

REPUBLIQUE ALGERIENNE DÉMOCRATIQUE & POPULAIRE.  
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE  
UNIVERSITÉ HASSIBA BENBOUALI - CHLEF  
FACULTÉ DES SCIENCES EXACTES & INFORMATIQUE  
**Département de Mathématiques**

---

# Module : ANALYSE DE DONNÉES

## CHAP01 : ANALYSE EN COMPOSANTES PRINCIPALES

Cours destiné aux étudiants de Master Mathématiques  
(en cours de rédaction)

PRÉSENTÉ PAR DR :HAMEL ELHADJ<sup>1</sup>

**2020/2021**

---

1. e-mail : hamel\_2@yahoo.fr

# Table des matières

<b>1</b>	<b>Analyse en Composantes Principales</b>	<b>2</b>
1.1	Introduction . . . . .	2
1.2	Les objectifs . . . . .	2
1.3	Notations . . . . .	4
1.3.1	Trois Matrices de données . . . . .	5
1.3.2	matrice de variance-covariance et de corrélation . . . . .	6
1.3.3	La métrique . . . . .	8
1.3.4	L'inertie . . . . .	10
1.4	L'analyse du nuage des individus dans l'espace des variables : . . . . .	14
1.4.1	Le principe : . . . . .	14
1.4.2	les composantes principales . . . . .	17
1.4.3	L'inertie associées aux axes . . . . .	23
1.4.4	Choix de dimension . . . . .	24
1.5	L'analyse du nuage des variables dans l'espace des individus : . . . . .	25
1.5.1	La métrique. . . . .	26
1.5.2	Distances entre les points-variables . . . . .	27
1.5.3	Le principe . . . . .	27
1.5.4	les coordonnées factorielles sur le 1 er axe . . . . .	29
1.5.5	les coordonnées factorielles sur le 2eme axe . . . . .	29
1.5.6	Formule de passage . . . . .	30
1.5.7	conclusion . . . . .	30
1.6	Représentations graphiques et aide à l'interprétation . . . . .	30
1.6.1	Représentations graphiques . . . . .	30
1.6.2	Aide à l'interprétation . . . . .	31
1.6.3	Exemple numérique . . . . .	36
1.7	Exercice . . . . .	37
	<b>Bibliographie</b>	<b>40</b>

# Chapitre 1

## Analyse en Composantes Principales

### 1.1 Introduction

L'analyse en composantes principales - que nous notons par la suite ACP - est une des premières méthodes de l'analyse factorielle, et certainement aujourd'hui l'une des plus employées dans divers domaines de recherche.

L'objectif de l'Analyse en Composantes Principales (ACP) est de revenir à un espace de dimension réduite (par exemple 2) en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé (réduction de dimension) le plus pertinent possible des données initiales. L'analyse en Composantes Principales (ACP) est une méthode de base de l'analyse des données pour l'étude exploratoire ou la compression d'un grand tableau de dimension  $n \times p$  de données quantitatives. D'un point de vue mathématique, l'ACP correspond à l'approximation d'une matrice de taille  $(n, p)$  par une matrice de même dimensions mais de rang  $q < p$ ;  $q$  étant souvent de petite valeur 2, 3 pour la construction de graphiques facilement compréhensibles

Les principales variantes de l'ACP viennent des différences de transformations du tableau de données. Ainsi, le nuage de points peut être centré ou non, réduit ou non. Le cas le plus étudié, et que nous présentons ici, est lorsque le nuage de point est centré et réduit; dans ce cas nous parlons d'ACP normée.

### 1.2 Les objectifs

L'analyse en composantes principales (ACP) est une méthode de traitement des données multidimensionnelles qui poursuit les objectifs suivants :

- La représentation graphique "optimale" des individus (lignes), minimisant les déformations du nuage des points, dans un sous-espace  $\mathcal{E}_q$  de dimension  $q$  ( $q < p$ ),
- La représentation graphique des variables dans un sous-espace  $\mathcal{F}_q$  en explicitant au "mieux" les liaisons initiales entre ces variables,
- La réduction de la dimension (compression), ou approximation d'un tableau  $X$  par un tableau de rang inférieur  $q$  ( $q < p$ )
- Recherche à représenter graphiquement les relations entre individus par l'évaluation de leurs

- ressemblances*<sup>1</sup>, ainsi que les relations entre variables par l'évaluation de leurs *liaisons*<sup>2</sup>.
- Visualiser les données et l'étude des individus et les variables doit se faire simultanément.
  - *Geométriquement*, les données multidimensionnelles forment un nuage des points dans  $\mathbb{R}^p$  (un point de ce nuage correspond à un  $x_i$ ). Si la dimension  $p$  est supérieure à 3, ' ce qui est le plus souvent le cas, on ne peut pas visualiser ce nuage. Le seul moyen de visualiser les données est alors de considérer leurs projections sur des droites, sur des plans ou éventuellement sur des espaces de dimension 3. Ainsi, si  $a = (a_1, \dots, a_p) \in \mathbb{R}^p$  est une direction de projection (c'est-à-dire un vecteur de **norme un** :  $\|a\|^2 = a_1^2 + \dots + a_p^2 = 1$ ), les données projetées ( $a^T x_1, \dots, a^T x_n$ ) forment un échantillon de dimension 1 que l'on peut visualiser et qui est donc plus facile à interpréter que l'échantillon de départ  $(x_1, \dots, x_n)$ . Si la dimension  $p$  est grande, elle est d'habitude redondante. En réalité la "vraie" dimension des données  $q$  est souvent beaucoup plus petite que  $p$  ( $q \ll p$ ). L'ACP a pour but de trouver un sous-espace linéaire de  $\mathbb{R}^p$  de dimension  $q \ll p$  tel que la projection sur ce sous-espace balayé ou bien "capte" presque toute la structure des données.

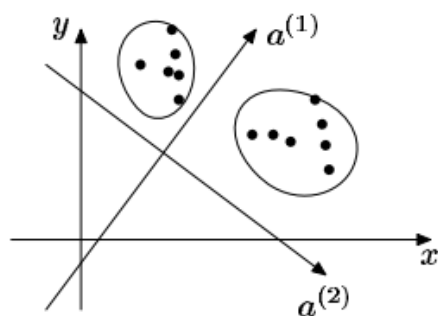


FIGURE 1.1 – Bonne et mauvaises directions de projection.

Dans la Figure 1.1, on voit clairement que si l'on projette les données  $x_i$  (représentées par des points noirs) sur la direction  $a^{(1)}$ , certaines projections coïncideront. Par contre, la projection de ces données sur la direction  $a^{(2)}$  donne des valeurs deux à deux distinctes. On voit que la projection sur cette dernière direction est plus informative que sur la première, donc plus intéressante.

L'idée de base de l'ACP est de chercher la direction  $a \in \mathbb{R}^p$  "la plus intéressante", pour laquelle les données projetées seront le plus dispersées possibles.

**Remarque 1.2.1.** Pour établir un bilan des ressemblances entre les individus, nous cherchons à répondre à des questions du type :

- Quels sont les individus qui se ressemblent ?
- Quelles sont ceux qui sont différents ?
- Existe-t-il des groupes homogènes d'individus ?

1. Deux individus se ressemblent, ou sont proches, s'ils possèdent des valeurs proches pour l'ensemble des variables. Cette définition sous-entend une notion de proximité qui se traduit par une distance.

2. Deux variables sont liées si elles ont un fort coefficient de corrélation linéaire.

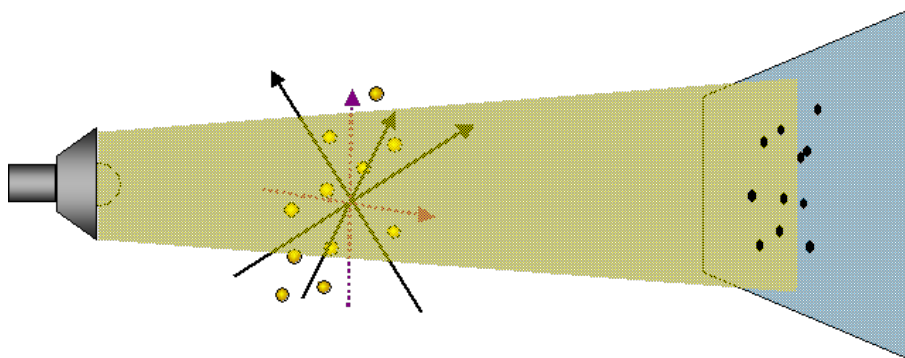


FIGURE 1.2 – La projection du nuages des points

- Est-il possible de mettre en évidence une typologie des individus ?

De la même façon que nous avons défini la ressemblance entre individus, il est essentiel de définir la liaison entre des variables.

### 1.3 Notations

Soit  $p$  variables statistiques réelles  $X^j$  ( $j = 1, \dots, p$ ) observées sur  $n$  individus  $i$  ( $i = 1, \dots, n$ ) affectés des poids  $\omega_i$  :

- $\forall i \in 1, \dots, n$  :  $\omega_i > 0$  et  $\sum_{i=1}^n \omega_i = 1$ .
- $\forall i \in 1, \dots, n$  :  $x_i^j = X^j(i)$  ; mesure de  $X^j$  sur le  $i^{eme}$  individu :

Ces mesures sont regroupées dans une matrice (tableau)  $X$  d'ordre  $(n \times p)$  :

	$X^1$	$\dots$	$X^j$	$\dots$	$X^p$
1	$x_1^1$	$\dots$	$x_1^j$	$\dots$	$x_1^p$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$i$	$x_i^1$	$\dots$	$x_i^j$	$\dots$	$x_i^p$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$n$	$x_n^1$	$\dots$	$x_n^j$	$\dots$	$x_n^p$

- A chaque individu  $i$  est associé le vecteur  $x_i$  contenant la  $i^{eme}$  ligne de  $X$  mise en colonne. C'est un élément d'un espace vectoriel noté  $E$  de dimension  $p$  ; nous choisissons  $\mathbb{R}^p$  muni de la base canonique  $\mathcal{E}$  et d'une métrique de matrice  $M$  lui conférant une structure d'espace euclidien :  $E$  est isomorphe à  $(\mathbb{R}^p, \mathcal{E}, M)$ ,  $E$  est alors appelé *espace des individus*.
- A chaque variable  $X^j$  est associé le vecteur  $x^j$  contenant la  $j^{eme}$  colonne "centrée" (la moyenne de la colonne est retranchée à toute la colonne) de tableau  $X$ . C'est un élément d'un espace vectoriel noté  $F$  de dimension  $n$  ; nous choisissons  $\mathbb{R}^n$  muni de la base canonique  $\mathcal{F}$  et d'une métrique notée  $N$  c'est la matrice diagonale des poids ( $\omega_i$ ) lui conférant une structure d'espace euclidien :  $F$  est isomorphe à  $(\mathbb{R}^n, \mathcal{F}, N)$  avec

$N = \text{diag}(\omega_1, \dots, \omega_n)$ ;  $F$  est alors appelé *espace des variables*.

— **Matrice de poids :**

Chaque individu de l'échantillon peut avoir un poids  $\omega_i$ , tel que  $\sum_{i=1}^n \omega_i = 1$ , notamment quand les individus n'ont pas la même importance (échantillon redressés, données regroupées,...). On a souvent  $\omega = 1/n$  ces poids qui sont des nombres positifs ( $\omega_i > 0$ ), sont regroupée par une matrice diagonale notée  $N$  de taille  $n \times n$  :

$$N = \begin{pmatrix} \omega_1 & & 0 \\ & \ddots & \\ 0 & & \omega_i \end{pmatrix}$$

### 1.3.1 Trois Matrices de données

pour faire l'ACP on distingue trois tableau( matrice ) : *le tableau initiale où la matrice des données brutes*, *les données centrés* et *les données centrés réduites* :

X=		1	...	j	...	p	,	Y=		1	...	j	...	p								
	1								1													
	⋮																	⋮				
	i																	i		...	$y_{ij}$	...
	⋮																	⋮				
	n																	n				
								$y_{ij}=x_{ij}-\overline{x_j}$														
		Données brutes						Données centrés														

#### Matrice des données centrées réduites

En divisant chaque colonne  $j$  du tableau centré  $Y$  par l'écarte-type  $\sigma_j$  de la variable  $j$ , on construit le tableau notée  $Z$  des **données centrées réduites**, tel que

$$Z_i^j = \frac{(x_i^j - \bar{x}^j)}{\sigma_j}$$

La matrice  $Z$  s'exprime aussi en fonction de  $X$  par :

$$Z = (X - \mathbf{1}g^t)D_{1/\sigma}$$

où  $D_{1/\sigma}$  est la matrice diagonale contenant  $\frac{1}{\sigma_1}; \dots; \frac{1}{\sigma_p}$  sur sa diagonale :

$$Z = \begin{array}{c|cccc} & 1 & \cdots & j & \cdots & p \\ \hline 1 & & & & & \\ \vdots & & & \vdots & & \\ i & & \cdots & z_{ij} & \cdots & \\ \vdots & & & \vdots & & \\ n & & & & & \\ \hline & & & z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} & & \end{array}$$

Données centrés réduites

$$D_{1/\sigma} = \begin{pmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_p \end{pmatrix}$$

**Remarque 1.3.1.** *Le terme réduite désigne que les variances des variables  $Z^j$  sont égales à 1 ( $\mathbb{V}(Z^j) = 1$ ).*

donc dans l'ACP ils existes trois matrices différentes pour décrire les même individus cela signifie qu'on on a trois nuages de points différentes .

Dans chaque matrice on a :

- un individu  $i$  est décrit par un vecteur de  $\mathbb{R}^p$  noté  $x_i, y_i, z_i$  .
- une variable  $j$  est décrite par un vecteur de  $\mathbb{R}^n$  noté  $x^j, y^j, z^j$  .

**Remarque 1.3.2.** *On parle toujours de vecteurs colonnes (même s'ils décrivent les lignes des tableaux  $X, Y, Z$  )*

### 1.3.2 matrice de variance-covariance et de corrélation

#### Définition 1.3.1. (*Rappel*)

Soient  $x = (x_1, x_2, \dots, x_n)$  et  $y = (y_1, y_2, \dots, y_n)$  deux échantillons de même taille  $n$  et de moyenne arithmétique respective

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{et} \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

.

