

Chapitre 2

Analyse en Composantes Principales

2.1 Introduction

L'analyse en Composantes Principales (ACP) est une méthode de base de l'analyse des données pour l'étude exploratoire ou la compression (réduction) d'un grand tableau de dimension $n \times p$ de données quantitatives et certainement aujourd'hui l'une des plus employées dans divers domaines de recherche.

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 est lorsque le nuage de point est centré et réduit, dans ce cas nous parlons d'ACP normée.

2.2 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*¹, ainsi que les relations entre variables par l'évaluation de leurs *liaisons*².
- Visualiser les données et l'étude des individus et les variables doit se faire simultanément.

2.3 Notations

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

- $\forall i \in 1, \dots, n \quad : \omega_i > 0 \quad \text{et} \quad \sum_{i=1}^n \omega_i = 1.$
- $\forall i \in 1, \dots, n \quad : x_i^j = X^j(i)$; 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 chois-

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.

sons \mathbb{R}^n muni de la base canonique \mathcal{F} et d'une metrique notée N c'est la matrice diagonale des poids (ω_i) lui conferant 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 ω_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 $n \times n$:

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

2.3.1 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		,
		1						
		:						
		i	...	x_{ij}	...			
		:						
		n						

$Y=$		1	...	j	...	p	
		1					
		:					
		i	...	y_{ij}	...		
		:					
		n					
		$y_{ij} = x_{ij} - \bar{x}_j$					

Données brutes

Données centrés

Matrice des données centrées réduites

En divisant chaque colonne j du tableau centré Y par l'écarte-type σ_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 :

$$Z = \begin{array}{c|cccccc} & 1 & \dots & j & \dots & p \\ \hline 1 & & & & & \\ \vdots & & & \vdots & & \\ i & & \dots & z_{ij} & \dots & \\ \vdots & & & \vdots & & \\ n & & & & & \\ \hline & & & z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} & & \end{array} \quad D_{1/\sigma} = \begin{pmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_p \end{pmatrix}$$

Données centrés réduites

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)

2.3.2 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 $cov(x, y)$ le nombre

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

On appelle variance de x et on note $Var(x)$ le nombre

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

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

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

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

- Inégalité de Cauchy-Schwarz : $(cov(x, y))^2 \leq Var(x)Var(y)$, 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 σ_j est son écarte-type .

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

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

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

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

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

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

$$\mathcal{R} = D_{1/\sigma} V D_{1/\sigma} \quad \text{où} \quad D_{1/\sigma} = \begin{pmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_p \end{pmatrix}$$

Exemple 1.3.1. Soit (f, μ_1, μ_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 Σ ?

