

1. ACM sous R (package « ade4 »)

Tutoriels de référence :

<http://tutoriels-data-mining.blogspot.fr/2012/12/analyse-des-correspondances-multiples.html>

<http://tutoriels-data-mining.blogspot.fr/2009/05/analyse-de-corresponances-multiples.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 ACM dans cet exercice.

- A. Importez les données du fichier « **races_canines_acm.xls** »
 1. La première ligne correspond aux noms des variables.
 2. La première colonne correspond aux noms des individus (des canidés). Attention : il faudra modifier en conséquence votre data frame pour que les noms des lignes correspondent bien à l'identifiant des canidés.
 3. Les variables (taille, ..., agressivité) correspondent aux variables actives, elles doivent être utilisées pour la construction des axes factoriels.
 4. La variable « fonction » est illustrative.
- B. Isolez dans un dataframe (**D1**) spécifique les variables actives.
- C. Pour chaque variable active, affichez son diagramme de fréquences (graphique).
- D. Lancez l'ACM avec l'outil dédié de la librairie « ade4 ».
- E. Affichez les valeurs propres issues de l'ACM. Si l'on applique la règle de Kaiser (on retient les facteurs pour lesquels les valeurs propres sont supérieures à la moyenne), combien de facteurs devons-nous retenir ?
- F. Affichez l'éboulis des valeurs propres. Combien de facteurs suggérez-vous de retenir à la vue du graphique ?
- G. Calculez le pourcentage d'inertie associé à chaque facteur, puis faites le cumul. Quelle est la part d'information véhiculée par les deux premiers facteurs ?
- H. Appliquez la correction de Benzécri. Elaborez de nouveau le graphique « éboulis des valeurs propres » après correction. Combien de facteurs devrions-nous retenir ?
- I. Finalement, nous nous en tenons à 2 facteurs.
- J. Affichez les modalités des variables dans le premier plan factoriel (graphique). Quelles proximités vous paraissent les plus évidentes ?
- K. Affichez les valeurs des coordonnées, cosinus carrés et contributions des modalités.
- L. Quelles sont les 5 modalités les mieux représentés sur le premier facteur, sur le second facteur, sur les 2 premiers facteurs ?

- M. Quelles sont les 5 modalités qui contribuent le plus au premier facteur ? Au second facteur ?
- N. Affichez les étiquettes des individus dans le premier plan factoriel (graphique). Quelles proximités vous paraissent les plus évidentes ?
- O. Affichez les valeurs des coordonnées, cosinus carrés et contributions des individus.
- P. Quels sont les 5 individus les mieux représentés sur le premier facteur ? Sur le second facteur ? Sur les 2 premiers facteurs ?
- Q. Quels sont les 5 individus qui contribuent le plus au premier facteur ? Au second facteur ?
- R. Etude de la variable supplémentaire « fonction »
 1. Dans le plan factoriel des individus, coloriez les étiquettes des canidés selon les modalités de la variable supplémentaire.
 2. Calculez les coordonnées des modalités de la variable supplémentaire sur les 2 premiers axes factoriels.
 3. Placez ces modalités dans la carte factorielle des modalités. Que pouvons-nous dire ?
 4. Sur lequel des deux facteurs les modalités de la variable illustrative se distinguent-elles le mieux ? Pouvez-vous confirmer cela par le calcul ? En vous appuyant sur le rapport de corrélation par exemple.
 5. Calculez les valeurs tests des modalités de la variable illustrative sur le premier facteur. Quelles sont le ou les modalités qui se démarquent le plus ?
- S. Croisez la variable supplémentaire « fonction » avec chacune des variables actives de l'analyse. Calculez le KHI-2 du tableau de contingence. Triez les résultats par ordre de p-value croissante. Quelle est la variable la plus liée avec « fonction » ? Ce résultat confirme-t-il ce que nous avons constaté avec l'ACM ?

2. ACM 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 questions seront très faciles à réaliser, sauf lorsqu'il faudra introduire des calculs et tris additionnels...*).

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

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