On appelle covariance de  $x$  avec  $y$  et on note  $\text{cov}(x, y)$  le nombre

$$\text{cov}_{x,y} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

On appelle variance de  $x$  et on note  $\text{Var}(x)$  le nombre

$$\text{Var}_x = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

et  $\sigma = \sqrt{\text{Var}(x)}$  s'appelle l'écart-type de l'échantillon  $x$ .

On appelle *corrélation* (coefficient de corrélation, dépendance,...) de  $x$  avec  $y$  et on note  $\rho_{xy} = \text{cor}(x, y)$  le

$$\rho = \text{cor} x, y = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\text{Var}(x)} \sqrt{\text{Var}(y)}}$$

- Inégalité de Cauchy-Schwarz :  $(\text{cov}(x, y))^2 \leq \text{Var}(x)\text{Var}(y)$ , et donc  $\rho^2 < 1$ .
- Plus il est proche de  $\rho \simeq 1$ , plus les variables sont corrélées de sens positive.
- Plus il est proche de  $\rho \simeq -1$ , plus elles corrélées de sens négative.
- Un coefficient de corrélation nul i.e  $\rho \simeq 0$ , indique l'absence de corrélation linéaire ( $x$  et  $y$  sont linéairement indépendantes).

**Définition 1.3.2.** La matrice de variance-covariance est une matrice carrée symétrique qui comporte les variances et les covariances associées à plusieurs variables. Les éléments de diagonale de la matrice contiennent les variances des variables, tandis que les éléments hors diagonale contiennent les covariances entre toutes les paires possibles de variables.

On note :

$$\sigma_j^2 = \frac{1}{n} \sum_{i=1}^n (x_i^j - \bar{x}^j)^2$$

la **variance** de la variable  $j$ , sa racines carré noté  $\sigma_j$  est son écarte-type .

$$\text{cov}_{kl} = \frac{1}{n} \sum_{i=1}^n (x_i^k - \bar{x}^k)(x_i^l - \bar{x}^l)$$

est la **covariance** des variables  $k$  et  $l$ .

matrice de variance-covariance noté  $\mathcal{V}$  s'écrit :

$$\mathcal{V} = \frac{1}{n} X^t X - g g^t = \frac{1}{n} Y^t Y = Y^t N Y$$

$$\mathcal{V} = \begin{pmatrix} \text{cov}(x_1, x_1) & \cdots & \text{cov}(x_1, x_p) \\ \vdots & \ddots & \vdots \\ \text{cov}(x_p, x_1) & \cdots & \text{cov}(x_p, x_p) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \cdots & \text{cov}(x_1, x_p) \\ \vdots & \ddots & \vdots \\ \text{cov}(x_p, x_1) & \cdots & \sigma_p^2 \end{pmatrix}$$

**Définition 1.3.3.** La matrice de corrélation notée  $\mathcal{R}$  est une matrice carrée symétrique qui comporte les coefficients de corrélation entre chaque variable et les autres. c'est à dire elle est utilisée pour évaluer la dépendance entre plusieurs variables en même temps.

Donc elle regroupe les coefficients de corrélation linéaire entre les  $p$  variables prises deux à deux, elle résume la structure des dépendances linéaires entre les  $p$  variables. Elle est symétrique et sa diagonale est composée de 1 :

$$\mathcal{R} = \begin{pmatrix} 1 & \cdots & r(x_1, x_p) \\ \vdots & \ddots & \vdots \\ r(x_p, x_1) & \cdots & 1 \end{pmatrix}$$

Matrice de corrélation



La matrice de corrélation définie aussi par

$$\mathcal{R} = D_{1/\sigma} V D_{1/\sigma}$$

où :

$$D_{1/\sigma} = \begin{pmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_p \end{pmatrix}$$

**Exemple 1.3.1.** Soit  $(f, \mu_1, \mu_2)$  un vecteur gaussien  $\mathcal{N}(0, I)$  et  $\beta \in \mathbb{R}, \sigma > 0$ . On pose

$$\xi_1 = \beta f + \sigma \mu_1, \quad \xi_2 = -\beta f + \sigma \mu_2.$$

On note  $X = (\xi_1, \xi_2)^T$

- Donner la loi de  $X$ . ?
- Donner la matrice de covariance de  $X$  notée  $\Sigma$  ?
- Donner la matrice de corrélation de  $X$  ?

**Réponse :**

- Vu que  $X$  est une transformation linéaire d'un vecteur gaussien, lui aussi suit une loi gaussienne.

Les  $\ell$  paramètre de cette loi sont :

$$\begin{aligned} \mathbb{E}(\xi_1) &= \beta \mathbb{E}(f) + \sigma \mathbb{E}(\mu_1) = 0 \\ \mathbb{E}(\xi_2) &= -\beta \mathbb{E}(f) + \sigma \mathbb{E}(\mu_2) = 0 \\ \text{Var}(\xi_i) &= \beta^2 \text{Var}(f) + \sigma^2 \text{Var}(\mu_i) = \beta^2 + \sigma^2 \quad (\perp) \quad i = 1, 2 \\ \text{Cov}(\xi_1, \xi_2) &= -\beta^2 \text{Var}(f) + \sigma^2 \text{Cov}(\mu_1, \mu_2) = -\beta^2 \end{aligned}$$

$$\text{Donc } X \sim \mathcal{N}(0, \Sigma), \text{ avec } \Sigma = \begin{pmatrix} \beta^2 + \sigma^2 & -\beta^2 \\ -\beta^2 & \beta^2 + \sigma^2 \end{pmatrix}, \quad \mathcal{R} = \begin{pmatrix} 1 & \frac{-\beta^2}{\beta^2 + \sigma^2} \\ \frac{-\beta^2}{\beta^2 + \sigma^2} & 1 \end{pmatrix}$$

**Exemple 1.3.2.** illustrer sous R où Matlab ?

`x <- c(4, 6, 2, 9, 0, 1, 2)`

`y <- c(9, 2, 5, 1, 6, 0, 3)`

`mean(x); mean(y)`

`median(x); median(Y)`

`cov(x, y)`

`cor(x, y)`

### 1.3.3 La métrique

L'espace  $\mathbb{R}^p$  est muni d'une structure euclidienne afin de pouvoir définir des distances entre individus .

La distance  $d(i, j)$  entre deux individus  $i$  et  $j$  peut s'écrire de façon générale :

$$d^2(i, j) = d^2(x_i, x_j) = \|x_i - x_j\|_M^2 = \langle x_i - x_j, x_i - x_j \rangle_M = (x_i - x_j)^t M (x_i - x_j)$$

Où

- $M$  est la matrice symétrique définie positive de taille  $p$  spécifiant la distance choisie,
- $\langle ., . \rangle$  et  $\|.\|$  étant le produit scalaire et la norme associée à la métrique  $M$ .

On note que lorsque toutes les variables sont exprimées **dans la même unité**, et que les **variances ne sont pas trop différentes**, la distance euclidienne convient. La métrique est alors la matrice identité  $M = I$  de taille  $p \times p$  et la distance euclidienne s'écrit classiquement :

$$d(i, j) = \left( \sum_{k=1}^p (x_i^k - x_j^k)^2 \right)^{\frac{1}{2}}$$

Dans le cas contraire, la métrique la plus couramment utilisée est **la métrique des inverses des variances**  $M = D_{1/\sigma^2}$ , où  $D_{1/\sigma^2}$  est la matrice diagonale contenant  $\frac{1}{\sigma_1^2}, \dots, \frac{1}{\sigma_p^2}$  sur sa diagonale :

$$M = D_{1/\sigma^2} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_p^2 \end{pmatrix}$$

**Remarque 1.3.3.** *On utilise ces métriques*

- *Pour que les distance soient indépendantes des unités de mesure.*
- *Pour qu'elles ne privilégient pas les variables dispersées.*

Donc la métrique permet à la fois de s'affranchir de l'unité de mesure et de donner à chaque variable la même importance dans le calcul de la distance.

### Équivalence avec les données réduites

:

La matrice de la métrique des inverses des variances peut s'écrire :

$$M = D_{1/\sigma^2} = D_{1/\sigma} D_{1/\sigma}$$

Ainsi, la distance entre deux individus  $i$  et  $j$  est :

$$d^2(i, j) = (x_i - x_j)^t D_{1/\sigma} D_{1/\sigma} (x_i - x_j) = (D_{1/\sigma} x_i - D_{1/\sigma} x_j)^t (D_{1/\sigma} x_i - D_{1/\sigma} x_j)$$

Donc travailler avec la métrique  $M = D_{1/\sigma^2}$  est équivalent à diviser chaque variable par écart-type et à utiliser la métrique identité sur tableau  $D_{1/\sigma} X$ , où encore sur le tableau centré réduit  $Z$ .

L'ACP usuelle revient donc à centrer et réduire les variables puis à utiliser la métrique identité :

c'est ce qu'on appelle une ACP normée.

### Résumé :

on muni l'espace

- $\mathbb{R}^p$  d'une metrique M de dimension  $p \times p$ .
- $\mathbb{R}^n$  d'une metrique N de dimension  $n \times n$ .

L'ACP va consister à analyser les  $n$  points- individus (les lignes) et les  $p$  points-variables ( les colonnes) de la matrice des données centrés(Y) où centré réduite (Z) , avec les métriques

$$M_p = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = I_p \text{ sur } \mathbb{R}^p$$

$$N = \begin{pmatrix} \omega_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \omega_n \end{pmatrix} \text{ sur } \mathbb{R}^p \text{ avec en général } \omega = \frac{1}{n} \implies N = \frac{1}{n} \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = \frac{1}{n} I_n$$

Fait l'analyse avec la triplet  $(\mathbf{Z}, \mathbf{I}_p, \mathbf{N}) \implies \begin{cases} \text{ACP normé.,} \\ \text{ACP avec la matrice de corrélation.,} \end{cases}$

Fait l'analyse avec la triplet  $(\mathbf{Y}, \mathbf{M}_{1/\sigma}, \mathbf{N}) \implies \begin{cases} \text{ACP non normé.,} \\ \text{ACP avec la matrice de variance-covariance.,} \end{cases}$

**Remarque 1.3.4.** On recommande **l'ACP centré** lorsque les variables sont homogène c'est à dire même signification, même unité de mesure, même ordre de grandeur. dans le cas échéant On recommande **l'ACP centré réduit** lorsque les variables sont hétérogène.

### 1.3.4 L'inertie

L'inertie est une notion fondamentale en ACP, puis qu'elle est une mesure de la dispersion du nuage des points autour de son centre de gravité  $g$ .

**Définition 1.3.4.** : On appelle inertie totale du nuage de points  $I_g$ , la moyenne ponderée des carrées des distances des points au  $g$ .

$$\mathbb{I}_g = \frac{1}{n} \sum_{i=1}^n d^2(x_i, g) = \frac{1}{n} \sum_{i=1}^n (x_i - g)^t M (x_i - g)$$

Plus l'inertie est grande, plus le nuage est dispersé, et au contraire plus elle est petite, plus le nuage est concentré sur son centre de gravité.

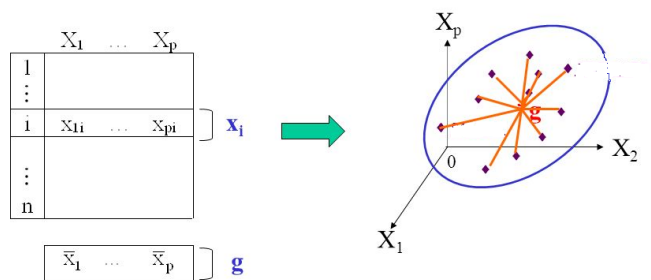
**Proposition 1.3.1.** L'inertie totale est la trace de la matrice  $\mathcal{V}M$  (ou  $MV$ ) :

$$\mathbb{I}_g = \text{tr}(\mathcal{V}M) = \text{tr}(MV)$$

**Preuve :**

$$\mathbb{I}_g = \frac{1}{n} \sum_{i=1}^n (x_i - g)^t M (x_i - g) = \frac{1}{n} \sum_{i=1}^n y_i^t M y_i = \frac{1}{n} \sum_{i=1}^n \text{tr}(y_i^t M y_i)$$

### Inertie totale du nuage de points



$$\begin{aligned} \text{Inertie totale} = I(N, g) &= \frac{1}{n} \sum_{i=1}^n d^2(x_i, g) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_{ji} - \bar{x}_j)^2 = \sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 = \sum_{j=1}^p \sigma_j^2 \end{aligned}$$

FIGURE 1.3 – Inertie Total

$$= \text{tr}\left(\frac{1}{n} \sum_{i=1}^n y_i y_i^t M\right) = \text{tr}\left(\frac{1}{n} Y^t Y M\right) = \text{tr}(VM)$$

**Remarque 1.3.5.**

- Si  $M = I$  inertie totale peut s'écrire :

$$\mathbb{I}_g = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p (x_i - \bar{x}^j)^2 = \sum_{j=1}^p \sum_{i=1}^n (x_i^j - \bar{x}^j)^2 = \sum_{j=1}^p \sigma_j^2$$

L'inertie est donc égale à la somme de variance des variables étudiées qui n'est autre que le trace de la matrice de variance  $\mathcal{V}$ .

- Si  $M = D_{1/\sigma^2}$ , on a :

$$\mathbb{I}_g = \text{tr}(M\mathcal{V}) = \text{tr}(D_{1/\sigma^2}\mathcal{V}) = \text{tr}(D_{1/\sigma}\mathcal{V}D_{1/\sigma})$$

Ce qui égale à

$$\text{tr}(R) = p$$

l'inertie est donc égale au nombre  $p$  de variables et ne dépend pas leurs valeurs .