— 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 t' parametre 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 sour 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)
```

2.3.3 La métrique

L'espace \mathbb{R}^p est muni d'une structure euclidienne afin de pouvoire 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,
- $\langle \cdot, \cdot \rangle$ et $\|\cdot\|$ 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/σ^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 métrique M de dimension $p \times p$.
- \mathbb{R}^n d'une métrique 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.*

2.3.4 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 pondérée 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 $M\mathcal{V}$) :*

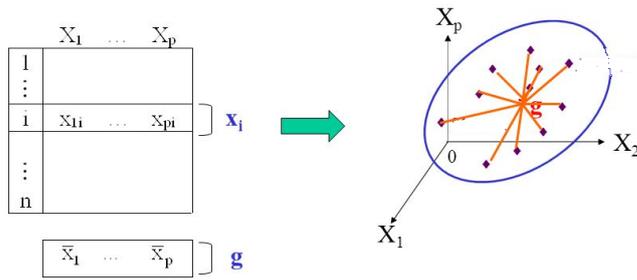
$$\mathbb{I}_g = \text{tr}(\mathcal{V}M) = \text{tr}(M\mathcal{V})$$

Preuve :

$$\begin{aligned} \mathbb{I}_g &= \frac{1}{n} \sum_{i=1}^n (x_i - g)^t M (x_i - g) = \frac{1}{n} \sum_{i=1}^n y_i^t M y_i = \frac{1}{n} \sum_{i=1}^n \text{tr}(y_i^t M y_i) \\ &= \text{tr}\left(\frac{1}{n} \sum_{i=1}^n y_i y_i^t M\right) = \text{tr}\left(\frac{1}{n} Y^t Y M\right) = \text{tr}(VM) \end{aligned}$$

Remarque 1.3.5.

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 1.1 – Inertie Total

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

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

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 .

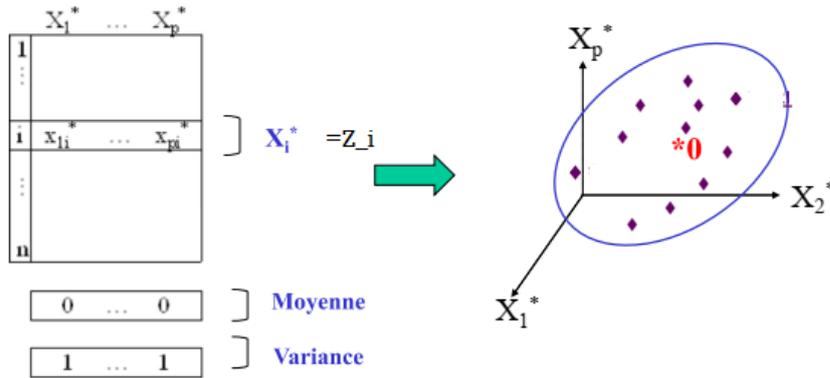
L'inertie du nuage par rapport à un axe

L'inertie du nuage par rapport à un axe Δ 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_{Δ_i} est la projection orthogonale de x_i sur l'axe Δ , 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 1.2 – 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 Huygens³ :

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

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

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

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 Δ_u de vecteur directeur u est :

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

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 :

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

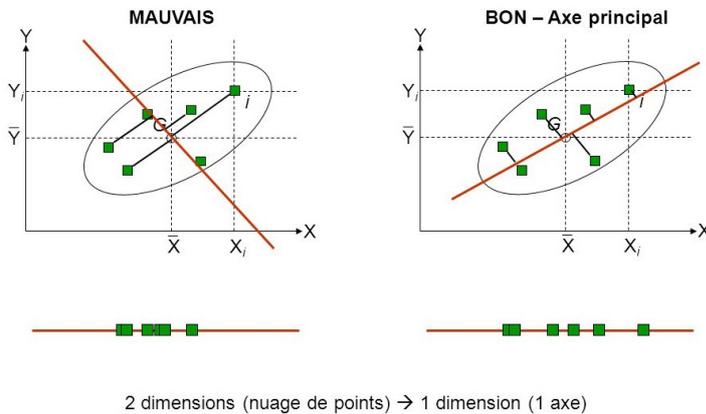


FIGURE 1.3 – Axe principal

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 (unité arbitraire)	4	10	4	13	25
	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 suppose que la matrice de variance-covariance \mathcal{V} et de corrélation \mathcal{R} sont mises sous la forme ci-dessous : (sont des matrices symétriques).

$$\mathcal{V} = \begin{pmatrix} & \begin{matrix} lg & hg & pd & O_2 \end{matrix} \\ \begin{matrix} lg \\ hg \\ pd \\ O_2 \end{matrix} & \begin{pmatrix} 9 & & & \\ 2,5 & 2,66 & & \\ 316,66 & 91,66 & 117,9 & \\ -6 & -9,83 & -27,5 & 184,58 \end{pmatrix} \end{pmatrix}, \quad \mathcal{R} = \begin{pmatrix} & \begin{matrix} lg & hg & pd & O_2 \end{matrix} \\ \begin{matrix} lg \\ hg \\ pd \\ O_2 \end{matrix} & \begin{pmatrix} 1 & & & \\ 0,51 & 1 & & \\ 0,97 & 0,51 & 1 & \\ -0,14 & -0,44 & -0,01 & 1 \end{pmatrix} \end{pmatrix}$$

2.4 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.

2.4.1 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 $x_i, x_{i'}$:

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

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 $\Delta_1, \dots, \Delta_k(I_p)$ - orthogonaux) tels que le nuage de n points

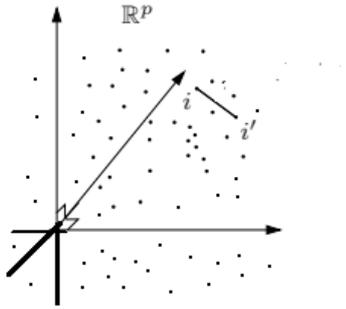
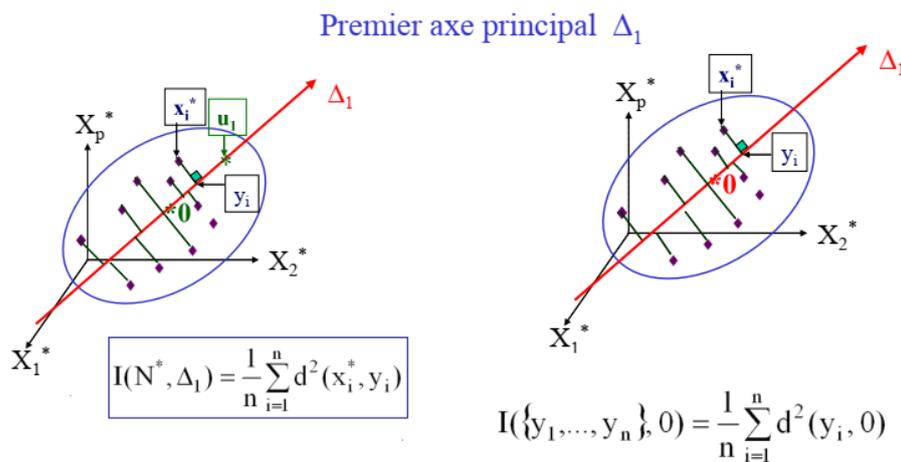


FIGURE 1.4 – Représentation graphique des données.



individus I_p projetés sur F_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 Δ_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 Δ_1 qui ajuste au mieux le nuage N_i donne la dispersion ou inertie maximale le long de la droite Δ_1 .

L'objectif : On cherche l'axe Δ_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 Δ_1 :

Donc à la recherche de Δ_1 :

- Maximiser la dispersion le long de la droite Δ_1 revient à minimiser les distances des points du nuage N_i^* à la droite Δ_1 , c'est -à-dire que la droite Δ_1 passe au plus près de tous les points du nuage N_i^* .

-
- Soit u_1 le vecteur unitaire de la droite Δ_1 (vecteur de direction) alors : maximiser la dispersion le long de la droite Δ_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 Δ_1 défini par le vecteur u_1 solution de ce problème.

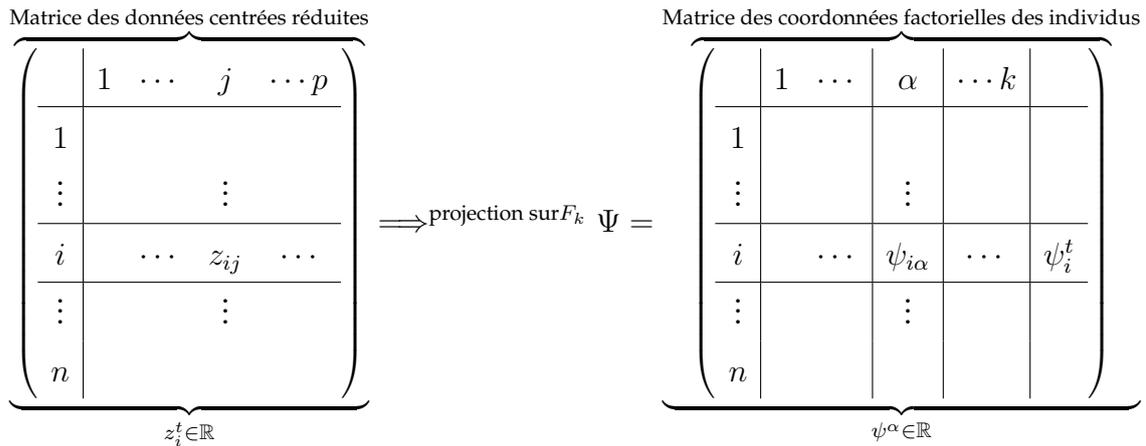
ii-) Cherchons maintenant à déterminer le sous-espace à deux dimensions (Δ_1, Δ_2) s'ajustant au mieux au nuage N_i^* , nous considérons donc le cas où $k = 2$.

- 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 (Δ_1, Δ_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)
 - ${}^tu_2u_1 = 0$ (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 - 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.



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 α^{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 $k=\text{rang}(Z)$, $I(\Psi) = I(Z) = p$

2.4.2 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 Δ_1 (resp Δ_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 moindre carrés*. Nous demandons que la droite Δ_1 (resp Δ_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 Δ_1 (resp Δ_2) soit minimale.

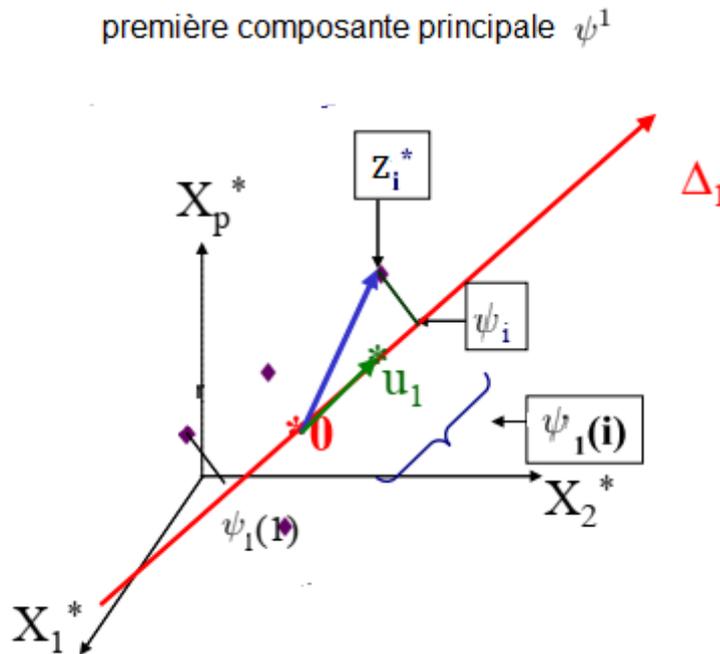


FIGURE 1.5 – Première composante principale ψ_1

ψ_1 est une nouvelle variable définie pour chaque individu i par :

