le numéro des questions, le code, les résultats et les commentaires) :

Attention, vous devez utiliser exclusivement le package « ade4 » pour la partie AFC dans cet exercice.

- A. Importez les données du fichier « media_prof_afc.xls »
 1. C'est un tableau de contingence croisant la profession avec les médias consultés par 12388 « contacts médias » (Lebart et al., page 103).
 2. La première ligne correspond aux étiquettes de colonnes.
 3. La première colonne correspond aux étiquettes de lignes. Attention : il faudra modifier en conséquence votre data frame pour que les noms des lignes correspondent bien à des étiquettes.
- B. Calculez et affichez les profils lignes du tableau de contingence. Pour chaque profil, affichez la modalité colonne qui est la plus fréquente. Que peut-on dire en observant les résultats ?
- C. Calculez la distance à l'origine de chaque profil ligne. Quelle est la modalité ligne qui s'écarte le plus du profil moyen ?
- D. Calculez l'inertie de chaque modalité ligne. Effectuez la somme des inerties et notez la valeur obtenue.
- E. Reproduisez les trois questions précédentes en travaillant sur les profils colonnes. Commentez les résultats.
- F. Effectuez le test d'indépendance du KHI-2. Le tableau observé s'éloigne significativement de la situation d'indépendance à 5% ? (**chisq.test**, sans correction de continuité)
- G. Calculez les contributions au Khi-2 dans le tableau de contingence. Quelles sont les couples de modalités les plus contributives ? (cf. **les propriétés fournies par chisq.test**)
- H. Calculez les indices d'attraction et de répulsion. Quelles sont les associations (positives ou négatives) les plus remarquables ? (cf. **les propriétés fournies par chisq.test**)
- I. Lancez l'analyse factorielle des correspondances avec la fonction dédiée du package « ade4 » (**dudi.coa**). Affichez les pourcentages d'inerties (individuels et cumulés) associés aux axes factoriels. Affichez le graphique des éboulis. Combien d'axes retiendriez-vous finalement ?
- J. Affichez les coordonnées, cosinus carrés et contributions des modalités lignes. Quelles sont les trois modalités qui contribuent le plus au premier facteur ? Au second facteur ? Quelles sont les

modalités les mieux représentés sur le premier facteur ? Sur le second facteur ? Sur les deux premiers facteurs ?

- K. Affichez les modalités lignes dans le premier plan factoriel. Quels commentaires pouvez-vous faire à la vue du graphique ? Est-ce que ces résultats corroborent ce que nous avons pu voir plus haut (ex. distance à l'origine et inertie des modalités lignes)
- L. Répéter les deux questions précédentes pour les modalités colonnes.
- M. Affichez le graphique conjoint des modalités lignes et colonnes. Que peut-on dire au sujet des attractions et répulsions à la vue des positions relatives des modalités dans le plan factoriel ? Rapprochez ces résultats avec ce que nous avons pu voir concernant l'étude des indices et contributions au Khi-2 plus haut.

2. AFC sous Python (package « fanalysis »)

Refaire l'analyse à l'identique (contentez-vous de reproduire les calculs, il n'est pas nécessaire de commenter les résultats) en utilisant le package « fanalysis » (*je crois que certaines parties seront très faciles à réaliser, tout comme pour l'ACM, l'idée est de se familiariser avec la pratique de l'AFC sous Python*).

https://github.com/OlivierGarciaDev/fanalysis/blob/master/doc/ca_tutorial.ipynb

A rendre : un fichier PDF imprimé de Jupyter Notebook, incluant les numéros des questions, le code, les résultats et les éventuels commentaires.