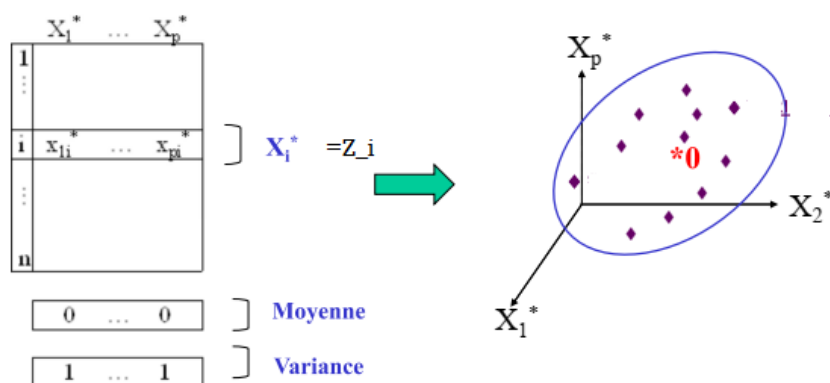
### L'inertie du nuage par rapport à un axe

L'inertie du nuage par rapport à un axe  $\Delta$  passant par le centre de gravité  $g$  est :

$$I_\Delta = \frac{1}{n} \sum_{i=1}^n d^2(\Delta, x_i) = \frac{1}{n} \sum_{i=1}^n d^2(h_{\Delta_i}, x_i)$$

Où  $h_{\Delta_i}$  est la projection orthogonale de  $x_i$  sur l'axe  $\Delta$  , cette inertie mesure la proximité du nuage des individus à l'axe .

### Le nuage de points associé aux données réduites



$$X^* = \{x_1^*, \dots, x_i^*, \dots, x_n^*\}$$

**Centre de gravité :**  $g^* = 0$ , **Inertie totale :**  $I(x^*, 0) = p$

FIGURE 1.4 – Inertie Total des données centrés réduites

### L'inertie du nuage par rapport à un sous-espace vectoriel

l'inertie du nuage par rapport à un sous-espace vectoriel  $F$  passant par le centre de gravité  $g$  est :

$$I_F = \frac{1}{n} \sum_{i=1}^n d^2(h_{F_i}, x_i)$$

où  $h_{F_i}$  est la projection orthogonale de  $x_i$  sur le sous-espace  $F$ . Soit  $F^\perp$  le supplémentaire orthogonal de  $F$  dans  $\mathbb{R}^p$ , alors d'après le théorème de Pythagore (Figure 1 en dimension 3) on a :

$$d^2(h_{F_i}, x_i) + d^2(h_{F^\perp_i}, x_i) = d^2(g, x_i) = d^2(g, h_{F_i}) + d^2(g, h_{F^\perp_i})$$

On déduit le théorème de Huygens<sup>3</sup> :

$$I_F + I_{F^\perp} = I_g$$

Lorsque  $F$  est un axe, l'inertie  $I_F$  par rapport à son supplémentaire orthogonal mesure alors l'allongement du nuage de points selon cet axe  $F$ , et est appelée l'inertie portée par l'axe ou inertie expliquée par l'axe. Ainsi, en projetant le nuage des individus sur un sous-espace  $F$ , on perd l'inertie mesurée par  $I_F$  et on ne conserve que  $I_{F^\perp}$ . Attention à ne pas confondre  $I_F$ , l'inertie par rapport à l'axe  $F$ , et  $\mathbb{I}_{F^\perp}$  l'inertie expliquée ou portée par ce même axe, qui n'est autre que l'inertie par rapport à son orthogonal. Expression de l'inertie  $I_{\Delta_u^\perp}$  portée par un axe  $\Delta_u$  de vecteur directeur  $u$  est :

$$\mathbb{I}_{\Delta_u^\perp} = \frac{1}{n} \sum_{i=1}^n d^2(h_{\Delta_u^\perp_i}, x_i)$$

3. Christian Huygens, est un mathématicien Néerlandais, né le 14 avril 1629 à Haye et mort le 8 juillet 1695 dans la même ville.

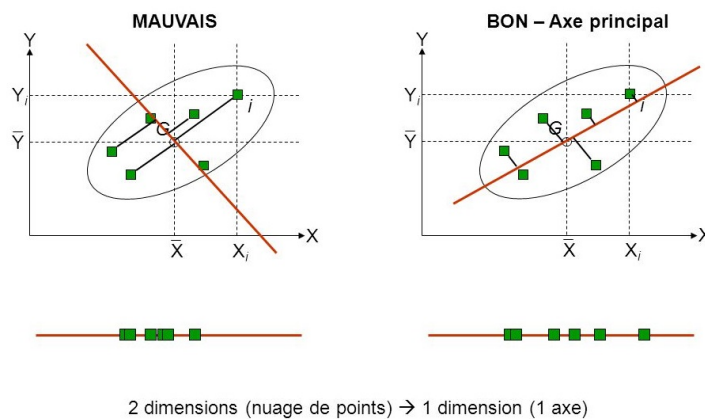


FIGURE 1.5 – Axe principal

Ainsi ,

$$\mathbb{I}_{\Delta_u^\perp} = \frac{1}{n} \sum_{i=1}^n \langle x_i, u \rangle_M^2 = \frac{1}{n} \sum_{i=1}^n (x_i^t M u)^2 = \frac{1}{n} \sum_{i=1}^n u^t M x_i x_i^t M u$$

car  $(x_i^t M u)^t = (x_i^t M u)$  puisque c'est un produit scalaire et que  $M$  est symétrique,

$$\mathbb{I}_{\Delta_u^\perp} = u^t M \left( \frac{1}{n} \sum_{i=1}^n x_i x_i^t \right) M u = \frac{1}{n} u^t M X^t X M u$$

Ainsi, si  $X$  est centrée de matrice de variance  $\mathcal{V}$ , on a :

$$\mathbb{I}_{\Delta_u^\perp} = u^t M \mathcal{V} M u$$

**Exemple 1.3.3.** Des océanographes ont prélevé dans une baie un échantillon de 100 bivalves d'une même espèce. Ils se demandent si les différents facteurs environnementaux dans la baie peuvent avoir une influence sur les caractéristiques morphologiques et biologiques de ces mollusques. Pour cela, ils ont mesuré une série de variables sur chaque individu (tableau ci-dessous).

Les variables	Individu	lg	hg	pd	$O_2$
lg= Longueur (mm)	1	12	4	14	30
hg= Hauteur (mm)	2	19	8	42	10
pd= Poids (cg)	3	16	3	30	22
$O_2$ = Consommation d' $O_2$	4	10	4	13	25
(unité arbitraire)	5	12	6	16	10
	6	15	5	32	50
	variance	.....	.....	117,9	.....

On calcul la matrice de var-cov et de corrélation de notre tableau de données brute. On supposons que la matrice de variance-covariance  $\mathcal{V}$  et de corrélation  $\mathcal{R}$  sont mettent sous la forme ci-dessous :(sont des matrices symétriques ).

$$\mathcal{V} = \left( \begin{array}{c|cccc} & lg & hg & pd & O_2 \\ \hline lg & 9 & & & \\ hg & 2,5 & 2,66 & & \\ pd & 316,66 & 91,66 & 117,9 & \\ O_2 & -6 & -9,83 & -27,5 & 184,58 \end{array} \right), \quad \mathcal{R} = \left( \begin{array}{c|cccc} & lg & hg & pd & O_2 \\ \hline lg & 1 & & & \\ hg & 0,51 & 1 & & \\ pd & 0,97 & 0,51 & 1 & \\ O_2 & -0,14 & -0,44 & -0,01 & 1 \end{array} \right)$$

## 1.4 L'analyse du nuage des individus dans l'espace des variables :

L'objectif est de fournir des images approchées du nuage des individus dans  $\mathbb{R}^p$  l'espace des variables.

### 1.4.1 Le principe :

On considère le nuage des  $n$  points individus dans  $\mathbb{R}^p$  :  $N(I) = \{x_i \in \mathbb{R}^p | i \in 1, \dots, n\}$ . Supposons que la distance entre individus est la distance  $d$  induite par le produit scalaire usuel  $\langle \cdot, \cdot \rangle$  sur  $\mathbb{R}^p$ . Pour toute paire d'individus  $x_i, x_{i'}$  :

$$d^2(x_i, x_{i'}) = \|x_i - x_{i'}\|^2 = \sum_{j=1}^n (x_{ij} - x_{i'j})^2.$$

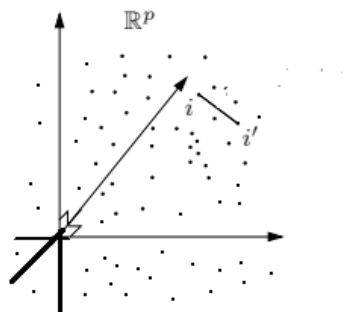


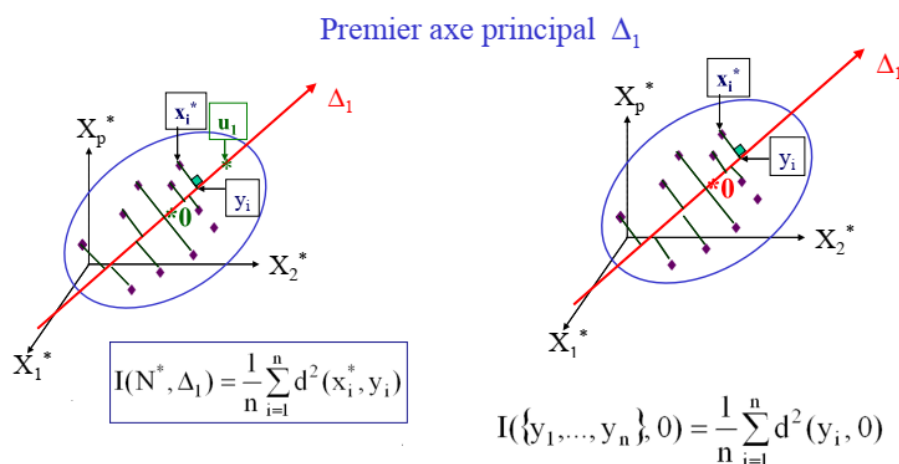
FIGURE 1.6 – Représentation graphique des données.

Donc on va chercher un sous espace vectoriel  $E_k$  de  $\mathbb{R}^p$  de dimension  $k \leq r$  ( avec  $r = \text{rang}(Z)$  c'est à dire  $k$  axes  $\Delta_1, \dots, \Delta_k, (I_p - \text{orthogonaux})$  tels que le nuage de  $n$  points individus  $I_p$  projetés sur  $E_k$ , "déforme le moins possible" les distances euclidiennes entre les individus.

- i-) Dans un premier temps, nous cherchons un sous-espace vectoriel à une dimension, c'est à dire une droite  $\Delta_1$  passant par l'origine, qui ajuste au mieux le nuage des individus  $N_i$ . Nous considérons donc le cas où  $k = 1$ . La projection sur la droite  $\Delta_1$  qui ajuste au mieux le nuage  $N_i$  donne la dispersion ou inertie maximale le long de la droite  $\Delta_1$ .

**L'objectif :** On cherche l'axe  $\Delta_1$  passant le mieux possible au milieu du nuage  $N_i^*$ . On cherche à **minimiser** l'inertie du nuage  $N_i^*$  par rapport à l'axe  $\Delta_1$  :

Donc à la recherche de  $\Delta_1$  :



- Maximiser la dispersion le long de la droite  $\Delta_1$  revient à minimiser les distances des points du nuage  $N_i^*$  à la droite  $\Delta_1$ , c'est -à-dire que la droite  $\Delta_1$  passe au plus près de tous les points du nuage  $N_i^*$ .
- Soit  $u_1$  le vecteur unitaire de la droite  $\Delta_1$  (vecteur de direction) alors :  
maximiser la dispersion le long de la droite  $\Delta_1$  revient à maximiser une forme quadratique définie par

$${}^t(Xu_1)Xu_1$$

- **Le problème** revient donc à trouver  $u_1$  qui maximise cette forme quadratique avec la contrainte  ${}^tu_1u_1 = 1$ . Le sous-espace à une dimension optimal au sens de l'inertie maximale est donc l'axe  $\Delta_1$  défini par le vecteur  $u_1$  solution de ce problème.
- ii-) Cherchons maintenant à déterminer le sous-espace à deux dimensions  $(\Delta_1, \Delta_2)$  s'ajustant au mieux au nuage  $N_i^*$ , nous considérons donc le cas où  $k = 2$ .

- Le sous-espace à deux dimensions qui ajuste au mieux le nuage  $N_i^*$  contient  $u_1$ .
- Le sous-espace à deux dimensions est donc caractérisé par les axes  $(\Delta_1, \Delta_2)$  défini par le vecteur  $u_2$  orthogonal à  $u_1$  vérifiant donc :
  - la forme quadratique  ${}^t(Xu_1)Xu_1$  est maximal.
  - ${}^tu_2u_2 = 1$  (contrainte de normalité)
  - ${}^tu_1u_2 = 0$  (contrainte d'orthogonalité)

Par récurrence, le sous-espace  $F_k$  à  $k$  dimensions s'ajustant au mieux au nuage  $N_i^*$  contient les vecteurs  $u_1, u_2, \dots, u_k$ . Ce sous-espace est engendré par le sous-espace vérifiant :

$$\begin{cases} {}^t(Xu_1)Xu_1, \text{ est maximal} \\ {}^tu_2u_2 = 1, (\text{orthonormée}) \end{cases}$$

**Résumé :** on va chercher un sous espace vectoriel  $F_k$  de  $\mathbb{R}^p$  de dimension  $k \leq r$  ( avec  $r = \text{rang}(Z)$  c'est à dire  $k$  axes  $\Delta_1, \dots, \Delta_k, (I_p - \text{orthogonaux})$  tels que le nuage de  $n$  points individus  $I_p$  projetés sur  $F_k$ , " déforme le moins possible " les distances euclidiennes entre les individus.



$$\begin{array}{ccc}
 \text{Matrice des données centrées réduites} & & \text{Matrice des coordonnées factorielles des individus} \\
 \underbrace{\left( \begin{array}{c|cccc} & 1 & \cdots & j & \cdots & p \\ \hline 1 & & & & & \\ \vdots & & & & & \\ \hline i & & \cdots & z_{ij} & \cdots & \\ \hline \vdots & & & & & \\ n & & & & & \end{array} \right)}_{z_i^t \in \mathbb{R}} & \xRightarrow{\text{projection sur } F_k} \Psi = & \underbrace{\left( \begin{array}{c|ccccc} & 1 & \cdots & \alpha & \cdots & k \\ \hline 1 & & & & & \\ \vdots & & & & & \\ \hline i & & \cdots & \psi_{i\alpha} & \cdots & \psi_i^t \\ \hline \vdots & & & & & \\ n & & & & & \end{array} \right)}_{\psi^\alpha \in \mathbb{R}}
 \end{array}$$

**Définition 1.4.1.** — On appelle  $u_1, u_2, \dots, u_k$  les vecteurs directeurs de  $\Delta_1, \dots, \Delta_k$  les axes principaux .

