# CHAP. 1 : INTRODUCTION

Trois types d'analyse factorielle :
- Analyse en composantes principales (ACP)
- Analyse factorielle des correspondances (AFC)
- Analyse des composantes multiples (ACM)

## I. TYPES DE DONNEES

### 1. Tableau de données

- Lignes : correspondent aux ind. stat. (sf AFC)
- Colonnes : correspondent : - aux variables stat.
  - aux modalités de var. stat.

### 2. Variables

- Variables : - En ACP → variables quantitatives de m importance.
  - En AFC → 2 var. sur une population
  - En ACM → des var. sur une population

## II. CALCUL MATRICIEL & ANALYSE DES DONNEES

### 1. Matrices symétriques

Prop : Soit M, une mat. quelconque.
Alors ${}^tM \times M$ et $M \times {}^tM$ st symétriques.

### 2. Distance

U et V, deux vect. colonnes ds une base orthon.

- Produit scalaire : ${}^tU \times V$
- Distance euclidienne : $\|U\| = \sqrt{{}^tU \times U}$

## 3. Diagonalisation

**Prop:** Tte mat. symétrique est diagonalisable dans $\mathbb{R}$.

# CHAP. 2 : PARAMETRES D'UN NUAGE DE POINTS APPROCHE STATISTIQUE

## I. PARAMETRES STATISTIQUES

$n$ : effectif total  $X$ : variable  $x$ : modalité de $X$

- **Moyenne :** $\boxed{\overline{X} = E(X) = \frac{1}{n}\sum x}$
- **Variance :** $\boxed{V(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2 = \frac{1}{n}\sum (x - \overline{x})^2}$
- **Ecart-type :** $\boxed{\sigma_x = \sqrt{V(X)}}$
- **Covariance :** $\boxed{cov(X,Y) = E((X - E(X))(Y - E(Y))) = \frac{1}{n}\sum (x - \overline{x})(y - \overline{y})}$
- **Corrélation :** $\boxed{\rho(X,Y) = \frac{cov(X,Y)}{\sigma_x \sigma_y}}$  $-1 \leq \rho(X,Y) \leq 1$
- **Variable centrée :** $\boxed{X^c = X - \overline{X}}$  $\overline{X^c} = 0$
- **Variable centrée-réduite :** $\boxed{X^s = \frac{X - \overline{X}}{\sigma_x}}$  $\overline{X^s} = 0$  $\sigma_x^2 = 1$

## II. MATRICES FONDAMENTALES

### 1. Matrice de variance-covariance

- **Matrice de var-covar :** $X^c$ désigne la mat. des données centrées. On appelle mat. de var-covar la mat. $\Sigma$ :

$$\boxed{\Sigma = \frac{1}{n}\, {}^tX^c \cdot X^c}$$

$\Sigma$ est une mat. symétrique.
Sa trace est dite variance totale.
$Tr\,\Sigma = var(A) + var(B) + var(C)$

$$\Sigma = \begin{bmatrix} \mathrm{var}(x_1^c) & \mathrm{cov}(x_1^c, x_2^c) & \mathrm{cov}(x_1^c, x_3^c) \\ \mathrm{cov}(x_2^c, x_1^c) & \mathrm{var}(x_2^c) & \mathrm{cov}(x_2^c, x_3^c) \\ \mathrm{cov}(x_3^c, x_1^c) & \mathrm{cov}(x_3^c, x_2^c) & \mathrm{var}(x_3^c) \end{bmatrix}$$

**Démonstration:** Données init: $M = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{bmatrix}$

Matrice centrée:

$$M^c = \begin{bmatrix} a_1-\bar{A} & b_1-\bar{B} & c_1-\bar{C} \\ a_2-\bar{A} & b_2-\bar{B} & c_2-\bar{C} \end{bmatrix} \qquad {}^tM^c = \begin{bmatrix} a_1-\bar{A} & a_2-\bar{A} \\ b_1-\bar{B} & b_2-\bar{B} \\ c_1-\bar{C} & c_2-\bar{C} \end{bmatrix}$$

$$\Sigma = \frac{1}{2} \begin{bmatrix} a_1-\bar{A} & a_2-\bar{A} \\ b_1-\bar{B} & b_2-\bar{B} \\ c_1-\bar{C} & c_2-\bar{C} \end{bmatrix} \begin{bmatrix} a_1-\bar{A} & b_1-\bar{B} & c_1-\bar{C} \\ a_2-\bar{A} & b_2-\bar{B} & c_2-\bar{C} \end{bmatrix}$$

Faire les produits $\Rightarrow \Sigma = \begin{bmatrix} \mathrm{var}(A) & \mathrm{cov}(A,B) & \mathrm{cov}(A,C) \\ \mathrm{cov}(B,A) & \mathrm{var}(B) & \mathrm{cov}(B,C) \\ \mathrm{cov}(C,A) & \mathrm{cov}(C,B) & \mathrm{var}(C) \end{bmatrix}$

## 2. Matrice des corrélations R

▫ **Mat. des corrélations:** $X^s$ désigne la mat. des variables centrées réduites. On appelle la mat. des corrélat° R, mat tq:

$$R = \frac{1}{n}\, {}^tX^s \cdot X^s$$ mat. symétrique

$$R = \begin{bmatrix} 1 & \rho(x_1^s, x_2^s) & \rho(x_1^s, x_3^s) \\ \rho(x_2^s, x_1^s) & 1 & \rho(x_2^s, x_3^s) \\ \rho(x_3^s, x_1^s) & \rho(x_3^s, x_2^s) & 1 \end{bmatrix}$$

Demo?