- $\psi_1(i)$: longueur algébrique du segment "0 ψ_i "
- $\psi_1(i)$: la coordonnée de Z_i sur l'axe Δ_1
- $\psi_1(i)$: produit scalaire entre les vecteurs Z_i et u_1 .

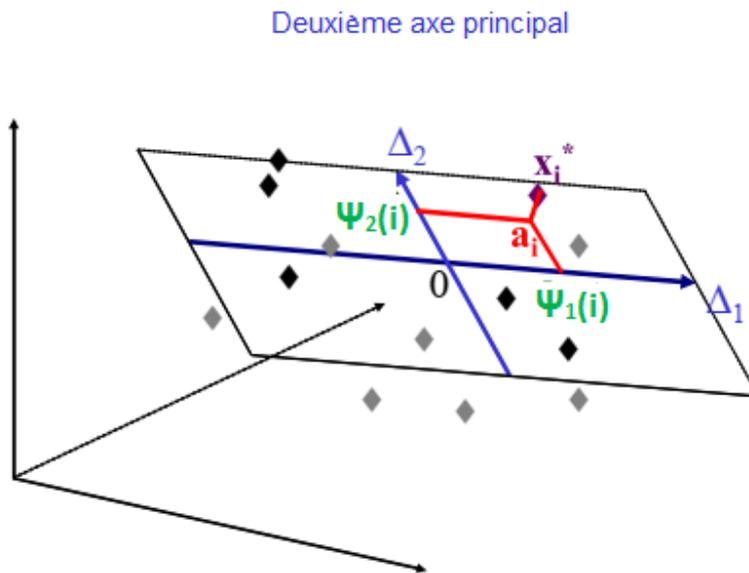


FIGURE 1.6 – Deuxième composante principale ψ_1

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

par définition :

— $\psi^1 = ZM_p u_1 = ZI_p u_1 = Z u_1$ donc

$$\psi^1 = u_{11} \begin{pmatrix} z_{11} \\ \vdots \\ z_{i1} \\ \vdots \\ z_{n1} \end{pmatrix} + \dots + u_{1p} \begin{pmatrix} z_{11} \\ \vdots \\ z_{i1} \\ \vdots \\ z_{n1} \end{pmatrix}$$

$\underbrace{\hspace{10em}}_{1^{ere} \text{ colonnes de } Z}$
 $\underbrace{\hspace{10em}}_{p^{ere} \text{ colonnes de } Z}$

= { combinaison linéaires des colonnes de Z }

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

Propriété 1.4.1. Quelques propriétés de ψ_i

- $\psi^1 = ZM_p u_1 = ZI_p u_1 = Z u_1$
- La moyenne $\bar{\psi} = 0$
- La variance (Inertie expliquée par l'axe Δ_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 général, 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 :

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

λ étant un multiplicateur de Lagrange. Par suite :

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

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

$$A u = \lambda M u \quad (1.1)$$

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 λ est donc le maximum cherché.

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

$$M^{-1} A u = \lambda u \quad (1.2)$$

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

Appelons désormais u_1 le vecteur u correspondant à la plus grande valeur λ_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 λ_2 et μ_2 sont deux multiplicateurs de Lagrange.

La condition d'extremum s'écrit pour μ_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 λ_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 effectue 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 associés 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, μ_1, μ_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 Σ ? et de de corrélation de X :

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

1. Calculer les valeurs propres $\lambda_1 \geq \lambda_2$ de la matrice de covariance de X ?
2. Calculer, en fonction de ξ_1 et ξ_2 , puis en fonction de (f, μ_1, μ_2) les composantes principales ψ_1 et ψ_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 ξ_i et ψ_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 λ_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 λ_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 ξ_i et ψ_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

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

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

2. Les vecteurs propres unitaire associés à λ_1 et λ_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) :

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

2.4.3 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 Δ_α

(a) - Le cas d'un ACP normé

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

(b) Le cas d'un ACP non normé

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

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

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

par exemple le pourcentage d'inertie des 03 variables.

	λ_α	% d'inertie	% inertie cumulé
Axe1	1.73	57.7	57.7
Axe2	1.06	35.4	93.3
Axe3	0.20	6.70	100

2.4.4 Choix de dimension

La qualité 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 représentation.

De nombreux critères de choix pour les axes de représentation ont été proposés dans la littérature. Nous présentons ici ceux, les plus courants, basés sur une heuristique et un reposant sur une quantification de la stabilité 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 Δ_α :

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

La valeur de k est choisie de sorte que cette part d'inertie expliquée I_{Δ_α} 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 :

	λ_α	% d'inertie	% inertie cumulé
Axe1	1.73	57.7	57.7
Axe2	1.06	35.4	93.3
Axe3	0.20	6.70	100

On va choisir 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 ana-

lyse 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 1.7 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 .

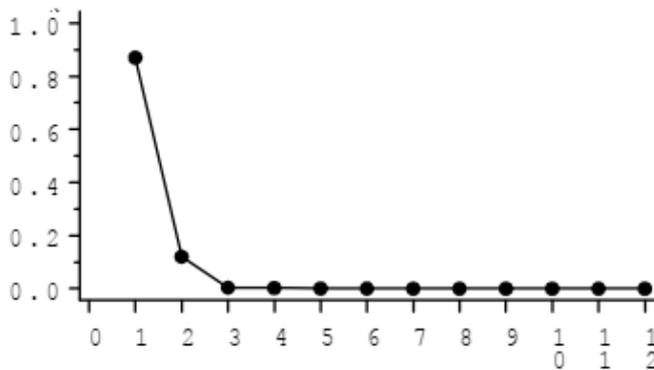


FIGURE 1.7 – éboulis des valeurs propres.

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

2.5.1 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

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

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

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

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

Le cosinus de l'angle θ 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.*

2.5.2 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 :

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

Dans l'espace \mathbb{R}^n , le cosinus de l'angle de deux vecteurs-variables est le coefficient de corrélation entre ces deux variables ($\rho_{jj'} = \cos(\mathbf{j}, \mathbf{j}')$). 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.

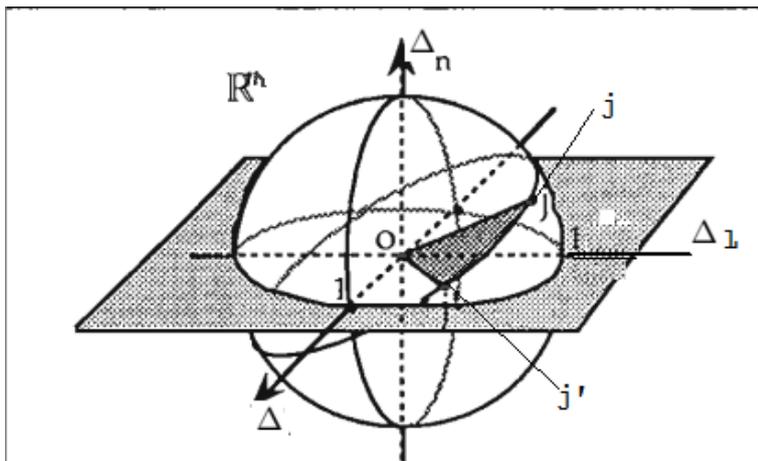


FIGURE 1.8 – La proximité des points variables

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.

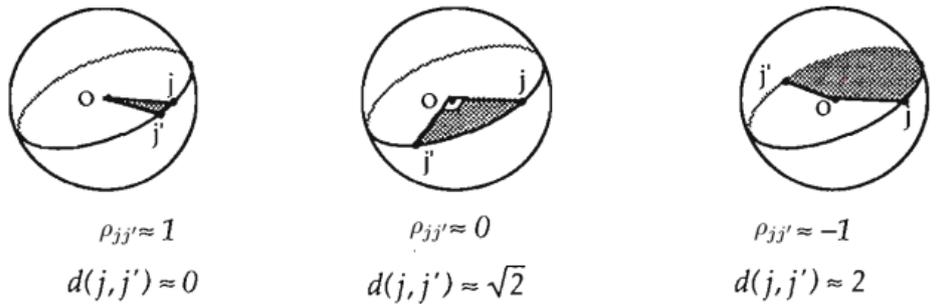


FIGURE 1.9 – Corrélations et distances entre points-variables