- On appelle  $\Psi_{n \times k}$  la matrice des coordonnées de  $n$  individus projetés sur ces axes ou bien
  - La matrice des coordonnées factorielles des individus.
  - La matrice des Scores des individus sur les  $k$  premières composantes principales.
- On appelle  $\psi^\alpha \in \mathbb{R}^n$  la  $\alpha^{\text{eme}}$  composante principale.
- On a  $\psi_i \in \mathbb{R}^k$  le vecteur des coordonnées factorielles de l'individu  $i$  sur les  $k$  premier axes c'est à dire le vecteur des scores de l'individu  $i$  sur les  $k$  premières composantes principales.

**Remarque 1.4.1.** pour quoi la phrase " déformer le moins " et qu'el critère vent on optimiser ?

- On la relation suivante :

$$\underbrace{\sum_i \sum_{i'} \omega_i \omega_{i'} d^2(z_i, z_{i'})}_{\text{la somme pondéré}} = \underbrace{2 \sum_i \omega_i d^2(z_i, \bar{z})}_{\text{Inertie}=(Z)=\text{Somme des carrés des distances des distances}}$$

entre paires d'individus                      entre les individus et le centre de gravité

Pour la dernière assertion, il suffit de remarquer la relation suivante entre la moyenne pondérée des carrés des distances entre les points du nuage projeté, et l'inertie du nuage projeté :

- On a toujours la distance entre les individus  $i$  et  $i'$  dans  $\mathbb{R}^p$  plus grande que la distance "en projection" c'est à dire entre  $i$  et  $i'$  dans  $\mathbb{R}^k$  :

$$d^2(z_i, z_{i'}) \geq d^2(\psi_i, \psi_{i'}) \implies I(Z) = P$$

est plus grand que  $I(\Psi)$ . et on va chercher les axes qui maximise  $I(\Psi)$ .

- déformer le moins possible le nuage de  $n$  points individus dans  $\mathbb{R}^p$  (centrés - réduites)  $\iff I(\Psi)$ . or

$$I(\Psi) = \sum_i \omega_i \sum_{\alpha=1}^k d^2(\psi_{i\alpha} - \bar{\psi}^\alpha)^2 = \sum_{\alpha=1}^k \underbrace{\sum_{i=1}^n \omega_i (\psi_{i\alpha} - \bar{\psi}^\alpha)^2}_{\text{var}(\psi^\alpha)} = \sum_{\alpha=1}^k \text{var}(\psi^\alpha)$$

- si  $k=\text{rang}(Z)$  ,  $I(\Psi) = I(Z) = p$

### 1.4.2 les composantes principales

Dans la suite on va utiliser la matrices  $Z$  ( les données centrés réduites ) pour faire l'ACP avec la métrique  $M = I_p$ .

Supposons que l'on souhaite ajuster le nuage de points  $N(I)$  par une droite  $\Delta_1$  (resp  $\Delta_2$ ) qui passe par l'origine.

Notons que  $u_1$  (resp  $u_2$  ) le vecteur unitaire qui engendre cette droite.

Le critère d'ajustement que nous retenons est le critère *des moindres carrés*. Nous demandons que la droite  $\Delta_1$  (resp  $\Delta_2$  )soit le plus proche possible de l'ensemble des points, c'est-à-dire que la somme des carrés des distances entre les points et leur projection orthogonale sur la droite  $\Delta_1$  (resp  $\Delta_2$ ) soit minimale.

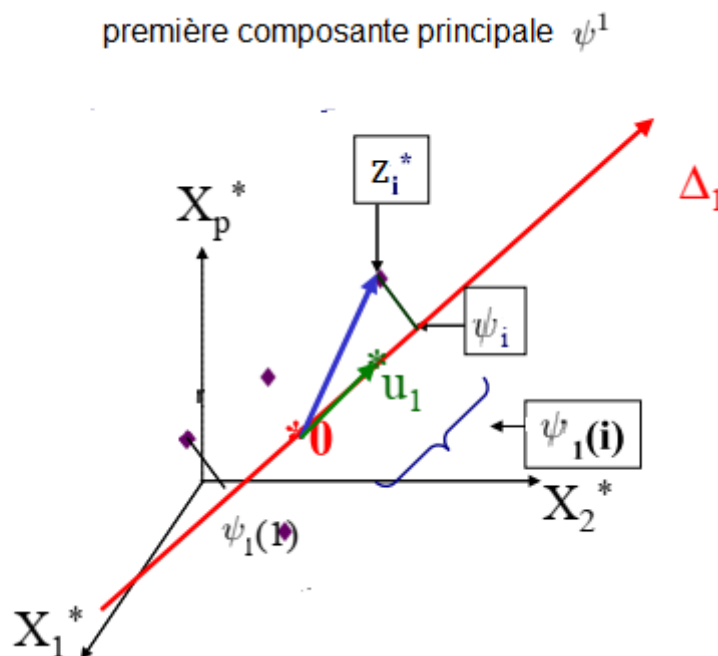
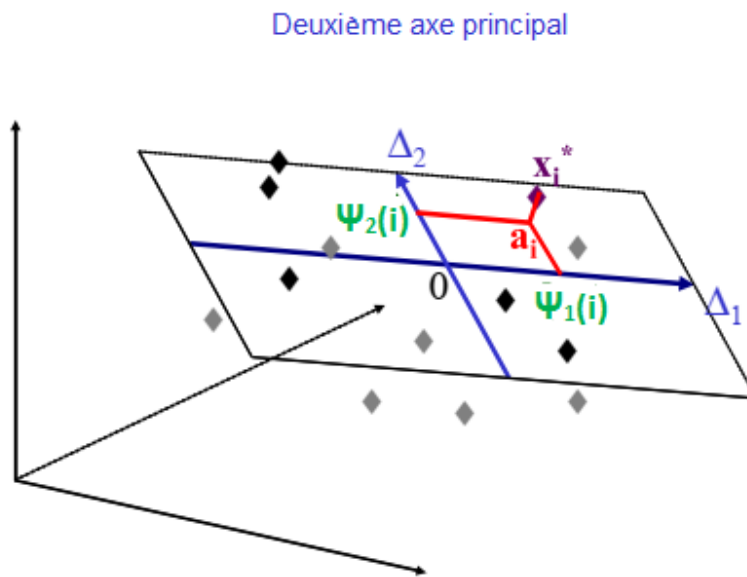


FIGURE 1.7 – Première composante principale  $\psi_1$

$\psi_1$  est une nouvelle variable définie pour chaque individu  $i$  par :

- $\psi_1(i)$  : longueur algébrique du segment " $0\psi_i$ "
- $\psi_1(i)$  : la coordonnée de  $Z_i$  sur l'axe  $\Delta_1$
- $\psi_1(i)$  : produit scalaire entre les vecteurs  $Z_i$  et  $u_1$ .

FIGURE 1.8 – Deuxième composante principale  $\psi_1$ 

$$\underbrace{\begin{pmatrix} \psi_{11} & \psi_{12} \\ \vdots & \vdots \\ \psi_{i1} & \psi_{i2} \\ \vdots & \vdots \\ \psi_{n1} & \psi_{n2} \end{pmatrix}}_{\text{1}^{ere} \text{ et } 2^{eme} \text{ colonnes de } \Psi}$$

**par définition :**

$$- \psi^1 = ZM_p u_1 = ZI_p u_1 = Zu_1 \text{ donc}$$

$$\psi^1 = u_{11} \underbrace{\begin{pmatrix} z_{11} \\ \vdots \\ z_{i1} \\ \vdots \\ z_{n1} \end{pmatrix}}_{\text{1}^{ere} \text{ colonnes de } Z} + \dots + u_{1p} \underbrace{\begin{pmatrix} z_{11} \\ \vdots \\ z_{i1} \\ \vdots \\ z_{n1} \end{pmatrix}}_{p^{eme} \text{ colonnes de } Z}$$

= { combinaison linéaires des colonnes de Z }

$$- \overline{\psi^1} = 0 \implies \text{var}(\psi^1) = \|\psi^1\|_N^2$$

**Propriété 1.4.1.** Quelques propriétés de  $\psi_i$

- $\psi^1 = ZM_p u_1 = ZI_p u_1 = Zu_1$
- La moyenne  $\overline{\psi} = 0$
- La variance (Inertie expliquée par l'axe  $\Delta_1$ )  $\text{Var}(\psi_i) = \lambda_i$ ???

$$- \text{cor}(Z_j, \psi_1) = \sqrt{\lambda_1} u_{1j}$$

**Problème :** comment trouver le vecteur  $u_1, u_2, \dots, u_k$  ? ? ?

**Solution :** soit  $A$  une matrice symétrique . De manière générale, le problème est la recherche du vecteur  $u$  qui rend maximale la quantité  ${}^t u A u$ , avec la contrainte  ${}^t u M u = l$  ,  $\| u \|_M = 1$  , et  $A$  et  $M$  sont des matrices symétriques ;  $M$  est de plus définie positive et définit la métrique dans  $\mathbb{R}^P$ .

La forme quadratique  ${}^t u A u$  s'écrit :

$${}^t u A u = \sum_{ij} a_{ij} u_i u_j$$

En dérivant cette quantité successivement par rapport aux  $p$  composantes du vecteur  $u$ , on voit que le vecteur des dérivées partielles de  ${}^t u A u$  s'écrit sous forme matricielle :

$$\frac{\partial({}^t u A u)}{\partial u} = 2 A u$$

De même :

$$\frac{\partial({}^t u M u)}{\partial u} = 2 M u$$

La recherche d'un maximum lié implique que s'annulent les dérivées du Lagrangien :

$$\mathcal{L} = {}^t u A u - \lambda({}^t u M u - I)$$

$\lambda$  étant un multiplicateur de Lagrange. Par suite :

$$\frac{\partial \mathcal{L}}{\partial u} = 2 A u - 2 \lambda M u = 0$$

exprime la condition d'extremum. On en déduit la relation :

$$A u = \lambda M u \tag{1.1}$$

Prémultipliant les deux membres de cette relation par  ${}^t u$ , et tenant compte du fait que  ${}^t u M u = 1$ , il vient :

$$\lambda = {}^t u A u$$

La valeur du paramètre  $\lambda$  est donc le maximum cherché.

Lorsque la matrice  $M$  est définie positive, donc inversible, la relation 1.1 s'écrit alors :

$$M^{-1} A u = \lambda u \tag{1.2}$$

$u$  est le vecteur propre de la matrice  $M^{-1} A$  correspondant à la plus grande valeur propre  $\lambda$  (si celle-ci est unique, ce qui sera le cas général).

Appelons désormais  $u_1$  le vecteur  $u$  correspondant à la plus grande valeur  $\lambda_1$  telle que la relation 1.1 soit vérifiée.

Cherchons maintenant le vecteur  $u_2$  unitaire et M-orthogonal à  $u_1$  (c'est-à-dire tel que  $u_2 M u_2 = 1$  et  $u_1 M u_2 = 0$ ), qui rend maximale la forme quadratique  ${}^t u_2 A u_2$

$$\mathcal{L} = {}^t u_2 A u_2 - \lambda_2 ({}^t u_2 M u_2 - 1) - \mu_2 {}^t u_2 M u_1$$

où  $\lambda_2$  et  $\mu_2$  sont deux multiplicateurs de Lagrange.

La condition d'extremum s'écrit pour  $\mu_2$  :

$$\frac{\partial \mathcal{L}}{\partial u_2} = 2A u_2 - 2\lambda_2 M u_2 - \mu_2 M u_1 = 0$$

En multipliant les divers membres de cette relation par  ${}^t u_1$ , on voit que  $\mu_2 = 0$  (puisque  ${}^t u_1 A u_2 = \lambda_1 {}^t u_1 M u_2 = 0$ )

Il reste donc comme précédemment :

$$M^{-1} A u_2 = \lambda_2 u_2 \quad (1.3)$$

Quand M est inversible,  $u_2$  est le second vecteur propre de  $M^{-1}A$ , relatif à la seconde plus grande valeur propre  $\lambda_2$  si celle-ci est unique.

