

1. ACP sous R (princomp)

Tutoriel de référence :

<http://tutoriels-data-mining.blogspot.fr/2009/05/analyse-en-composantes-principales-avec.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 ne devez pas utiliser de package spécialisé dans cette partie. Vous passerez par la fonction princomp() du package standard « stats ».

- A. Importez les données du fichier AUTOS2005.TXT ([read.table](#))
 1. La première ligne correspond aux noms des variables ;
 2. La première colonne correspond aux noms des individus (des véhicules) ;
 3. Les 4 dernières variables (prix, origine, carburant, type4X4) correspondent aux variables illustratives, elles ne doivent pas être utilisés pour la construction des axes factoriels ;
 4. Les 5 derniers véhicules (P607, VELSATIS, BMW_530, MERC_E, AUDIA8) correspondent aux individus illustratifs, ils ne doivent pas être utilisés pour la construction des axes factoriels.
- B. Isolez dans un data.frame (**D1**) spécifique les variables et individus actifs.
- C. Construire les graphiques nuages de points sur les variables prises 2 à 2 sur D1 ([pairs](#)).
- D. Lancez l'ACP sur **D1** à l'aide de [princomp\(\)](#). Attention, vous devez utiliser l'ACP normée (cf. les options de princomp). Affichez les résultats. Affichez les attributs de l'objet issu de [princomp\(\)](#).
- E. Détection du nombre adéquat de composantes à retenir :
 1. Calculez et affichez les valeurs propres associées aux axes. Si on applique la règle du Kaiser, combien de facteurs conserverions-nous ?
 2. Affichez le graphique « scree plot » (éboulis des valeurs propres). Combien de facteurs conserveriez-vous pour à la vue de ce graphique ?
 3. Traduisez-les valeurs propres en pourcentages d'inertie. Quel est le pourcentage d'inertie porté par le premier axe, par le second, par les 2 premiers ?
 4. Construisez le graphique des pourcentages d'inerties cumulées ([cumsum peut être ?](#)). Combien de facteur sélectionnerons-nous à la vue de ce graphique ?
 5. Calculer les seuils du « test des bâtons brisés » (cf. Google). Combien de composantes sélectionnerions-nous dans ce cas ?
- F. Finalement, nous nous en tenons à 2 composantes principales. Calculer les corrélations des variables actives avec les axes. Affichez le cercle des corrélations. Que constatez-vous ?
- G. Quelles sont les 2 variables qui sont le mieux représentées sur la première composante ? Sur la seconde composante ? Sur le cumul des 2 premières composantes ? (**cf. du côté des cos²**)
- H. Quelles sont les 2 variables qui contribuent le plus à la première composante ? A la seconde composante ? (**cf. les contributions**)

- I. Quelle interprétation pouvons-nous dès lors faire à l'issue de l'analyse du rôle des variables ?
- J. Projetez les individus dans le premier plan factoriel. Utilisez les noms des véhicules en guise d'étiquettes dans le graphique. Que distinguez-vous ?
- K. Quelles sont les 5 véhicules les mieux représentées dans le plan factoriel ? (les 2 premières composantes prises en bloc) (**cos²**)
- L. Quelles sont les 5 véhicules qui contribuent le plus à la première composante ? A la seconde composante ? (**contributions**)
- M. Calculez la corrélation de la variable quantitative supplémentaire « prix » avec les 2 premières composantes. Quel commentaire peut-on faire ? (**cor**). Placez la variable dans le cercle des corrélations.
- N. Calculez les moyennes des 2 composantes principales conditionnellement aux modalités des variables qualitatives supplémentaires « origine », « carburant » et « type4X4 » (**tapply**). Placez les modalités dans le plan factoriel.
- O. Projetez les individus dans le plan factoriel sans y faire figurer les étiquettes, coloriez les points selon les modalités de la variable « Type4X4 ». Que constatez-vous ?
- P. Quelles remarques pouvons-nous faire à l'issue de cette analyse des variables supplémentaires ?
- Q. Calculez les coordonnées factorielles des 5 véhicules supplémentaires, placez-les dans le plan factoriel. Que constatez-vous ?

2. ACP sous Python (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 ».

<https://github.com/OlivierGarciaDev/fanalysis>

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