2.5.3 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

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 $G_{\mathbb{R}}$ "déforme le moins possible" les distances entre les variables et donc leurs corrélations.

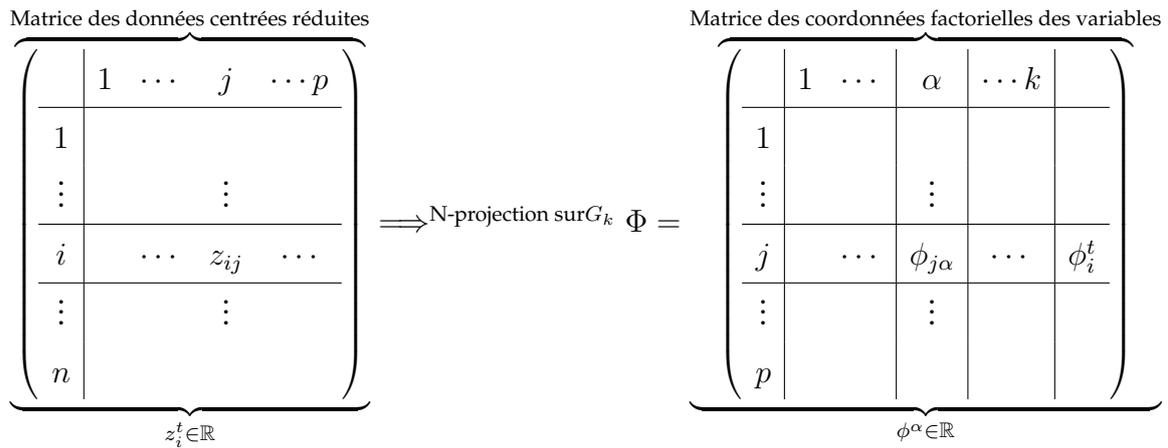
En effet, dans une matrice centrée- réduite :

- 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.



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 α^{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.

les coordonnées factorielles sur le 1 er axe

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

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 λ_1 de la matrice ZZ^tN .

Remarque 1.5.2. On verra plus tard que λ_1 est aussi la plus grande valeur propre de $R = {}^t ZNZ$.(matrice de corrélation) .

Excercice : démentrer le résultat "la solution du problème" ?

Les coordonnées factorielles sur le 2 ème axe

par définition $\phi^2 = {}^t Z N v_2$

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 λ_1 de la matrice $Z Z^t N$.

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

Conclusion

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

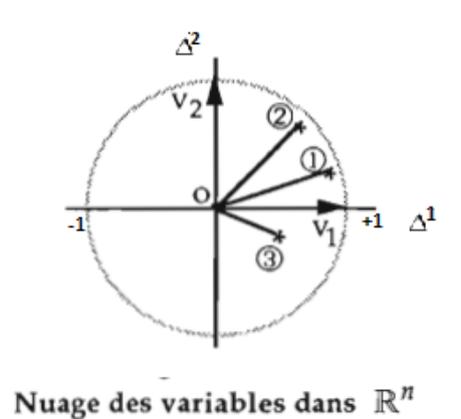
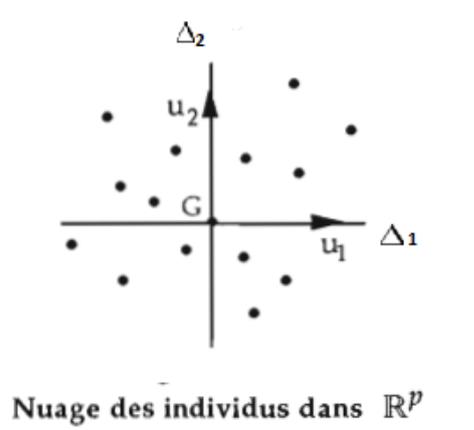
2.6 Représentations graphiques et l'interprétation

2.6.1 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-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 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

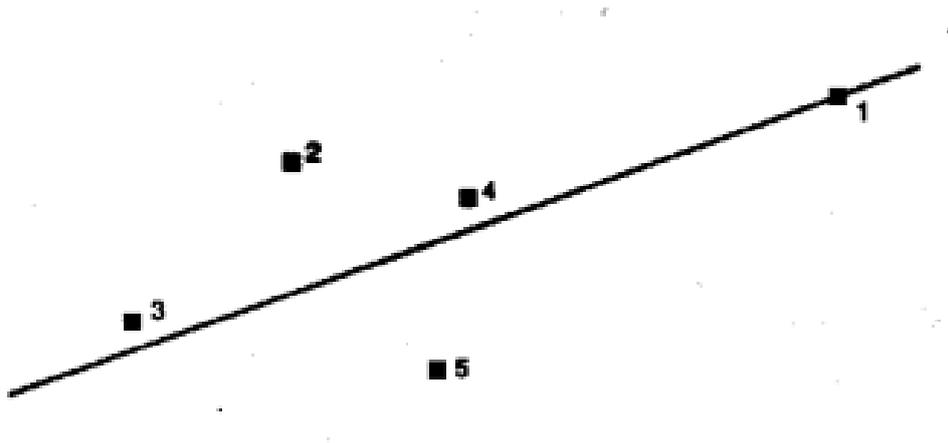


FIGURE 1.10 – L'axe d'un nuage de points

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 1.10 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 ω_i) à l'inertie de l'axe Δ_α . Celle-ci s'exprime par la formule :

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

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

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

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 varié 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 qualites de representation des individus :

L'exemple represente sur la figure 1.11 montre l'intérêt des qualités de représentation. Les deux points M_i et M_j ont des projections confondues sur l'axe Δ 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.

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

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

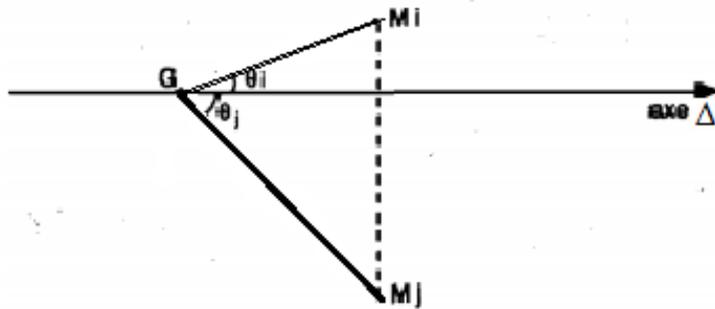


FIGURE 1.11 – Exemple de deux individus différentes ayant la même coordonnée sur Δ

où θ_i désigne l'angle entre le vecteur individu GM_i et l'axe Δ_α .

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 situé sur l'axe Δ_α lui-même,
- un cosinus de 0 pour un angle de 90° correspondant au cas où le vecteur individu est perpendiculaire à l'axe Δ_α .

Le développement de la formule permet d'en donner une interprétation en d'autres termes. Le cosinus d'un triangle rectangle est égal au rapport du côté adjacent sur l'hypothénuse. 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 $d^2(G, M_i)$ représente la distance qui sépare l'individu M_i du centre de gravité G du nuage. Etant donné que les axes factoriels sont orthogonaux, ce terme se décompose de la façon suivante :

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

ce résultat 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.

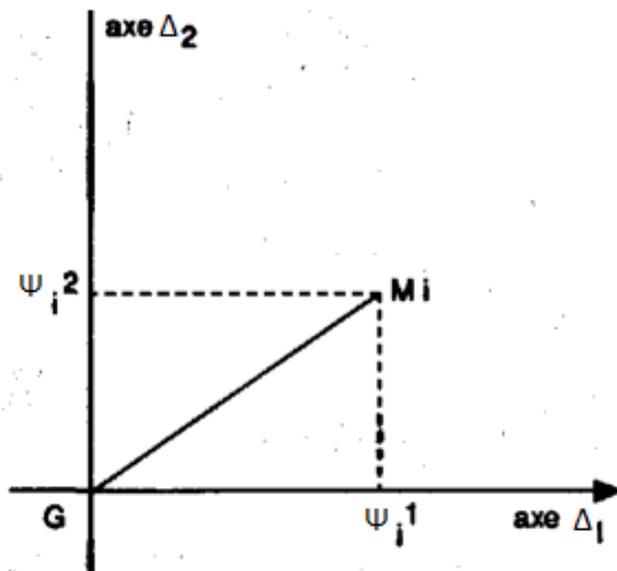


FIGURE 1.12 – Le théoreme de Pythagore : Le carré de l’hypathenuse (G, Mi) est égale la somme carrée des deux autres cotés

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

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

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

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 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 ϕ_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 $\alpha - \text{ème}$ axe factoriel, si la corrélation de X^j et ϕ_α 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 $\alpha - \text{ème}$ axe factoriel si le cosinus de l'angle formé par le vecteur X^j et celui de ϕ_α est proche de 0.

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 existe 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 variables 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.

Exemple 1.6.1. Nous supposons que les données du tableau suivant représentent deux variables avec six individus notées 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
σ	

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
σ	

les données centrées

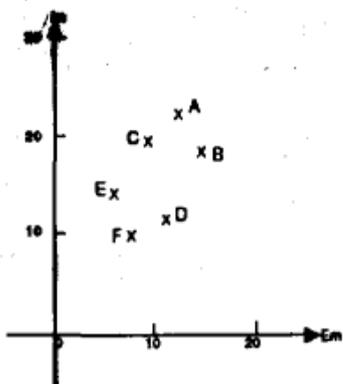
Nous effectuerons d'abord l'analyse en composantes principales de ce nuage étape par étape puis commenterons les résultats obtenus en montrant quelle structure l'analyse a mise en évidence.

A-) Centrage et réduction des variables : cette opération a eu deux conséquences :

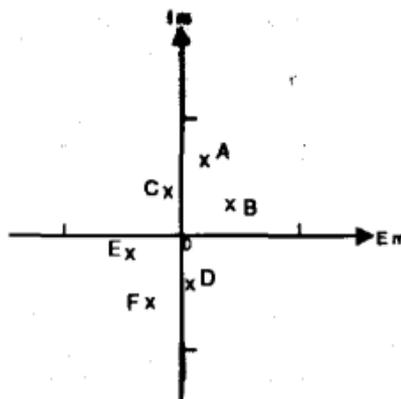
- L'origine est placée au centre de gravité du nuage,
- La forme du nuage de points est légèrement modifiée, les variables étant réduites, et elles ont donc toutes la même dispersion.

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).

$$V = {}^t Y N Y = \frac{1}{5} \begin{pmatrix} 16 & 8 & 6 \\ 8 & 14 & 8 \\ 6 & 8 & 16 \end{pmatrix} \implies \lambda_1 = 6 \quad \lambda_2 = 2 \quad \lambda_3 = 6/5. \implies \sum \lambda_i = \text{tr}(V)$$



Le nuage de points brut.



Le nuage de points centré.

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