On a alors :

$$A u_\alpha = \lambda_\alpha u_\alpha$$

Et si M est inversible :

$$M^{-1} A u_\alpha = \lambda_\alpha u_\alpha$$

**Remarque 1.4.2.** — *Le problème est ici un peu plus général, pour lequel on a  $A = R$  où  $V$  et  $M = I$  où  $(D_{\frac{1}{\sigma}})$  où  $I$  est la matrice unité. Mais cette formulation plus large, avec une métrique et des critères quelconques tels que des masses affectées aux points, sera utile à propos de l'analyse des correspondances et de l'analyse discriminante.*

— **Matrice de variance-covariance et de corrélation**

*L'ensemble des variances  $V_{jj} = \text{Var}(X^j)$  et des covariances  $V_{jj'} = \text{Cov}(X^j, X^{j'})$  est regroupé dans une matrice  $V$ , symétrique, de rang égal à  $p$  en général, si  $n > p$ , sauf s'il existe des relations exactes entre les  $p$  variables :*

$$\begin{pmatrix} V_{11} & V_{12} & \cdot & V_{1p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ V_{p1} & \cdot & \cdot & V_{pp} \end{pmatrix}$$

$$V = {}^t X D_p X - G {}^t G$$

*( ${}^t X$  est la matrice transposée de  $X$ ,  ${}^t G$  le vecteur transposé du vecteur colonne  $G$ ).*

La matrice  $R$  de corrélation est la matrice suivante où  $r_{ij}$  est le coefficient de corrélation linéaire entre les variables  $X^j$  et  $X^{j'}$  :

$$\begin{pmatrix} r_{11} & r_{12} & \cdot & r_{1p} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{p1} & \cdot & \cdot & r_{pp} \end{pmatrix}$$

Si  $D_{\frac{1}{\sigma}}$  est la matrice des inverses des écarts-types, la matrice  $R$  s'écrit :

$$R = D_{\frac{1}{\sigma}} V D_{\frac{1}{\sigma}}$$

### Conclusion

Pour calculer la matrice des coordonnées factorielle des individus (scores des individus) on effectuer la décomposition spectrale de la matrice  $R = {}^t Z N Z$  des corrélations en cherche les  $k$  vecteurs propres  $u_1, u_2, \dots, u_k$  de  $\mathbb{R}^p$   $I_p$  normés à 1 et associes aux plus grandes valeurs propres triées par ordre croissant  $\lambda_1, \dots, \lambda_k$  et on calcule

$$\begin{cases} \psi^\alpha = Z u_\alpha, \\ \text{var}(\psi^\alpha) = \lambda_\alpha, \end{cases}$$

Ces calcul permettent :

- Des représentation graphiques des individus projetés
- Une réduction de la dimension

**Exemple 1.4.1.** D'après l'exemple 1.3.1 on rappelle que :

Soit  $(f, \mu_1, \mu_2)$  un vecteur gaussien  $\mathcal{N}(0, I)$  et  $\beta \in \mathbb{R}, \sigma > 0$ . On pose

$$\xi_1 = \beta f + \sigma \mu_1, \quad \xi_2 = -\beta f + \sigma \mu_2.$$

On note  $X = (\xi_1, \xi_2)^T$

la matrice de covariance notée  $\Sigma$  ? et de de corrélation de  $X$  :

Donc  $X \sim \mathcal{N}(0, \Sigma)$  , avec  $\Sigma = \begin{pmatrix} \beta^2 + \sigma^2 & -\beta^2 \\ -\beta^2 & \beta^2 + \sigma^2 \end{pmatrix}$ ,  $\mathcal{R} = \begin{pmatrix} 1 & \frac{-\beta^2}{\beta^2 + \sigma^2} \\ \frac{-\beta^2}{\beta^2 + \sigma^2} & 1 \end{pmatrix}$

1. Calculer les valeurs propres  $\lambda_1 \geq \lambda_2$  de la matrice de covariance de  $X$ . ?
2. Calculer, en fonction de  $\xi_1$  et  $\xi_2$ , puis en fonction de  $(f, \mu_1, \mu_2)$  les composantes principales  $\psi_1$  et  $\psi_2$  associées à  $X$ . ?
3. Montrer que  $\text{Var}(\psi_i) = \lambda_i$ ,  $\text{Cov}(\psi_1, \psi_2) = 0$ . ?
4. Calculer les corrélations  $\tilde{\rho}_{ij}$  entre  $\xi_i$  et  $\psi_j$ . Montrer que  $\tilde{\rho}_{i1}^2 + \tilde{\rho}_{i2}^2 = 1$ ,  $i = 1, 2$ . ?

### réponse :

1. Pour calculer les valeurs propres de cette matrice, on factorise son polynôme caractéristique :

$$\det(\Sigma - \lambda \mathbb{I}) = (\beta^2 + \sigma^2 - \lambda)^2 - \beta^4 = (2\beta^2 + \sigma^2 - \lambda)(\sigma^2 - \lambda).$$

D'où  $\lambda_1 = 2\beta^2 + \sigma^2$  et  $\lambda_2 = \sigma^2$ .

Pour déterminer le vecteur propre associé à  $\lambda_1$ , il faut résoudre le système linéaire  $(\Sigma - \lambda_1 \mathbb{I})u_1 = 0$  sous la contrainte  $\|u_1\| = 1$ .

On trouve facilement  $u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

De la même façon, on trouve le vecteur propre associé à  $\lambda_2$  :  $u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

2. les composantes principales :

$$\begin{aligned}\psi_1 &= {}^t u_1 X = \frac{\xi_1 - \xi_2}{\sqrt{2}} = \frac{2\beta f + \sigma(u_1 - u_2)}{\sqrt{2}} \\ \psi_2 &= {}^t u_2 X = \frac{\xi_1 + \xi_2}{\sqrt{2}} = \frac{\sigma(u_1 + u_2)}{\sqrt{2}}\end{aligned}$$

3.  $Var(\psi_i) = \lambda_i$ ,  $Cov(\psi_1, \psi_2) = 0$  ?

En utilisant l'indépendance de  $f$ ,  $u_1$  et  $u_2$ , ainsi que la bilinéarité de la covariance, on trouve

$$\begin{aligned}Var(\psi_1) &= \frac{4\beta^2 + \sigma^2 + \sigma^2}{2} = 2\beta^2 + \sigma^2 = \lambda_1 \\ Var(\psi_2) &= \frac{\sigma^2 + \sigma^2}{2} = \sigma^2 = \lambda_2 \\ Cov(\psi_1, \psi_2) &= \frac{Cov(\xi_1 - \xi_2, \xi_1 + \xi_2)}{2} = \frac{Var(\xi_1) + Var(\xi_2)}{2} = 0\end{aligned}$$

4. Calculons les corrélations  $\tilde{\rho}_{ij}$  entre  $\xi_i$  et  $\psi_j$ .

$$\begin{aligned}\tilde{\rho}_{11} &= \frac{Cov(\xi_1, \psi_1)}{\sqrt{Var(\xi_1)}\sqrt{Var(\psi_1)}} = \frac{Var(\xi_1) - Cov(\xi_1, \xi_2)}{\sqrt{2Var(\xi_1)\lambda_1}} = \frac{2\beta^2 + \sigma^2}{\sqrt{2\lambda_1(\beta^2 + \sigma^2)}} = \sqrt{\frac{2\beta^2 + \sigma^2}{2(\beta^2 + \sigma^2)}} \\ \tilde{\rho}_{12} &= \frac{Cov(\xi_1, \psi_2)}{\sqrt{Var(\xi_1)}\sqrt{Var(\psi_2)}} = \frac{Var(\xi_1) + Cov(\xi_1, \xi_2)}{\sqrt{2Var(\xi_1)\lambda_2}} = \frac{\sigma^2}{\sqrt{2\lambda_2(\beta^2 + \sigma^2)}} = \frac{\sigma}{\sqrt{2(\beta^2 + \sigma^2)}} \\ \tilde{\rho}_{21} &= \frac{Cov(\xi_2, \psi_1)}{\sqrt{Var(\xi_2)}\sqrt{Var(\psi_1)}} = \frac{Cov(\xi_1, \xi_2) - Var(\xi_2)}{\sqrt{2Var(\xi_2)\lambda_1}} = \frac{-2\beta^2 - \sigma^2}{\sqrt{2\lambda_1(\beta^2 + \sigma^2)}} = -\sqrt{\frac{2\beta^2 + \sigma^2}{2(\beta^2 + \sigma^2)}} \\ \tilde{\rho}_{22} &= \frac{Cov(\xi_2, \psi_2)}{\sqrt{Var(\xi_2)}\sqrt{Var(\psi_2)}} = \frac{Cov(\xi_1, \xi_2) + Var(\xi_2)}{\sqrt{2Var(\xi_2)\lambda_2}} = \frac{\sigma^2}{\sqrt{2\lambda_2(\beta^2 + \sigma^2)}} = \frac{\sigma}{\sqrt{2(\beta^2 + \sigma^2)}}\end{aligned}$$

On peut vérifier que facilement que

$$\tilde{\rho}_{11}^2 + \tilde{\rho}_{12}^2 + \tilde{\rho}_{21}^2 + \tilde{\rho}_{22}^2 = 1$$

**Exemple 1.4.2. Application numérique :**

soit les matrices

$$X = \begin{pmatrix} 3 & 4 \\ 7 & 2 \\ 5 & 6 \\ 5 & 4 \end{pmatrix}, \quad Z = \begin{pmatrix} -\sqrt{2} & 0 \\ \sqrt{2} & -\sqrt{2} \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix} \quad R = \frac{1}{4} Z Z^t = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

1. Les deux valeurs propres sont  $\lambda_1 = 3/2$  et  $\lambda_2 = 1/2$ .

2. Les vecteurs propres unitaire associés à  $\lambda_1$  et  $\lambda_2$  sont  $u_1 = \begin{pmatrix} \sqrt{2}/2 \\ -\sqrt{2}/2 \end{pmatrix}$ ,  $u_2 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}$

3. Les nouvelles coordonnées des individus (les composantes principales) :

$$\Psi = [\psi_1, \psi_2] = Z.U = \begin{pmatrix} -\sqrt{2} & 0 \\ \sqrt{2} & -\sqrt{2} \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix} = \begin{pmatrix} -1 & -1 \\ 2 & 0 \\ -1 & 1 \\ 0 & 0 \end{pmatrix}$$

### 1.4.3 L'inertie associées aux axes

Rappelons que la valeur propre (ou l'inertie liée à une composante principale) est la variance des coordonnées des points-individus sur l'axe correspondant. C'est un indice de dispersion du nuage des individus dans la direction définie par l'axe.

Dans une analyse normée, la somme des inerties est égale au nombre de variables et donc l'inertie moyenne vaut  $p$ . Chaque axe étant une combinaison particulière des variables d'origine, on s'intéresse en général aux axes ayant une inertie supérieure. On observe souvent une décroissance assez irrégulière des premières valeurs propres.

Si les données sont peu structurées (i.e : les variables ne sont pas fortement corrélées entre elles), le nuage a une forme "régulière". Dans ce cas, les valeurs propres sont régulièrement décroissante et l'analyse factorielle ne fournira pas des résultats intéressants.

**Définition 1.4.2.** Les pourcentages d'inertie représentent la part de la variance (ou inertie) totale prise en compte par chaque facteur (composante). Son appréciation doit cependant tenir compte du nombre de variables et du nombre d'individus. Un taux d'inertie (relatif à un axe) égal à 10% peut être une valeur importante si le tableau possède 100 variables et faible s'il n'en a que 10. ce pourcentage vaut :

1. **Pourcentage de l'inertie du nuage de points initiale expliqué par l'axe  $\Delta_\alpha$**

(a) - Le cas d'un **ACP normé**

$$\Rightarrow \frac{L'_{\text{inertie (variance) du nuage des inds projeté sur } \Delta_\alpha = I(\psi^\alpha)}{L'_{\text{inertie de nuage initiale}}} \times 100 = \frac{\lambda_\alpha}{\lambda_\alpha + \dots + \lambda_p} =$$



$$\frac{\lambda_\alpha}{p}$$

(b) Le cas d'un **ACP non normé**

$$\Rightarrow \frac{L'inerie (variance) \text{ du nuage des inds projeté sur } \Delta_\alpha = I(\psi^\alpha)}{L'inerie \text{ de nuage initiale}} \times 100 = \frac{\lambda_\alpha}{tr(VM)}$$

## 2. Pourcentage de l'inertie du nuage de points initiale expliqué par les k premiers axes

ce pourcentage vaut  $\frac{\lambda_1 + \dots + \lambda_k}{\lambda_1 + \dots + \lambda_p} \times 100$

par exemple le poucentage d'inertie des 03 variables.

	$\lambda_\alpha$	% d'inetie	% inertie cumulé
Axe1	1.73	57.7	57.7
Axe2	1.06	35.4	93.3
Axe3	0.20	6.70	100

### 1.4.4 Choix de dimension

La qualite de représentation aux quelles conduit l'ACP dépend, de façon évidente, du choix de nombre des axes à retenir pour visualiser le nuage de points (ou réduire la dimension ) sans manquer trop d'information , c'est-à-dire du nombre de composantes retenues pour reconstituer les données, ou encore de la dimension du sous-espace de representation.

De nombreux criteres de choix pour les axes de représentation ont été proposés dans la littérature. Nous présentons ici ceux, les plus courants, bases sur une heuristique et un reposant sur une quantification de la stabilite du sous-espace de représentation.

#### 1. Part d'inertie

La "qualité globale" des représentations est mesurée par la part d'inertie expliquée (pourcentage expliqué par l'axe  $\Delta_\alpha$  :

$$I_{\Delta_\alpha} = \frac{\sum_{\alpha=1}^k \lambda_\alpha}{\sum_{\alpha=1}^p \lambda_\alpha} \times 100$$

La valeur de  $k$  est choisie de sorte que cette part d'inertie expliquée  $I_{\Delta_\alpha}$  soit supérieure à une valeur seuil fixée a priori par l'utilisateur (par exemple : choisir k axes pour avoir plus de 80% d'inertie expliquée). C'est souvent le seul critère employé. par exemple :

	$\lambda_\alpha$	% d'inetie	% inertie cumulé
Axe1	1.73	57.7	57.7
Axe2	1.06	35.4	93.3
Axe3	0.20	6.70	100

On va choisit l'axe(1 et 2) pour la représentation (>80%).

#### 2. Règle de Kaiser

le critère de Kaiser (1961), qui stipule de ne retenir que les valeurs propres supérieures à la moyenne des valeurs propres (c'est-à-dire à 1 dans le cas d'une analyse en composantes principales sur matrices de corrélation), en s'appuyant notamment sur des travaux de

Guttman (1954). D'un emploi très répandu à cause de son extrême simplicité, il peut être facilement mis en défaut. Ainsi, une analyse en composantes principales sur matrice des corrélations en biométrie par exemple peut produire un facteur (composante) de taille très dominant. Comme la trace est constante, les autres valeurs propres sont condamnées à être très petites, ce qui pourrait interdire l'interprétation d'autres dimensions.

### 3. Éboulis

C'est le graphique 1.9 présentant la décroissance des valeurs propres.

Le principe consiste à rechercher, s'il existe, un **"coude"** (changement de signe dans la suite des différences d'ordre 2) dans le graphe et de ne conserver que les valeurs propres jusqu'à ce coude.