# III. ALGEBRE LINEAIRE

- Valeur propre d'une mat. M : $\lambda$ tq $\det(M - \lambda I) = 0$
- Vecteur propre associé à $\lambda$ : vecteur $u$ tq $Mu = \lambda u$

# CHAP. 3: ANALYSE EN COMPOSANTES PRINCIPALES

## I. APERCU GENERAL

p variables quantitatives : $X_1 \dots X_p$ Pop. de n indiv.

□ Idées de l'ACP : - Transformer les p var init en p nvelles var. → facteurs $f_1, \dots, f_p$
- Conserver k facteurs ($k \in p$) en conservant max d'info. Redondance = corrélation → facteurs pas corrél.

### 1. La transformation

Transf. $(X_1, X_2, \dots, X_p) \longrightarrow (F_1, F_2, \dots, f_p)$
Compression extrême p facteurs

Transf. mat. $\Sigma$ ou R → mat. diagonale ⇒ diagonalisat° de $\Sigma$ ou R.
U, la base de vecteurs propres → facteurs

□ Matrice en composantes principales : coordonnées des indiv. ds la nvelle base.
Chgmt base sur $A^c \rightarrow F = A^c U$
Données init $A^s$ $A^s U$

Donc : - $A^c$ ou $A^s$ : notions d'origine
- U mat. vect. propres de la diagonalisa° de $\Sigma$/R
- $\Lambda$ mat diagonale valeurs propres
- F mat composantes principales

### 2. L'interprétation

Identifier les facteurs → donne significat° concrète fcts.
⇒ Calculer corrélations entre les variables & facteurs

□ Matrice de saturation : pr calculer corrélations

$$S = \frac{1}{n} {}^t X^s . F . \Lambda^{-1/2}$$

# II. ANALYSE EN COMPOSANTES PRINCIPALES - METHODE

A : mat des données ac n lignes et p colonnes
n : nb indiv stat  p : nb variables

Première étape : détermination mat. centrée réduite $A^s$
- ACP centrée et réduite : si var. st ds unités diff, si ordres grandeurs très diff., ou si variances st très éloignées.
- ACP centrée : si cas contraire, déterminer mat. centrée $A^s$

Deuxième étape : détermination de la mat. des corrélations
- ACP centrée-réduite : mat. des corrélations

$$R = \frac{1}{n} {}^tA^s . A^s$$

- ACP centrée : mat. de variance-covariance

$$\Sigma = \frac{1}{n} {}^tA^c \times A^c$$

Troisième étape : diagonalisation de R (ou $\Sigma$)
Déterminer valeurs propres de R
- Mat diagonale $\Lambda$ : formée par vp par ordre ↘
Déterminer vecteurs propres normés associés aux vp.
- Mat chgmt base U : formée par les vecteurs propres.
- Qualité globale d'expression (qge) : permet choisir axes factoriels  qge ⩾ 60%

$$\left| \frac{\lambda_1}{p} + \frac{\lambda_2}{p} \right| \times 100 = qge$$

Quatrième étape : détermination des composantes prin
- Mat. des composantes principales : $F = A^s \times U$

- Qualité de représentation d'un ind. sur un axe : noté qlt ou $\cos^2$

$$\cos^2(m_i ; \text{axe } k) = \frac{\text{coord}(m_i ; \text{axe } k)^2}{\sum \text{coord}(m_i ; \text{axe } j)^2}$$

Si qlt ≈ 1 alr $m_i$ proche axe k
Si qlt ≈ 0 alr $m_i$ éloigné axe k

Cinquième étape : détermination mat. des saturations
- Mat. des saturations : $S = \frac{1}{n} {}^tA^s \times F \times \Lambda^{-1/2}$

## Sixième étape: représentations graphiques

□ Cercle de corrélation : pts variables st situés sur une hypersphère de centre 0 et de rayon 1.

$$S = [s_{ij}]_{\substack{1 \le i \le p \\ 1 \le j \le p}} \quad \sum_{j=1}^{p} s_{ij}^2 = 1 \quad \sum_{i=1}^{p} s_{ij}^2 = \lambda_j$$

Placer pts sur repère d'après mat. des comp. princip.

## Septième étape: interprétation

Élément d'interprétation majeur en ACP : axes correspondent aux nvelles variables → donner nom aux axes
Savoir ac quelle(s) variables les facteurs st corrél.

Deux variables corrélées : évoluent ds manière
Deux variables de corrélat° opposée : évolut° opposée

Ds cercle de corrélation, seuls pts intéressants st situés à proximité des axes ou du cercle.