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 ψ^α : Les nouvelles variables ψ^α nous sont données par la formule suivante :

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

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 $cov(\psi_1, \psi_2) = 0$

Exemple 1.6.2. On considère le tableau de données, noté X , qui est défini par : $X = \begin{pmatrix} 2 & 1 & -1 & 3 & -3 \\ 0 & 3 & -1 & 2 & -2 \\ 1 & 2 & -1 & 1 & -1 \end{pmatrix}$

ou la $i^{\text{ème}}$ ligne désigne l'individu x_i et la $j^{\text{ème}}$ colonne désigne la variable x^j . Chaque individu possède un poids égal à $1/3$. On considère les résultats de l'ACP du tableau X lorsque R^5 est muni de la métrique identité $M = I$.

1. Déterminer le tableau centré Y et le tableau centré réduit Z ? (03 individus et 05 variables.)

$$Y = \begin{pmatrix} 1 & -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

la matrice Z n'existe pas pourquoi ?

2. Sans calculer la matrice variance V du tableau X , combien existe-t-il d'axes factoriels ?

Le rang de Y est 2 puisque la colonne 1 est l'opposé de la colonne 2, la colonne 4 est l'opposé de la colonne 5 et la colonne 3 est nulle. Enfin la première et la cinquième colonne ne sont pas colinéaires. Donc le rang de Y est 2 donc celui de V aussi. Il y a donc deux axes factoriels.

3. Calculer V . En déduire l'inertie totale du nuage étudié.

$$V = \frac{1}{3}Y'Y = \frac{1}{3} \begin{pmatrix} 2 & -2 & 0 & 1 & -1 \\ -2 & 2 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 2 & -2 \\ -1 & 1 & 0 & -2 & 2 \end{pmatrix} \quad (1.5)$$

4. soit $U^t = (1 \ -1 \ 0 \ 1 \ -1)^t$

(a) Vérifier que U est un vecteur propre de V associé à la valeur propre 2.

On a

$$Vu = 2u.$$

(b) En déduire les valeurs propres et vecteurs propres de V .

Puisqu'il n'y a que deux axes factoriels non triviaux, il y a trois valeurs propres

nulle, 2 et une dernière valeur propre.

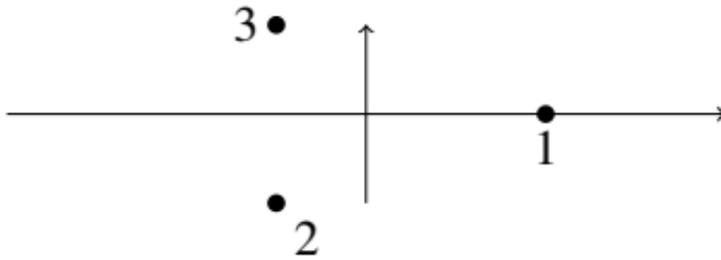
Or la trace vaut $8/3$ donc la dernière valeur propre est $2/3$.

5. Calculer les deux premières composantes principales, notée Ψ_1 et Ψ_2 .

les vecteurs propres : pour la valeur propre 2, on a $u_1 = \frac{1}{2}u$ et pour la valeur propre $2/3$, on a $u_2 = \frac{1}{2}(1 - 10 - 11)^t$; donc :

$$\Psi_1 = Yu_1 = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \text{ et } \Psi_2 = Yu_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}. \quad (1.6)$$

6. Représenter les trois individus dans le plan factoriel constitué des deux premiers axes.



7. Calculer la contribution relative de chaque individu à l'inertie du premier axe.

$$CTR_1(j_2) = \frac{1}{3} \frac{(\psi_1^{j_2})^2}{\lambda_1} \quad (1.7)$$

$$COR_1(j_2) = \frac{(\psi_1^{j_2})^2}{\|y^{j_2}\|^2} \quad (1.8)$$

La contribution de l'individu 1 à l'axe 1 est $2/3$, à l'axe 2 est 0.

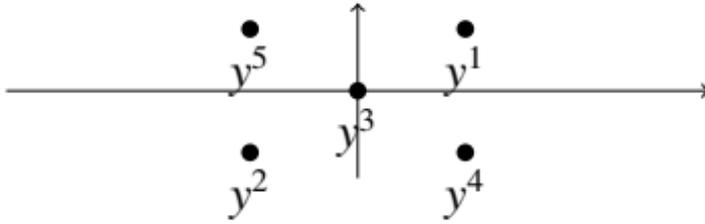
La contribution de l'individu 2 à l'axe 1 est $1/6$, à l'axe 2 est $1/2$.

La contribution de l'individu 3 à l'axe 1 est $1/6$, à l'axe 2 est $1/2$.

8. Représenter les 5 variables dans le plan factoriel constitué des deux premiers axes.

2.7 Exercice

Exercice 2.7.1. Six points du plan ont pour coordonnées, pour $\alpha > 0$ et $\beta > 0$ vérifiant $\alpha^2 + \beta^2 = 1$:



point	C_1	C_2
1	1	1
2	-1	-1
3	-1	1
4	1	-1
5	α	β
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 α et β .

Solution :

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

i- la matrice de variance-covariance empirique (S). Le tableau de données X est centré, ainsi

$Y = X$ et en posant $n = 6$ donc

$$(S) = \frac{1}{n} Y \cdot Y = \frac{1}{3} \begin{pmatrix} 2 + \alpha^2 & \alpha\beta \\ \alpha\beta & 2 + \beta^2 \end{pmatrix} = \frac{1}{3} \cdot A$$

ii- les valeurs propres de la matrice (S). En utilisant la relation $\alpha^2 + \beta^2 = 1$, et on cherche les valeurs propres de A qui sont 3 et 2. Les valeurs propres de (S) sont donc 1 et $2/3$.

iii- l'inertie du nuage de points associée au tableau de données.

D'après le cours, l'inertie du nuage de points est égale à la trace de (S) qui vaut $1/3 \times 5 = 5/3$

iv- Calculer les vecteurs propres de (S) ?. Un vecteur propre $V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ de A correspondant à la valeur propre 2 est solution de :

$$(2 + \alpha^2)v_1 + \alpha\beta v_2 = 2v_1$$

$$\alpha\beta v_1 + (2 + \beta^2)v_2 = 2v_2$$

$\Leftrightarrow \alpha v_1 + \beta v_2 = 0$ ainsi $V_2 = \begin{pmatrix} \beta \\ -\alpha \end{pmatrix}$ est vecteur propre de norme 1 pour la valeur propre 2, et, par orthogonalité des vecteurs propres correspondant à des valeurs propres différentes lorsqu'on diagonalise une matrice symétrique, le vecteur $V_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ est vecteur propre de norme 1 pour la valeur propre 3.

v- les composantes principales de l'ACP sur (S) .

Avec les notations du cours, les composantes principales de l'ACP sur (S) sont $\psi_1 = YV_1 = (\alpha + \beta - (\alpha + \beta) - \alpha + \beta \alpha - \beta 1 - 1)^t$ et $\psi_2 = YV_2 = (-\alpha + \beta(\alpha - \beta) - (\alpha + \beta)\alpha + \beta 00)^t$

vi- Représenter les points et les axes principaux dans le plan de départ. les quatre premiers points du nuages sont sur le carré à côtés parallèles aux axes, de longueur 2 et centré en 0, et les deux derniers points sont sur le cercle inscrit, de rayon 1 et centré en 0. Le premier axe principal passe par le centre du cercle et le point $(\alpha \ \beta)^t$.