Intuitivement, plus l'écart  $(\lambda_\alpha - \lambda_{\alpha+1})$  est significativement grand, par exemple supérieur à  $(\lambda_{\alpha-1} - \lambda_\alpha)$ , et plus on peut être assuré de la stabilité de  $E_k$ .

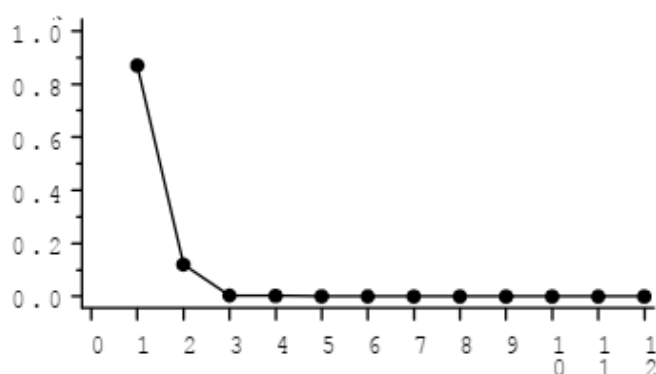


FIGURE 1.9 – éboulis des valeurs propres.

### 4. Diagrammes

Le graphique 1.10, les diagrammes des boîtes des variables principales illustre bien leurs qualités : stabilité lorsqu'une grande boîte est associée à de petites moustaches, instabilité en présence d'une petite boîte, de grandes moustaches et de points isolés.

Intuitivement, on conserve les premières "grandes boîtes". Les points isolés ou "outliers" désignent les points à forte contribution, ou potentiellement influents, dans une direction principale. Ils nécessitent une étude clinique : une autre analyse dans laquelle ils sont déclarés supplémentaires (poids nuls) afin d'évaluer leur impact sur l'orientation des axes

## 1.5 L'analyse du nuage des variables dans l'espace des individus :

L'analyse générale développée dans la section précédente nous a montré qu'en effectuant un ajustement dans un espace, on effectuait implicitement un ajustement dans l'autre espace. Nous avons choisi de commencer en travaillant dans  $\mathbb{R}^P$ . Dans cet espace, la transformation du tableau  $X$  initial avait deux objectifs :

- d'une part obtenir un ajustement qui respecte dans la mesure du possible les distances entre

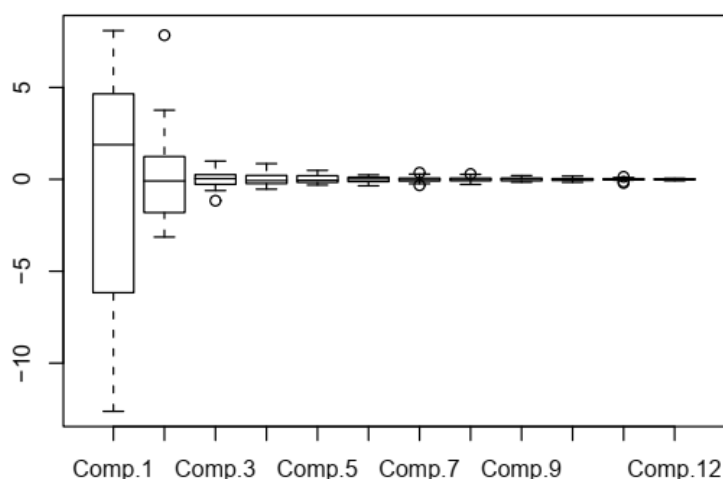


FIGURE 1.10 – éboulis des valeurs propres.

points-individus ;

- d'autre part, faire jouer des rôles similaires à toutes les variables dans la définition des distances entre individus.

Que signifie, dans  $\mathbb{R}^n$ , la proximité entre deux points-variables  $j$  et  $j'$  si l'on prend comme coordonnées de ces variables les colonnes du tableau transformé  $Z$  ?

### 1.5.1 La métrique.

Pour étudier la proximité des variable (caractères) entre eux il faut munir cet espace d'une métrique, c'est-à-dire trouver une matrice d'ordre  $n$  définie positive symétrique. Ici il n'y a pas d'hésitation comme pour l'espace des individus et le choix se porte sur la matrice diagonale des poids  $N$  pour les raisons suivantes :

- Le produit scalaire de deux caractères  $x^j$  et  $x^{j'}$  qui vaut

$${}^t x^j N x^{j'} = \sum_i i = 1^n \omega_i x_i^j N x_i^{j'}$$

n'est autre que la covariance  $S_{jj'}$  car les caractères sont centrés.

- La norme d'un caractère  $\|X^j\|_N$  est alors :  $\|X^j\|_N^2 = \sigma_j^2 = S_{jj}^2$
- La  $n$  longueur  $z$  d'un caractère est égale à son écart type.
- Dans un espace euclidien on définit l'angle  $\theta$  entre deux vecteurs par son cosinus qui est égal au quotient du produit scalaire par le produit des normes des deux vecteurs :

$$\cos \theta = \frac{\langle Z^j, Z^{j'} \rangle}{\|Z^j\| \|Z^{j'}\|} = \frac{S_{jj'}}{\sigma_j \sigma_{j'}}$$

Le cosinus de l'angle  $\theta$  entre deux variable(caractères) centrés n'est donc autre que leur coefficient de corrélation linéaire.

**Remarque 1.5.1.** *Donc Si dans l'espace des individus on s'intéresse aux distances entre les points, dans l'espace des caractères on s'intéressera plutôt aux angles en raison de la propriété précédente.*

### 1.5.2 Distances entre les points-variables

Calculons la distance euclidienne usuelle entre deux variables  $j$  et  $j'$  :

$$d^2(j, j') = \sum_{i=1}^n (Z_{ij} - Z_{ij'})^2 = \sum_{i=1}^n Z_{ij}^2 + \sum_{i=1}^n Z_{ij'}^2 + \sum_{i=1}^n Z_{ij} Z_{ij'}$$

Remplaçant  $Z_{ij}$  par son écriture et tenant compte du fait que :  $\sigma_j^2 = \frac{1}{n} \sum_{i=1}^n (X_{ij} - X_{ij'})^2$

on obtient :  $\sum_{i=1}^n Z_{ij}^2 = \sum_{i=1}^n Z_{ij'}^2 = 1$  et également :  $\sum_{i=1}^n Z_{ij} Z_{ij'} = \rho_{jj'}$

D'où la relation liant la distance dans  $\mathbb{R}^n$  entre deux points-variables  $j$  et  $j'$  et le coefficient de corrélation  $\rho_{jj'}$  entre ces variables :

$$d^2(j, j') = 2(1 - \rho_{jj'}) \implies 0 \leq d^2(j, j') \leq 4. \quad (1.4)$$

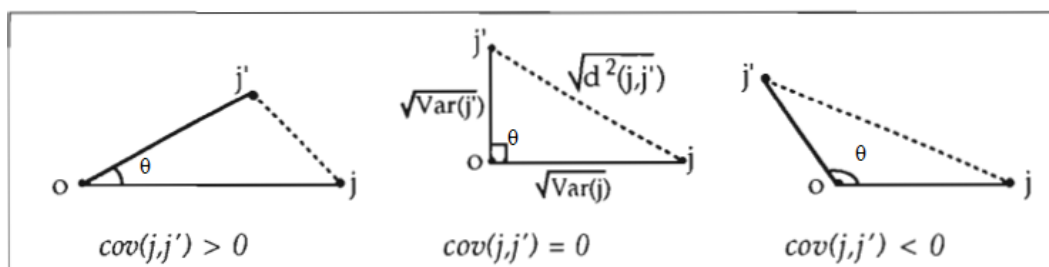


FIGURE 1.11 – Distance entre deux variables

Dans l'espace  $\mathbb{R}^n$ , le cosinus de l'angle de deux vecteurs-variables est le coefficient de corrélation entre ces deux variables ( $\rho_{jj'} = \cos(j, j')$ ). Si ces deux variables sont à la distance 1 de l'origine (i.e. si elles sont de variance unité), le cosinus n'est autre que leur produit scalaire.

Le système de proximités entre points-variables induit par la relation suivante :

- Deux variables fortement corrélées sont très proches l'une de l'autre ( $\rho_{jj'} = 1$ ) ou au contraire les plus éloignées possible ( $\rho_{jj'} = -1$ ) selon que la relation linéaire qui les lie est directe ou indirecte (inverse)
- Deux variables orthogonales ( $\rho_{jj'} \simeq 0$ ) sont à distance moyenne.

### 1.5.3 Le principe

De la même façon que pour le nuage des individus  $N_{i*}$ , nous cherchons une image du nuage des variables dans  $\mathbb{R}^n$

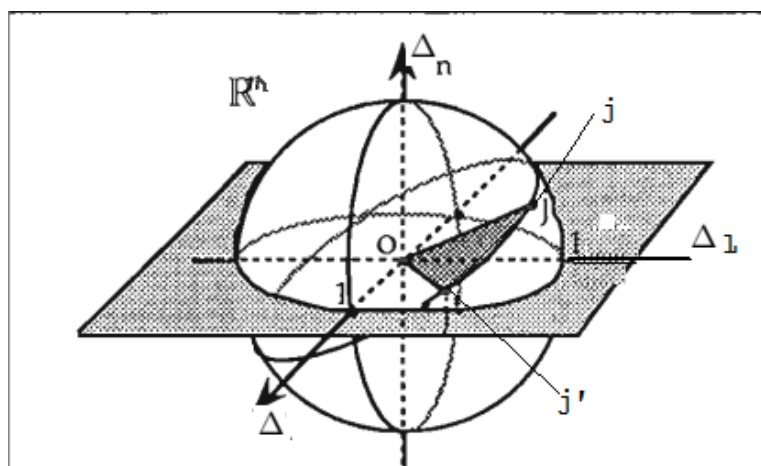


FIGURE 1.12 – La proximité des points variables

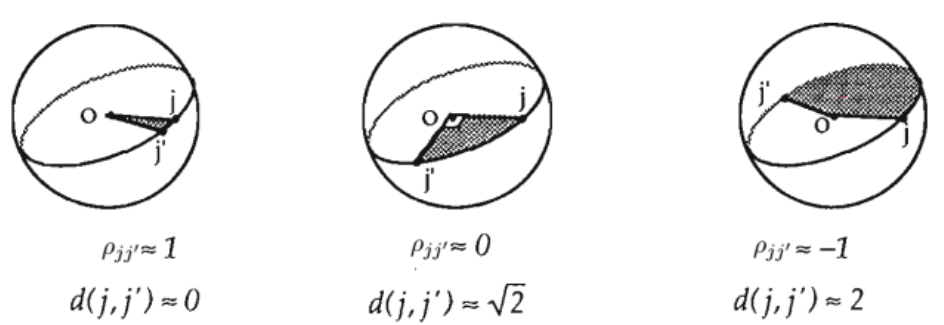


FIGURE 1.13 – Corrélations et distances entre points-variables

Donc on cherche un sous espace vectoriel  $G_{\mathbb{R}}$  de  $\mathbb{R}^n$  de dimension  $k \leq r$  ( $r \leq p = \text{rang de } Z$ ), c'est à dire  $k$  axes  $\Delta^1, \dots, \Delta^k$  ( $N$ -orthogonaux) ( $N$  = la métrique) tels que le nuage des  $p$  points variables  $N$ -projetés sur  $G_{\mathbb{R}}$  "deforme le moins possible" les distances entre les variables et donc leurs corrélations.

En effet, dans une matrice centrée- réduite :

- Les points variables  $z^j$  sont sur l'hypersphère unité  $\forall j, \|Z^j\|_N = \text{var}(z^j) = 1$
- la corrélation entre deux variables  $x^j$  et  $x^{j'}$  est égale au cosinus de l'angle entre  $z^j$  et  $z^{j'}$ .  
 $r_{jj'} = \langle z^j, z^{j'} \rangle_N = \cos_N(\theta)$ .  
 Dans  $\mathbb{R}^n$  si  $z^j$  et  $z^{j'}$  sont proches en terme de distance (avec la métrique  $N$ ) alors  $x^j$  et  $x^{j'}$  sont corrélées.

**Exercice :** montrer que  $d_N^2(z^j, z^{j'}) = 2(1 - r_{jj'})$  ou  $r_{jj'}$  le coefficient de corrélation entre  $x^j$  et  $x^{j'}$  et en déduire que la distance entre deux points variables  $z^j$  et  $z^{j'}$  varie entre 0 et 2 pour une corrélation entre  $x^j$  et  $x^{j'}$  qui varie entre -1 et 1.

$$\begin{array}{ccc}
 \text{Matrice des données centrées réduites} & & \text{Matrice des coordonnées factorielles des variables} \\
 \underbrace{\left( \begin{array}{c|cccc} & 1 & \cdots & j & \cdots & p \\ \hline 1 & & & & & \\ \vdots & & & & & \\ i & & \cdots & z_{ij} & \cdots & \\ \vdots & & & & & \\ n & & & & & \end{array} \right)}_{z_i^t \in \mathbb{R}} & \xRightarrow{\text{N-projection sur } G_k} \Phi = & \underbrace{\left( \begin{array}{c|cccc} & 1 & \cdots & \alpha & \cdots & k \\ \hline 1 & & & & & \\ \vdots & & & & & \\ j & & \cdots & \phi_{j\alpha} & \cdots & \phi_i^t \\ \vdots & & & & & \\ p & & & & & \end{array} \right)}_{\phi^\alpha \in \mathbb{R}}
 \end{array}$$

**Définition 1.5.1.** — On appelle  $v_1, v_2, \dots, v_k$  les vecteurs directeurs de  $\Delta^1, \dots, \Delta^k$  les axes principaux .

- On appelle  $\phi_{p \times k}$  la matrice des coordonnées de  $p$  variables projetés sur ces axes ou bien
- La matrice des coordonnées factorielles des variables.
- La matrice des loadings(saturation) des variable sur les  $k$  premières composantes principales.
- On appelle  $\phi^\alpha \in \mathbb{R}^n$  la  $\alpha^{eme}$  composante principale. on verra plus tard que les loadings sont des corrélations :  $\phi_{j\alpha} = \text{corr}(\psi^\alpha, x^j)$ .
- On a  $\phi_j \in \mathbb{R}^k$  le vecteur des coordonnées factorielles de l'individu  $i$  sur les  $k$  premier axes c'est à dire le vecteur des saturations de variable  $j$  sur les  $k$  premières composantes principales.

