# Analyse des données multidimensionnelles

## Objectifs

Ce cours aborde les trois/quatre méthodes les plus courantes de la statistique descriptive multidimensionnelle : l’Analyse en Composantes Principales (ACP), l’Analyse Factoriel des correspondances (AFC), l’Analyse des Correspondances Multiples (ACM) et l’Analyse Factorielle des Données Mixtes (AFMD)

## Volume horaire

CM : 12 h TD : TP : 12h

Les Tps sous R : Packages MExposition, FactMineR et/ou ADE4.

Références :

* http://wikistat.fr/
* Des jeux de données et exemple de TP sont disponibles à  : <http://factominer.free.fr/factomethods/> et http://wikistat.fr/

## Programme (voir aussi programme détaillé ci-dessous)

Chapitre 1 : Analyse en Composantes Prinicipales (ACP)

1. Introduction

2. Problématique et données

3. Espaces vectoriels / espace euclidien

4. Modèle : Observation =Modèle+Bruit

5. Estimation

6. Graphiques

7. Choix des axes

8. Individus et variables supplémentaires

9. Un exemple d’étude. Interpréter les sorties du TP : http://factominer.free.fr/factomethods/analyse-en-composantes-principales.html

Chapitre 2 : Analyse Factorielle des Correspondances (AFC)

1. Introduction

2. Problématique et données

3. Les profils lignes et colonnes

4. AFC peut être vue comme une double ACP

5. Interprétation

6. Un exemple d’étude.

Chapitre 3 : Analyse des Correspondances Multiples (ACM)

1. Introduction

2. Problématique et données, codage des variables qualitatives

3. AFC du tableau disjonctif complet : cas p=2 et cas p>2

4. AFC du tableau de Burt : cas p=2 et cas p>2

5. Etude de cas.

Chapitre 4 : Analyse Factorielle des données mixtes (AFDM)

1. Introduction

2. Problématique et données

3. AFDM à l’aide d’une ACP

4. AFDM à l’aide d’une ACM

5. Etude de cas

Chapitre 5 : Tableau multiples et Données Evolutives

1. Introduction

2. Problématique et données

3. STATIS (Structuration de Tableaux A Trois Indice de la Statistique)

4. STATIS DUALE

5. Etude de cas

## Programme détaillé

Chapitre 1 : Analyse en Composantes Principales (ACP)

1. Introduction

* Objectif : Etude de données comportant **p** variables observées sur **n** individus
* Résultats attendus :
  + Visualisation des corrélations entre les variables et interprétation des sorties graphiques
  + Visualisation de la répartition des individus et interprétation des sorties graphiques

2. Problématique et données

* Représentation de **n** individus dans un sous-espace vectoriel de R^p.
* Exemple illustratif pour l'A.C.P.

3. Espaces vectoriels / espace euclidien

* Choix d’un métrique (M) / Métrique des poids (D)
* Lien entre : covariance et produit scalaire, écart-type et norme, corrélation et cos
* Projection

4. Cadre théorique et modèle (Observation =Modèle+Bruit)

5. Estimation

* Valeurs propres, Vecteurs propres
* Vecteurs principaux, Axes principaux, Composantes principales.
* Cas particulier (usuel) où M et D sont des matrices identités.
* ACP non-normé, ACP normé
* Inertie totale
* Contribution des axes à l’inertie totale

6. Graphiques

* Les individus : projection/qualités de représentation/contribution
* Les variables : projection/qualités de représentation/corrélation, cerle de corrélation
* Représentation Biplot

7. Choix des axes

* Critères les plus courants : Part d’inertie/ Règle de Kaiser/règles de Cattell/éboulis des valeurs propres/diagrammes en boîtes
* Autres critères : critère utilisant des tests d’hypothèses.
* Interprétation des axes en fonctions des anciennes variables

8. Individus et variables supplémentaires

9. Un exemple d’étude de cas. Interpréter les sorties du TP : <http://factominer.free.fr/factomethods/analyse-en-composantes-principales.html>

10. Les limites d’utilisation de l’ACP

Chapitre 2 : Analyse Factorielle des Correspondances (AFC)

1. Introduction

* Objectif : Etude des correspondances entre les modalités de deux variables qualitatives : X et Y respectivement à q et p modalités.
* Résultats :
  + Visualisation des modalités de X et Y séparément
  + Visualisation des correspondances des modalités de X et de Y.

2. Problématique et données

* Le tableau de contingence
* Les profils lignes et colonnes
* Mesure de l’indépendance : métrique de chi2 et le test du chi2 d’indépendance.
* Objectif de l’AFC

4. AFC peut être vue comme une double ACP

* ACP sur les profils lignes.
  + Le tableau/ la métrique (ML)/la matrice de poids (DL)
  + Les éléments propres
* ACP sur les profils colonnes
  + Le tableau/ la métrique (ML)/la matrice de poids (DL)
  + Les éléments propres
* Dualité

5. Interprétation

6. Un exemple d’étude.

* Voir par exemple Wikistat ou http://factominer.free.fr/factomethods/analyse-factorielle-des-correspondances.html

Chapitre 3 : Analyse des Correspondances Multiples (ACM)

1. Introduction

* Objectif : Etude des correspondances entre les modalités de p>2 variables qualitatives

2. Problématique et données, codage des variables qualitatives

* Tableau disjonctif complet
* Tableau de Burt

3. AFC du tableau disjonctif complet : cas p=2 et cas p>2

4. AFC du tableau de Burt : cas p=2 et cas p>2

5. Etude de cas.

* Voir par exemple Wikistat ou http://factominer.free.fr/factomethods/analyse-des-correspondances-multiples.html

Chapitre 4 : Analyse Factorielle des données mixtes (AFDM)

1. Introduction
   * Mettre en œuvre une analyse factorielle dans laquelle des variables de types quantitatives et des variables types qualitatives jouent un rôle actif
2. Problématique et données, codage des variables
   * Coder les variables quantitatives en qualitatives ou les indicatrices en quantitatives
3. AFDM à l’aide d’une ACP
4. AFDM à l’aide d’une ACM
5. Etude de cas
   * Voir https://eric.univ-lyon2.fr/~ricco/cours/slides/AFDM.pdf

Chapitre 5 : Tableau multiples et Données Evolutives ?

1. Introduction
   * Objectif : Etude de la stabilité des relations entre les individus/variables dans les tableaux
2. Problématique des données structurées en blocs de variables
3. STATIS

ACP particulière sur la synthèse des individus

* + Inter-structure : Définition d’objets représentatifs des tableaux.
  + Compromis : Définition d’une matrice compromis.
  + Intra-structure : Etude des différences entre les tableaux
  + Evolution des individus ou variables suivant les tableaux.

1. STATIS DUALE

Même démarche que STATIS en prenant comme matrice un tableau Vt la matrice des covariances des variables à la place des Wt

1. Etude de cas
   * https://www.utdallas.edu/~herve/Abdi-Statis2007-pretty.pdf