vii- Reprendre i) et ii) quand on n'observe que les points 1, 2, 3, 4. Le tableau de données X est toujours centré, ainsi $Y = X$ et Y est cette fois égale à l'identité en dimension 2. Il y a donc une valeur propre double égale à 1.

viii- déduire le résultat de iv) . Tous les axes orthonormés sont axes principaux du nuage des quatre premiers points. De plus les deux derniers points sont alignés avec le centre du nuage de points. Ainsi ce qui détermine le premier axe principal du nuage des six points, c'est la direction engendré par ces deux derniers point et $V_1 = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ doit être vecteur propre.

Exercice 2.7.2. 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>Qualité</i>	<i>Prix</i>
<i>E1</i>	-2	3	-1
<i>E2</i>	-1	1	0
<i>E3</i>	2	-1	-1
<i>E4</i>	1	-3	2

1. Calculer le vecteur moyen des données. Qu'en conclure ?
2. Calculer la covariance entre x^1 et x^2 . 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 λ_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.

Solution :

1. le vecteur moyen des individus. les données sont centrés
2. Si on veut faire une ACP centrée avec des poids uniformes. il faut travailler avec la matrice des données centrés et la matrice de variance covariance ?
3. Si On veut faire une ACP centrée réduite avec des poids uniformes. il faut travailler avec la matrice des données centrés réduite et la matrice de corrélation

4. la matrice de variance-covariance et de corrélation. $S = 1/6({}^tY.Y) =$

$$\begin{pmatrix} & \text{service} & \text{qualité} & \text{prix} \\ \text{service} & 2,5 & & \\ \text{qualité} & -3 & 5 & \\ \text{prix} & 0,5 & -2 & 1,5 \end{pmatrix}$$

$$R = 1/6({}^tZ.Z) = \begin{pmatrix} & \text{service} & \text{qualité} & \text{prix} \\ \text{service} & 1 & & \\ \text{qualité} & -0,848 & 1 & \\ \text{prix} & 0,258 & -0,730 & 1 \end{pmatrix}$$

5. pour ACP centrée réduite On donne ($\lambda_1 = 1,639$; $\lambda_3 = 0,106$) En déduire λ_2 . D'après le cours la somme des valeurs propres est égale au nombre de variables donc $\lambda_2 = 3 - 0,106 - 1,639 = 1,25$

6. Calculer les pourcentages d'inertie des axes.

valeurs propre	% d'inertie	% d'inertie cumulé
1,639	0,546333333	0,546333333
1,25	0,416666667	0,963
0,106	0,035333333	0,998333333

d'après le critère de pourcentage d'inertie on a retenu deux axes pour la représentation .

7. avec ACP centrée réduite on a les vecteurs propres $(V_1)^t = (0,531, -0,761, -0,371)^t$ et $(V_2)^t = (0,633, 0,065, 0,770)^t$

d'après le cours on trouve les composantes des individus avec la relation $\psi_i = Z.V_i$ et les composantes des variables avec la relation $\phi_i = \sqrt{\lambda_i}.\psi_i$

8. la Représentation graphique des individus et les variables dans le plan principal(1,2).

Exercice 2.7.3. 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 cumulée .
- Combien d'axes doit être retenu 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. ?

Solution :

1. S est la matrice de covariance donc les coefficients a, b, c signifient les covariances deux à deux entre les variables. on a $-1 \leq \rho(\xi_i, \xi_j) \leq 1$ et $\rho(\xi_i, \xi_j) = \frac{\text{cov}(\xi_i, \xi_j)}{\sqrt{\text{Var}(\xi_i)} \cdot \sqrt{\text{Var}(\xi_j)}} = \text{cov}(\xi_i, \xi_j) = a, b, c$ donc $-1 \leq a, b, c \leq 1$

2. OUI on peut supposer que S est une matrice de corrélation puisque tout les coefficient entre -1 et 1 .

3. si $a=b=c=0$ donc S est une matrice diagonale et le déterminant de $|S - \lambda \cdot I_4| =$ le produit des éléments diagonaux donc on a quatre valeurs propres telle que $\lambda_i = 1_{\{i=1, \dots, 4\}}$

4. $a = 0,1 ; b = 0,4 ; c = 0,6$. si on suppose que S est une matrice de corrélation .

$\rho(\xi_1, \xi_2) = 0.1$ très faible corrélation $\rho(\xi_1, \xi_4) = 0.6$ il \exists une relation de sens positive

$\rho(\xi_1, \xi_3) = 0.4$ faible corrélation $\rho(\xi_2, \xi_3) = 0.6$ il \exists une relation de sens positive

$\rho(\xi_2, \xi_4) = 0.4$ faible corrélation $\rho(\xi_3, \xi_4) = 0.1$ très faible corrélation

on utilise la matrice corrélation dans le cadre de l'ACP normée puisque on a réduit toutes les variables initiales on a des variables avec même variance qui est égale à un.

a- $\lambda_4 = 4 - \sum \lambda_i = 0.001$ la somme des valeurs propres=4 cette valeur (la somme) représente l'inertie totale de nuage de points.

b- OUI les vecteurs propres sont orthonormés par rapport à la métrique identité puisque ${}^t V_i \cdot V_i = 1_{\{i=1, \dots, 4\}}$.

c- les pourcentages d'inertie pour chacun des axes et les pourcentages d'inertie cumulés .

	λ_i	% inertie	% inertie cumulés
axe1	2,83	70,75%	70,75%
axe2	1,146	28,65%	99,40%
axe3	0,023	0,58%	99,98%
axe4	0,001	0,03%	100,00%

d- deux axes retenus pour la représentation (% inertie axe1 + % inertie axe2 = 99,40%) et $(\lambda_1, \lambda_2 > 1)$.

e- la formule permettant de calculer les composantes principales associées au nuage des variables est $\phi_i = \sqrt{\lambda_i} \cdot V_i$ est la matrice

$$\phi = \begin{pmatrix} \xi_1 & -0,841 & 0,535 & -0,076 & 0,016 \\ \xi_2 & -0,841 & -0,535 & 0,076 & 0,016 \\ \xi_3 & -0,841 & -0,535 & -0,076 & -0,016 \\ \xi_4 & -0,841 & 0,535 & 0,076 & -0,016 \end{pmatrix}$$

f- Représentation graphique du nuage de points variables et l'interprétation.

5. la matrice inverse de la matrice des vecteur propres V est tV puisque on a la relation $A^{-1}.A = I$ équivalent a ${}^tV.V = I_{\text{identité}}$ puisque les vecteurs propres sont orthonormé .