#### 1.5.4 les coordonnées factorielles sur le 1 er axe

par définition

$$\phi^1 = {}^t Z.N.v_1$$

Le problème d'optimisation : Trouver le vecteur  $v_1 \in \mathbb{R}^n$  N -normé à 1 telque  $\|\phi^1\|_{I_p}^2$  soit maximale.

**solution** :  $v_1$  est le vecteur propre associe à la plus grande valeur propre  $\lambda_1$  de la matrice  $ZZ^t N$  .

**Remarque 1.5.2.** On verra plus tard que  $\lambda_1$  est aussi la plus grande valeur propre de  $R = {}^t ZNZ$  .(matrice de corrélation ) .

**Exercice** : démontrer le résultat "solution" ?

#### 1.5.5 les coordonnées factorielles sur le 2eme axe

par définition  $\phi^2 = {}^t ZNv_2$

**Le problème d'optimisation** : Trouver le vecteur  $v_2 \in \mathbb{R}^n$  N -normé à 1 c'est à dire  ${}^t v_2 N v_2 = 1$  et  ${}^t v_2 N v_1 = 0$  qui maximise  $\|\phi^2\|_{I_p}^2$  .

**solution** :  $v_2$  est le vecteur propre associe à la plus grande valeur propre  $\lambda_1$  de la matrice  $ZZ^t N$  .

En suivre la même procédure pour les axes suivant c'st à dire pour  $d=3 \dots k$  .

### 1.5.6 Formule de passage

### 1.5.7 conclusion

Pour calculer la matrice  $\phi_{p \times k}$  des *coordonnées factorielles des variables (loadings des variable)*, on effectue la décomposition spectrale de la matrice  $ZZ^tN$  (appelée matrice des produits scalaire des individus) : on cherche les  $k$  vecteurs propre  $v_1, \dots, v_k$  de  $ZZ^tN$ ,  $N$ -normés à 1 associés aux valeurs propre  $\lambda_1, \dots, \lambda_k$  triées par ordre décroissant et on calcule  $\phi^\alpha = {}^tZNv_\alpha$ .

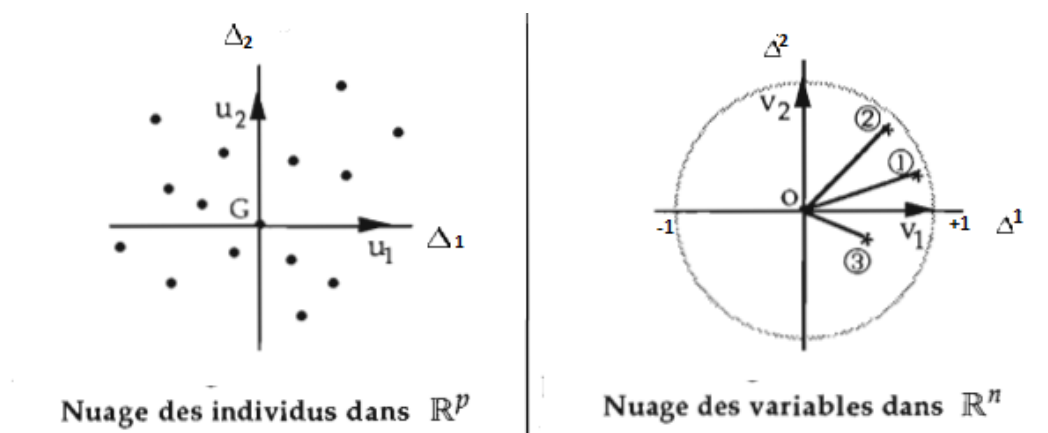
## 1.6 Représentations graphiques et aide à l'interprétation

### 1.6.1 Représentations graphiques

L'analyse du nuage des variables est déduite de celle du nuage des individus : la représentation des variables sur les axes factoriels dans  $\mathbb{R}^n$  aide l'interprétation des axes factoriels dans  $\mathbb{R}^P$  et réciproquement.

Notons que les deux nuages ne sont pas dans le même repère, ce qui rend impossible la représentation simultanée des individus et des variables.

Dans l'espace $\mathbb{R}^P$	Dans l'espace $\mathbb{R}^n$
L'analyse du nuage des $n$ points- L'analyse du nuage des $p$ points individus se fait dans le repère : $\{G, u_1, u_2, \dots, u_p\}$	L'analyse du nuage des $p$ points - variables se fait dans le repère : $\{O, v_1, v_2, \dots, v_p\}$
La représentation des individus sur les axes factoriels fournit la meilleure visualisation approchée des distances entre les individus.	La représentation des variables sur les axes factoriels fournit une synthèse graphique de la matrice de corrélations.



**Remarque 1.6.1.** Les proximités entre individus s'interprètent en termes de similitudes de comportement vis-à-vis des variables et les proximités entre variables en termes de corrélations. Il faut bien se garder d'interpréter la distance séparant un point-variable d'un point-individu car

*ces deux points ne font pas partie d'un même nuage dans un même espace : la superposition de ces deux plans factoriels est dénuée de sens.*

### 1.6.2 Aide à l'interprétation

L'analyse en composantes principales est principalement utilisée pour donner une représentation graphique des individus et des variables.

#### a-) Les individus :

Pour faciliter l'interprétation des axes factoriels trois données sont en général fournies pour chaque individu : *sa coordonnée sur chacun des axes, sa contribution à la formation de chaque axe et sa qualité de représentation par chacun d'entre eux.*

Notons que la valeur propre (ou l'inertie liée à un facteur) est la variance des coordonnées des points-individus sur l'axe correspondant. C'est un indice de dispersion du nuage des individus dans la direction définie par l'axe.

Dans une analyse ACP normée, la somme des inerties est égale au nombre de variables et donc l'inertie moyenne vaut 1. Chaque axe étant une combinaison particulière des variables d'origine, on s'intéresse en général aux axes ayant une inertie "notablement" supérieure à la moyenne.

En pratique, on projette orthogonalement les observations  $Z_i$  sur les plans factoriels. Les coordonnées de  $Z_i$  sur le sous espace de dimension  $q$  sont les  $q$  premiers éléments de la matrice  $R$ . Les graphiques obtenus permettent de représenter au mieux les distances euclidiennes inter-individus.

#### 1-) Les contributions des individus aux axes factoriels.

Les individus contribuent différemment à la formation de chaque axe factoriel. Ainsi la figure 1.14 montre clairement que l'individu 1 contribue beaucoup plus que les autres à déterminer la direction de l'axe d'allongement principal du nuage. donc il est utile d'avoir connaissance de ces contributions lors de l'interprétation des résultats de l'analyse. Comme la  $Var(\psi_\alpha) = \sum \omega_i \frac{(\psi_i^\alpha)^2}{n} = \lambda_\alpha$

on s'intéresse à ceux qui participent à la formation des axes. On calcule LA CONTRIBUTION de chaque point  $i$  (de poids  $\omega_i$ ) à l'inertie de l'axe  $\Delta_\alpha$ . Celle-ci s'exprime par la formule :

$$Ctr_\alpha(i) = \frac{\omega_i \psi_{\alpha i}^2}{\lambda_\alpha}$$

où  $\lambda_\alpha$  est l'inertie de l'axe  $\Delta_\alpha$  et  $\omega_i \psi_{\alpha i}^2$  est la contribution de l'individu  $i$  à l'inertie de cet axe. On a :



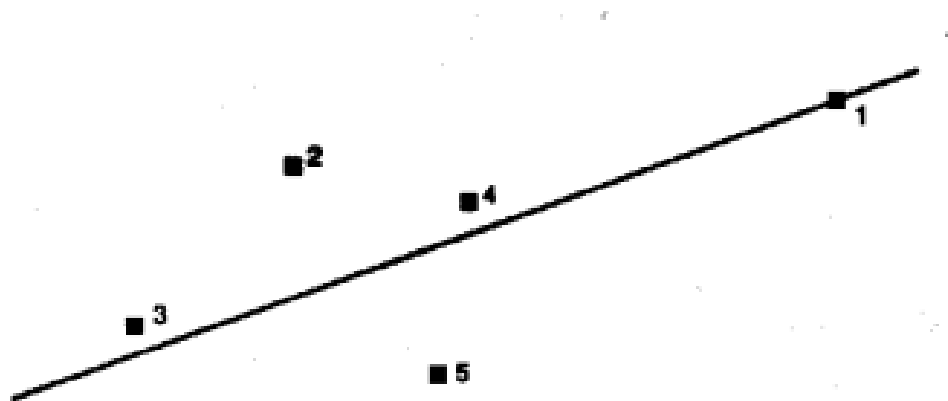


FIGURE 1.14 – L'axe d'un nuage de points

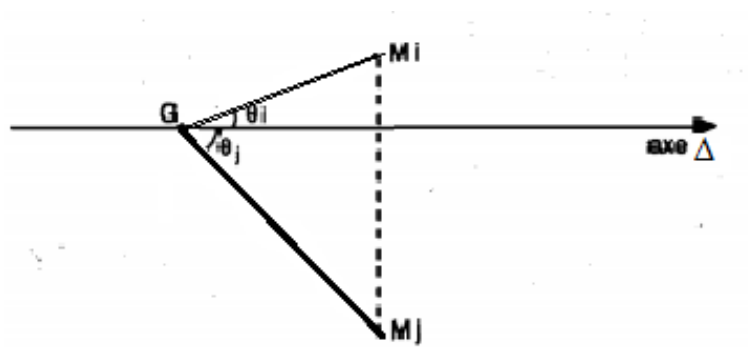
$$\sum_{i=1}^n Ctr_{\alpha}(i) = 1$$

On s'intéressera surtout aux individus qui ont les plus fortes contributions relatives aux axes.

Lorsque les  $n$  individus sont affectés d'une même masse égale à  $1/n$ , l'inertie d'un point varie comme sa distance au centre de gravité. Les individus qui contribuent le plus à la détermination de l'axe sont les plus excentrés et l'examen des coordonnées factorielles ou la lecture du graphique suffisent à interpréter les facteurs dans ce cas.

## 2-) Les qualités de représentation des individus :

L'exemple représenté sur la figure 1.15 montre l'intérêt des qualités de représentation. Les deux points  $M_i$  et  $M_j$  ont des projections confondues sur l'axe  $\Delta$  et ont par conséquent la même coordonnée sur cet axe alors que leurs positions sont très différentes :  $M_i$  est proche de l'axe,  $M_j$  beaucoup plus éloigné. L'individu  $M_i$  sera donc beaucoup mieux caractérisé par l'axe que  $M_j$  mais le seul examen de leurs coordonnées sur l'axe ne le montre pas.

FIGURE 1.15 – Exemple de deux individus différentes ayant la même coordonnée sur  $\Delta$

La qualité de representation notée  $QLT_\alpha(i)$  d'un point  $i$  sur un axe  $\Delta_\alpha$  mesure la proximité de l'individu  $i$  a l'axe  $\Delta_\alpha$  de la façon suivante :

$$QLT_\alpha(i) = \cos^2(\theta_i)$$

où  $\theta_i$  désigne l'angle entre le vecteur individu  $GM_i$  et l'axe  $\Delta_\alpha$ .

En effet, plus un angle est petit, plus son cosinus est grand. Les cas extrêmes sont :

- un cosinus de 1 pour un angle nul correspondant au cas où l'individu est situe sur l'axe  $\Delta_\alpha$  lui- même,
- un cosinus de  $\theta$  pour un angle de 90 correspondant au cas où le vecteur individu est perpendiculaire a l'axe  $\Delta_\alpha$ .

Le developpement de la formule permet d'en donner une interprétation en d'autres termes. Le cosinus d'un triangle rectangle est egal au rapport du côté adjacent sur l'hypothenuse. On a donc :

$$QLT_\alpha(i) = \cos^2(\theta_i) = \frac{d^2(G, \psi_\alpha(i))}{d^2(G, M_i)}$$

où  $d$  représente la distance euclidienne.

Le terme  $d^2(G, Mi)$  représente la distance qui sépare l'individu  $Mi$  du centre de gravité  $G$  du nuage. Etant donné que les axes factoriels sont orthogonaux, ce terme se décompose de la façon suivante :

$$d^2(G, Mi) = \sum d^2(G, \psi_\alpha(i))$$

ce resultat est connu en dimension 2 sous la forme de théorème de Pythagore, mais il est valable quelle que soit la dimension de l'espace étudié.

La qualité de représentation d'un individu par un plan d'axes s'obtient en additionnant celles des deux axes formant ce plan.

$$\sum QLT_\alpha(i) = 1$$

Donc la qualité globale des représentations est mesurée par *la part de dispersion expliquée* ou la *portion d'inertie expliquée* :

$$QLT_\alpha(i) = \frac{\sum_{\alpha=1}^q \lambda_\alpha}{\sum_{\alpha=1}^p \lambda_\alpha}$$

Tandis que la qualité de la représentation de chaque point est donnée par

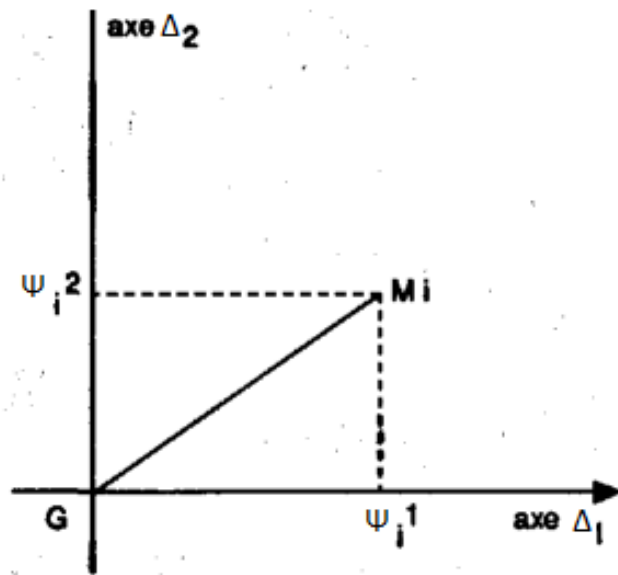


FIGURE 1.16 – Le théorème de Pythagore : Le carré de l'hypothénuse ( $G, Mi$ ) est égale la somme carrée des deux autres cotés

$$QLT_{\alpha}(i) = \cos^2(i) = \frac{\sum_{\alpha=1}^q \lambda_{\alpha}}{\sum_{\alpha=1}^p \lambda_{\alpha}}$$

$$\sum QLT_{\alpha}(i) = 1$$

### b-) Les variables

La plupart des études présentent les résultats relatifs aux variables de manière symétrique à ceux concernant les individus, en spécifiant pour chaque variable *sa coordonnée, sa qualité de représentation et sa contribution par rapport à chacun des axes factoriels*.

