

Analyses des données :

Analyse en composantes principales (Compléments)

L'Analyse en Composantes Principales est une technique de statistique descriptive multidimensionnelle qui cherche de à restituer l'information contenue dans une Base de données comprenant **n** individus et **p** variables quantitatives.

Critères pour un ACP dite 'intéressante' :

Afin d'obtenir une ACP intéressante il faut s'assurer de la vérification de certains critères :

Un très bon coefficient KMO (kaiser Meyer Olkin) :

Un coefficient de (kaiser Meyer Olkin) KMO proche de 1 c'est-à-dire que les corrélations partielles sont faibles

Il indique jusqu'à quel point l'ensemble de variables retenu est un ensemble cohérent et permet de constituer une ou des mesures adéquates de concepts. Un KMO élevé indique qu'il existe une solution factorielle statistiquement acceptable qui représente les relations entre les variables. Une valeur de KMO de moins de 0.5 est inacceptable.

Formule :

$$KMO = \frac{\sum_i \sum_j r_{ij}^2}{\sum_i \sum_j r_{ij}^2 + \sum_i \sum_j a_{ij}^2}$$

r_{ij} : Les corrélations Totales et a_{ij} : Les corrélations Partielles issues de la matrice Anti – image

Car L'ACP souhaite que les corrélations soient expliquées par d'autres variables que celles concernées il ne seait pas intéressant d'étudier des variables uniquement corrélées deux a deux .

Un test de sphéricité de Barlett ayant une significativité proche de zéro:

$$\begin{cases} H_0: R = I \\ H_1: R \neq I \end{cases} \quad I: \text{matrice identité}$$

Statistique de décision vaut :

$$T = - \left(n - 1 - \frac{2p+5}{6} \right) \ln(|\det(R)|) \hookrightarrow \text{de } \chi^2 \left(\frac{p(p+1)}{2} \right) \text{ sous } H_0$$

Ce test permet de savoir si variables sont corrélées entre elles.

On souligne la sphéricité du nuage indique que ce dernier se dilate dans tout les sens.

❖ Application sur R (Suite de l'ACP sur les voitures) :

```

> kmos(X)
Error in kmos(X) : could not find function "kmos"
> KMOS(X)

Kaiser-Meyer-Olkin Statistics

Call: KMOS(x = X)

Measures of Sampling Adequacy (MSA):
  cylindr  puissanc  vitesse    poids  largeur  longueur
0.7512436 0.5881367 0.6236736 0.7271066 0.6167531 0.5421397

KMO-Criterion: 0.6363657

> bart_spher(X)

Bartlett's Test of Sphericity

Call: bart_spher(x = X)

      X2 = 167.607
      df = 15
p-value < 2.22e-16

```

La p-valeur du test de Barlett est inf rieur a 5%
ET KMO vaut 0,63 m diocre mais acceptable.

Interpr tations :

Le but ultime de l'ACP est d'aboutir   des r sultats qui pr sente une aisance   l'interpr tation d'o  la difficult  des m thodes factorielles n anmoins, il arrive souvent qu'il y'a de nombreuses variables avec corr lations moyennes sur plusieurs axes factoriels ce qui rend l'interpr tations de ces axes laborieuse.

A cet effet il existe des m thodes dite de rotations qui rendent les valeurs de ces corr lations plus tranch es. On distingue quatre m thodes de rotations :

- **La m thode VARIMAX** : s'applique lorsque la plupart des variables sont repr sent s sur un seul axe elle minimise le nombre de variables qui ont des corr lations importantes avec un facteur.Cette rotation donc est destin e   maximiser les variances des poids factoriels bruts au carr  pour chaque facteur ; cela revient   maximiser les variances dans les colonnes de la matrice des poids factoriels bruts au carr .
- **La m thode QUARTIMAX** : s'applique lorsqu'une variable est fortement corr l e   plusieurs axes   la fois c'est une m thode de rotation qui minimise le nombre de facteurs requis pour expliquer une variable.
- **La m thode EQUAMAX** : est une combinaison de deux m thodes pr c dentes il s'agit d'une m thode de rotation qui minimise   la fois le nombre de variables qui p se fortement sur un

facteur est le nombre de facteurs requis pour expliquer une variable.

- **La méthode OBLIMIN** : permet d'effectuer des rotations oblique sur les axes factoriels elle permet de mettre en évidence des phénomènes qui déterminent les directions d'allongement orthogonal des nuages des points.

Variables Supplémentaires :

Dans la même optique de rendre aisée une lecture une autre méthode s'ajoute qui consiste à ajouter des variables (non pas dans la base de données) mais vers la fin de l'étude, On définit souvent comme variables statistiques supplémentaires les centres de gravité de groupes formés à priori, définis par les moyennes des variables de ces groupes. Par ailleurs, Ce ne sont pas des variables ayant un rapport direct avec l'analyse mais que l'on souhaite voir représentées dans les graphiques notamment le cercle des corrélations.