Exercice 2.7.4. Soit un vecteur aléatoire $x = (\xi_1, \xi_2, \xi_3)^t$ de moyenne 0 et de matrice de covariance Σ .dont les variables ont pour matrice

$$\Sigma = \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix}$$

où $\rho > 0$ est une valeur donnée.

1. Chercher les valeurs de ρ telle que Σ soit bien une matrice de covariance
2. Déterminer les valeurs propres de Σ .
3. Chercher la plus grande valeur ρ_1 telle que les valeurs propres de Σ soit positives .
4. calculer la somme des valeurs propres ? que signifie cette somme ?
5. Vérifier que Σ admet pour vecteur propre $v_1 = \frac{1}{2}(1, \sqrt{2}, 1)^t$. associe au plus grand valeur propre de Σ .
6. soit $v_2 = \frac{1}{2}(\sqrt{2}, 0, -\sqrt{2})^t$. , $v_3 = \frac{1}{2}(1, -\sqrt{2}, 1)^t$. les deux autre vecteurs propre de Σ .
Déterminer les composantes principales des variables ϕ^j .
7. Combien d'axes doit être retenir pour la représentation ? Justifier votre réponse.
8. Faire la projection des variables sur le disque des corrélations dans les deux premiers axes pour $\rho = \rho_1$. Commenter le graphique obtenu .

Solution :

1. Si Σ une matrice de covariance alors les éléments diagonaux sont les valeurs des variances des variables les les hors diagonales sont les covariance deux a deux et d'autre part on le coefficient de corrélation entre deux variable qui est égale $\frac{\text{cov}(\xi_1, \xi_2)}{\sigma_{\xi_1} \cdot \sigma_{\xi_2}}$ donc en trouve que $-1 \leq \rho \leq 1$ et puisque $\rho > 0$ on a $0 < \rho \leq 1$.

2. les valeurs propres de Σ .

d'après le polynôme caractéristique on $P(\Sigma - \lambda I_3) = 0$ et on la classé par ordre croissant on trouve

$$\boxed{\lambda_1 = 1 + \rho\sqrt{2} \quad \lambda_2 = 1 \quad \lambda_3 = 1 - \rho\sqrt{2}}$$

3. la plus grande valeur ρ_1 telle que les valeurs propres de Σ soit positives

On a $\rho > 0$ donc λ_1, λ_2 sont positives reste $\lambda_3 = 1 - \rho\sqrt{2} \geq 0$ donc $\rho_1 = \frac{1}{\sqrt{2}}$

4. La somme des valeurs propres $\lambda_1 + \lambda_2 + \lambda_3 = 3$

Cette somme signifie l'inertie total.

5. Vérifiant que Σ admet pour vecteur propre $v_1 = \frac{1}{2}(1, \sqrt{2}, 1)^t$. associe au plus grand valeur propre de Σ . pour que v_1 soit vecteur propre on doit vérifier la relation $\Sigma v_1 = \lambda_1 v_1$ donc

$$\Sigma \cdot v_1 = \frac{1}{2} \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & \rho \\ 0 & \rho & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 + \sqrt{2}\rho \\ 2\rho + \sqrt{2} \\ 1 + \sqrt{2}\rho \end{pmatrix} = (1 + \sqrt{2}\rho) \frac{1}{2} \begin{pmatrix} 1 \\ \sqrt{2} \\ 1 \end{pmatrix} = \lambda_1 v_1$$

6. deux d'axes retenus pour la représentation ($\lambda_1, \lambda_2 \geq 1$).

7. soit $v_2 = \frac{1}{2}(\sqrt{2}, 0, -\sqrt{2})^t$, $v_3 = \frac{1}{2}(1, -\sqrt{2}, 1)^t$. les deux autre vecteurs propre de Σ .

Les composantes principales des variables ϕ^j .

D'après la formule de passage on a $\phi^j = \sqrt{\lambda_j} \cdot v_j$ on a

$$V^j = \frac{1}{2} \begin{pmatrix} v_1 & v_2 & v_3 \\ 1 & \sqrt{2} & 1 \\ \sqrt{2} & 0 & -\sqrt{2} \\ 1 & -\sqrt{2} & 1 \end{pmatrix} \quad \phi^j = \frac{1}{2} \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ \sqrt{1 + \rho\sqrt{2}} & \sqrt{2} & \sqrt{1 - \rho\sqrt{2}} \\ \sqrt{2}\sqrt{1 + \rho\sqrt{2}} & 0 & -\sqrt{2}\sqrt{1 - \rho\sqrt{2}} \\ \sqrt{1 + \rho\sqrt{2}} & -\sqrt{2} & \sqrt{1 - \rho\sqrt{2}} \end{pmatrix}$$

8. pour $\rho = \rho_1 = \frac{1}{\sqrt{2}}$. la projection des variables sur le disque des corrélations dans les deux

$$\text{premiers axes} \cdot \phi^j = \frac{1}{2} \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 \\ \sqrt{2} & \sqrt{2} & 0 \\ 2 & 0 & 0 \\ \sqrt{2} & -\sqrt{2} & 0 \end{pmatrix}$$

Représentation graphiquement du nuage de points variables dans le plan principal et l'interprétation. Commenter le graphique obtenu.

Exercice 2.7.5. On considère le tableau de données, noté X , qui est défini par :

$$X = \begin{array}{c|ccccc} & j_1 & j_2 & j_3 & j_4 & j_5 \\ \hline i_1 & 1 & 2 & 3 & 4 & 5 \\ i_2 & 6 & 7 & 8 & 9 & 10 \\ i_3 & 11 & 12 & 13 & 14 & 15 \\ i_4 & 16 & 17 & 18 & 19 & 20 \\ i_5 & 21 & 22 & 23 & 24 & 25 \end{array} \quad (1.9)$$

ou la $i^{\text{ème}}$ ligne désigne la variable x_i et la $j^{\text{ème}}$ colonne désigne l'individu x^j . Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X (i.e., ACP de X lorsque \mathbb{R}^5 est muni de la métrique identité et chaque individu possède un poids égal à $1/5$).

1. Déterminer g le centre de gravité du nuage des cinq individus associés au tableau X . qu'est ce vous remarquer ?

$$\text{le centre de gravité du nuage } g = \frac{1}{5}(15406590115)' = (38131823)' = x^3$$

2. Calculer le tableau centré, noté Y , qui est associé à X .

Toutes les lignes de Y étant égales $(-2 \ -1 \ 0 \ 1 \ 2)$

3. Déterminer l'inertie totale du nuage étudié.

$$\text{On a } V_X = V_Y = \frac{1}{5}Y'Y = \frac{10}{5}K_5'K_5 = 2K_5'K_5. \text{ D'où } I_T = \text{tr}(V_Y) = 10$$

$$\text{On a aussi : } I_T = \sum_{j \in J} p_j d^2(x^j, g) = \frac{1}{5} \sum_{j \in J} p_j \|x^j - g\|^2 = \frac{2}{5} (5 \times (-2)^2 + 5 \times (-1)^2 + 5 \times 0^2) = 10$$

4. Calculer la première composante principale de l'individu j_5 , notée $\Psi_1^{j_5}$.

$$\Psi_1^{j_5} = (y^{j_5})' u = \frac{1}{\sqrt{5}}(2 + 2 + 2 + 2 + 2) = \frac{10}{\sqrt{5}} \approx 4.472$$

5. Calculer la contribution de l'individu j_5 .

$$\text{CTR}_1(j_5) = \frac{1}{5} \frac{(\Psi_1^{j_5})^2}{\lambda} = \frac{1}{5} \frac{100}{10} = \frac{2}{5} = 0.4$$