On a vu que la coordonnée d'une variable sur un axe factoriel donné est représentée par la projection de cette variable sur l'axe qui n'est autre que le coefficient de corrélation entre la variable et cet axe. Les coordonnées varient donc entre  $-1$  et  $+1$ .

On s'intéresse par conséquent aux variables présentant les plus fortes coordonnées (ce qui les situent proches du cercle de corrélations) et l'on interprétera les composantes principales en fonction des regroupements de certaines de ces variables et de l'opposition avec les autres.

Rappelons que le cosinus de l'angle sous lequel on voit deux points variables actives dans  $\mathbb{R}^n$  n'est autre que le coefficient de corrélation de ces deux variables. Selon la qualité de l'ajustement, cette propriété sera plus ou moins bien conservée en projection. On se gardera d'interpréter la distance entre deux variables actives qui ne sont pas proches du cercle de corrélation.

### 1-) Qualité de représentation ou $\cos^2$ :

Le cercle des corrélations va donc nous permettre de visualiser la corrélation des variables initiales avec les composantes principales ; on pourra donc, à l'aide de ce graphique, déterminer comment sont construites les composantes  $\phi_k$  à partir des  $X^j$  et évaluer la qualité de représentation de chaque  $X^j$  dans le nouveau repère choisi (espace de dimension plus petite).

Une variable  $X^j$  aura une forte coordonnée en valeur absolue sur le  $\alpha - me$  axe factoriel, si la corrélation de  $X^j$  et  $\phi_\alpha$  est importante (c'est-à-dire proche de 1 en valeur absolue). Géométriquement, nous avons

$$r(X^j, \phi_\alpha) = \cos_N(Z^j, \phi_\alpha)$$

$X^j$  sera fortement représentée sur le  $\alpha - me$  axe factoriel si le cosinus de l'angle formé par le vecteur  $X^j$  et celui de  $\phi_\alpha$  est proche de 1.

Ainsi, une coordonnée importante (proche de +1 ou -1) correspond à une bonne qualité de représentation.

### Interprétation du cercle des corrélations :

On a vu précédemment que le cercle des corrélations permet donc de visualiser les variables qui contribuent le plus aux facteurs. Il permet aussi d'étudier les corrélations entre les variables.

- Deux vecteurs orthogonaux sur le cercle des corrélations correspondent à deux variables non-corrélées ; deux vecteurs d'extrémités proches, à des variables fortement corrélées entre elles et de façon positive ; deux vecteurs dont les extrémités sont placées symétriquement par rapport à l'origine, à des variables fortement corrélées négativement. Ces observations pourront être retrouvées sur la matrice des corrélations empiriques.
- Lorsque l'on observe un groupe d'extrémités, c'est qu'il existe un ensemble de variables fortement corrélées entre elles. Il existent donc une "redondance" d'information dans les variables choisies initialement. On pourra éventuellement se limiter à utiliser une variable représentative pour chaque groupe.
- On appelle facteur taille, un facteur (ou composante) où les corrélations de toutes les variables sont de même signe. On appelle facteur forme, un facteur où les va-

riables ayant des corrélations positives avec ce dernier sont opposées à des variables ayant des corrélations négatives avec ce facteur.

## 2-) La contribution d'une variable a un axe

La somme des carrés des coordonnées de l'ensemble des variables sur l'axe est égale à la valeur propre associée à l'axe. La contribution de chaque variable s'obtient en effectuant le rapport entre le carré de sa coordonnée et la valeur propre. Les contributions mesurent donc les rôles relatifs joués par chacune des variables dans la formation de l'axe, et l'addition de toutes les contributions vaut 1.

### 1.6.3 Exemple numérique

Nous supposons que les données du tableau suivant représentent deux variables avec six individus notés A,B,C,D,E et F.

	var1	var2	var3
A	0	1	0
B	2	0	2
C	0	2	5
D	4	2	4
E	4	5	4
moy	2	2	3
$\sigma$	...	....	

les données brutes

	var1	var2	var3
A	-2	-1	-3
B	0	-2	-1
C	-2	0	2
D	2	0	1
E	2	3	1
moy	0	0	0
$\sigma$	...	....	

les données centrées

Nous effectuerons d'abord l'analyse en composantes principales de ce nuage étape par étape puis commenterons les résultats obtenus en montrant quelle structure l'analyse a mise en évidence.

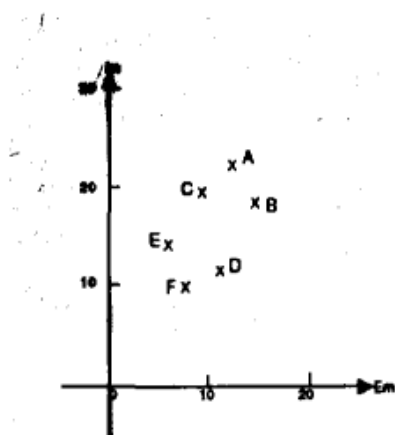
**A-) Centrage et réduction des variables :** L'opération a eu deux conséquences :

- L'origine est placée au centre de gravité du nuage,
- La forme du nuage de points est légèrement modifiée, les variables étant réduites, et elles ont donc toutes la même dispersion.

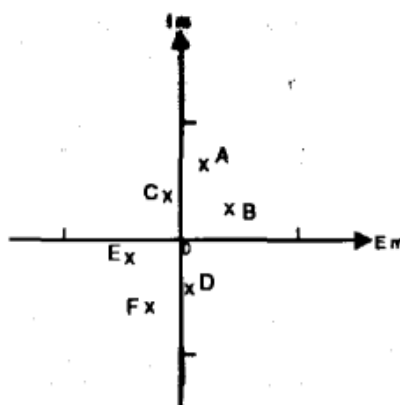
**B-) Les composantes principales :** Le calcul des composantes principales (ou axes factoriels) passe par celui des valeurs propres et vecteurs propres de la matrice des corrélations (ACP normée) ou de variance-covariance (ACP non normée).

$$V = {}^t Y N Y = \begin{pmatrix} 8 & 4 & 3 \\ 4 & 7 & 4 \\ 3 & 4 & 8 \end{pmatrix} \Rightarrow \lambda_1 = 15/2 \quad \lambda_2 = 5/2 \quad \lambda_3 = 3/2.$$

$$u_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad u_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$



Le nuage de points brut.



Le nuage de points centré.

C-) Choix des métriques : Les individus sont supposés tirés de manière équiprobable. On a donc pour l'espace des variables  $\mathbb{R}^n$ , la métrique  $N = \mathbb{I}/n$ . D'autre part, les tirages sont non corrélés et l'on a, comme dans tous les cas usuels, l'espace des individus,  $\mathbb{R}^p$  muni de la métrique  $M = \mathbb{I}_p$ .

D-) Calcul des composantes  $\psi^\alpha$  : Les nouvelles variables  $\psi^\alpha$  nous sont données par la formule suivante :

$$\psi^\alpha = YMu_j = Yu_j = \sum_{k=1}^p u_j^k Y^k$$

où  $M = \mathbb{I}_p$ .

$$\Psi = YU = \begin{pmatrix} \psi^1 & \psi^2 & \psi^3 \\ -6/\sqrt{3} & 1/\sqrt{2} & 3/\sqrt{6} \\ -3/\sqrt{3} & 1/\sqrt{2} & -3/\sqrt{6} \\ 0 & -4/\sqrt{2} & 0/\sqrt{6} \\ 3/\sqrt{3} & 1/\sqrt{2} & -3/\sqrt{6} \\ 6/\sqrt{3} & 1/\sqrt{2} & 3/\sqrt{6} \end{pmatrix} \cong \Psi = YU = \begin{pmatrix} -3.46 & 0.71 & 1.22 \\ -1.73 & 0.71 & -1.22 \\ 0 & -2.82 & 0 \\ 1.73 & 0.71 & -1.22 \\ 3.46 & 0.71 & 1.22 \end{pmatrix}$$

on trouve  $\overline{\psi_1} = 0$ ,  $var(\psi_1) = 7.5 = \lambda_1$  et  $cov(\psi_1, \psi_2) = 0$

E-) Calcul des composantes  $\phi^\alpha$  : on utilise la formule de passage :

F-) Représentation graphique :

## 1.7 Exercice

**Exercice 1.7.1.** Six points du plan ont pour coordonnées, pour  $\alpha > 0$  et  $\beta > 0$  vérifiant  $\alpha^2 + \beta^2 = 1$  :

<i>point</i>	$C_1$	$C_2$
1	1	1
2	-1	-1
3	-1	1
4	1	-1
5	$\alpha$	$\beta$
6	$-\alpha$	$-\beta$

- Calculer la matrice de variance-covariance empirique ( $S$ ) et les valeurs propres de cette matrice ?..
- Calculer l'inertie du nuage de points associée au tableau de données.
- Calculer les vecteurs propres de ( $S$ ) ? les composantes principales de l'ACP sur ( $S$ ) ?.
- Représenter les points et les axes principaux dans le plan de départ ?
- Reprendre i) et ii) quand on n'observe que les points 1, 2, 3, 4.

**Aide :** la matrice de variance-covariance est proportionnelle à la matrice  $A = \begin{pmatrix} 2 + \alpha^2 & \alpha\beta \\ \alpha\beta & 2 + \beta^2 \end{pmatrix}$  et ses valeurs propres ne dépendent pas de  $\alpha$  et  $\beta$ .

**Exercice 1.7.2.** Une étude sur des fournisseurs de matériel informatique a conduit à apprécier le service, la qualité et le prix de quatre fournisseurs. Pour cela un expert a noté ces entreprises avec des notes allant de -3 à 3. Les résultats sont consignés ci-dessous

<i>Ent</i>	<i>Service</i>	<i>Qualit</i>	<i>Prix</i>
E1	-2	3	-1
E2	-1	1	0
E3	2	-1	-1
E4	1	-3	2

1. Calculer le vecteur moyen des données. Qu'en conclure ?
2. Calculer la covariance entre  $x^1$  et  $x^1$ . Que représente cette quantité ?
3. Donner la matrice de variance-covariance et de corrélation. ?
4. Si on veut faire une ACP centrée avec des poids uniformes. Sur quelle matrice faut-il travailler ?
5. Si On veut faire une ACP centrée réduite avec des poids uniformes. Sur quelle matrice faut-il travailler ?
6. pour ACP centrée réduite On donne  $(\lambda_1 = 1,639; \lambda_3 = 0,106)$  En déduire  $\lambda_2$ .
7. Calculer les pourcentages d'inertie des axes. Quelle dimension retenir-vous ?
8. avec ACP centrée réduite calculer les vecteurs propres associés aux valeurs propres ?  
Calculer les deux composantes principales des individus et des variables.
9. Représenter graphiquement les individus et les variables dans le plan principal(1,2). Interpréter.

**Exercice 1.7.3.** Supposons qu'on ait un échantillon de  $n$  individus caractérisés par quatre variables  $\xi_1, \xi_2, \xi_3, \xi_4$  dont les moyennes et les variances sont finies. On se propose d'effectuer l'ACP sur la matrice de covariance  $\mathcal{S}$  d'un tableau  $X$  dont les colonnes sont  $(\xi_1, \xi_2, \xi_3, \xi_4)$  avec la métrique identité.

Supposons que cette matrice se met sous la forme ci-dessous on suppose que  $a, b$  et  $c$  sont des réels :

$$\mathcal{S} = \begin{pmatrix} 1 & a & b & c \\ a & 1 & c & b \\ b & c & 1 & a \\ c & b & a & 1 \end{pmatrix}$$

- Quelle est la signification des coefficients  $a, b, c$  et entre quelles valeurs varient-ils ?
- est ce que on peut supposer que  $\mathcal{S}$  soit une matrice de corrélation ?
- si  $a=b=c=0$  quelle sont les valeurs propres de  $\mathcal{S}$  ?
- soit  $a = 0,1$ ;  $b = 0,4$ ;  $c = 0,6$  . si on suppose que  $\mathcal{S}$  est une matrice de corrélation commenter, selon le contexte, ces coefficients. ? pourquoi on utilise cette matrice dans le cadre de l'ACP normée ? Déterminer les valeurs propres et les vecteurs propres associés de  $\mathcal{S}$
- Calculer pour chacun des axes les pourcentages d'inertie. et l'inertie cumulés .
- Combien d'axes doit être retenir pour la représentation ? Justifier votre réponse.
- Donner la formule permettant de calculer les composantes principales associées au nuage des variables et faire l'application numérique.
- Représenter graphiquement le nuage des individus dans le plan principal et faire l'interprétation. ?



# Bibliographie

- [1] L.Lebart, A. Morineau, M. Piron (1995). Statistique exploratoire multidimensionnelle, Dunod.
- [2] G. Saporta, (2005). Probabilités, analyses des données et statistiques , Editions Technip
- [3] Bouroche, J.-M. et Saporta, G. (1987). Analyse des Données. Que sais-je ?, Paris.
- [4] Escofier.B et J. pagès . (2008) .Analyses factorielles simples et multiples, Dunod, Paris, .
- [5] Sanders Lena,(1989). L'anaLyse des données appliquée à la géographie. Montpellier, G.I.P. RECLUS, .