6. Déterminer la représentation des variables dans le nouveau système d'axes factoriels de l'espace des variables.

Il n'existe qu'un seul "axe factoriel" pour représenter les variables. Les coordonnées des variables sur cet axe sont celles du vecteur $\phi = \sqrt{\lambda}u$. Donc, sur cet unique axe, chaque variable a la même coordonnée qui vaut $\frac{\sqrt{10}}{\sqrt{5}} = \sqrt{2} = 1,414$

Chapitre 3

Analyse Factorielle des Correspondances

3.1 Introduction

Dans le chapitre précédent, on a expliqué la méthode d'ACP. L'objet de ce chapitre concerne des variables non plus quantitatives mais qualitatives. Nous nous limiterons dans ce chapitre au cas de deux variables (Tableau de contingence), que l'on notera X_1 et X_2 , ayant chacune un nombre respectif de modalités égal à n_1 et n_2 , avec $n = n_{..} = n_1 + n_2 = n$ la taille de la population étudiée.

Les objectifs sont les mêmes que ceux de l'ACP dans le sens où l'AFC cherche donc à obtenir une typologie des lignes et une typologie des colonnes, puis de relier ces deux typologies. Il faut donc faire ressortir un bilan des ressemblances entre lignes (respectivement les colonnes).

3.2 Quelques définitions

Définition 2.2.1. Table de contingence - marges :

Un tableau de contingence est un tableau d'effectifs obtenus en croisant les modalités de deux variables qualitatives définies sur une même population de n individus.

Les données obtenues lors de l'étude d'une population à l'aide de deux variables qualitatives sont représentées sous la forme d'un tableau à n_1 lignes et n_2 colonnes (ou inversement), chaque ligne correspondant à une modalité d'une des deux variables et chaque colonne à une modalité de l'autre variable. Chaque case (i, j) de ce tableau comporte le nombre d'individus possédant la modalité i

pour la première variable et la modalité j pour la seconde :

$$T = (n_{ij})_{1 \leq i \leq n_1, 1 \leq j \leq n_2}$$

Les n_{ij} sont appelés **les effectifs dits conjoints**.

On note que les tables de contingence sont le résultat de ce qu'appellent les praticiens des enquêtes, les tables croisés.

La somme des effectifs des lignes et des colonnes, notées respectivement $n_{.j}$ et $n_{i.}$ sont appelées les marges en colonnes et les marges en lignes. Elles sont représentées dans le tableau ci-dessus .

Nous avons

$$\sum_i n_{i.} = \sum_j n_{.j} = n_{..} = n$$

$$T = \begin{array}{c|cccc|c} & 1 & \dots & j & \dots & n_2 & n_{i.} \\ \hline 1 & & & & & & . \\ \vdots & & & \vdots & & & . \\ i & & \dots & x_{ij} & \dots & & . \\ \vdots & & & \vdots & & & . \\ n_1 & & & & & & . \\ \hline n_{.j} & & & . & & & n = n_{..} \end{array}$$

Exemple 2.2.1. Le tableau suivant représente le type d'études poursuivies (université, classes préparatoires, autres) en fonction du parcours suivi au lycée (Lettres, Économie, Maths-Sciences, Technique).

	Univ	Prepa	Autres	$n_{i.}$
Lettres	13.00	2.00	5.00	20
Economie	20.00	2.00	8.00	30
Math-Sciences	10.00	5.00	5.00	20
Tech	7.00	1.00	22.00	30
$n_{.j}$	50	10	40	100

Définition 2.2.2. Tableau des fréquences :

c'est le tableau des fréquences relatives qui est considéré ou bien les fréquences conjointes sont données par :

$$f_{ij} = \frac{n_{ij}}{n_{..}}, \quad f_{i.} = \frac{n_{i.}}{n_{..}} \quad \text{et} \quad f_{.j} = \frac{n_{.j}}{n_{..}}$$

les fréquences conjointes et fréquences marginales.

	1	...	j	...	n_2	$f_{i.}$
1						.
\vdots						\vdots
i	...	$f_{ij} = \frac{n_{ij}}{n_{i.}}$
\vdots						\vdots
n_1						.
$f_{.j}$						1

Nous avons ainsi :

$$\sum_i f_{i.} = \sum_j f_{.j} = \sum_j \sum_i f_{ij} = 1$$

Exemple 2.2.2. Le tableau de l'exemple précédent.

	Univ	Prepa	Autres	$f_{i.}$
Lettres	13/100	2/100	5/100	20/100
Economie	20/100	2/100	8/100	30/100
Math-Sciences	10/100	5/100	5/100	20/100
Tech	7/100	1/100	22/100	30/100
$f_{.j}$	50/100	10/100	40/100	1

Définition 2.2.3. Tableau des profils-lignes et colonnes :

on appelle tableau des profils-lignes(matrice des profils-lignes) ,noté PL où L , le tableau correspondant aux fréquences conditionnelles

$$(L)_{ij} = n_{ij}/n_{i.} = f_{ij}/f_{i.}$$

et de même pour le tableau des profils colonnes, (PC) de terme générale ,

$$(C)_{ij} = n_{ij}/n_{.j} = f_{ij}/f_{.j}$$

On notera $L_i(\ell_i)$ le i -ème profil ligne (la i -ème ligne de de tableau PL)

$$\ell_i = (n_{i1}/n_{i.} \dots \dots n_{in_2}/n_{i.}) = (f_{i1}/f_{i.} \dots \dots f_{in_2}/f_{i.})$$

et C_j le j -ème profil colonne (la j -ème colonne de tableau PC)

$$C_j = (n_{1j}/n_{.j} \dots \dots n_{n_1j}/n_{.j}) = (f_{1j}/f_{.j} \dots \dots f_{n_1j}/f_{.j})$$

écriture marcielle : En notant $F = T/n$, la matrice des fréquences conjointes et $D_1 = D_r$ et $D_2 = D_c$ les matrices diagonales suivantes de taille respective $n_1 \times n_1$ et $n_2 \times n_2$.

$$D_1 = D_r = \begin{pmatrix} f_{i.} & & \\ & \ddots & \\ & & f_{n_1.} \end{pmatrix}, \quad D_2 = D_c = \begin{pmatrix} f_{.1} & & \\ & \ddots & \\ & & f_{.n_2} \end{pmatrix}$$

On a $L = D_1^{-1}.F$ et $C = D_2^{-1}.F'$.

où, F' correspond à la transposée de F .

3.3 AFC et indépendance

On a vu que l'AFC applicable sur un tableau de contingence ou de fréquence pour étudier les liaisons entre les deux variables qualitatives du tableau. Nous ne pouvons plus définir les liaisons (dépendance) par les coefficients de corrélation comme l'ACP.

Définition 2.3.1. Il y a indépendance entre les deux variables considérées si :

$$f_{ij} = f_{i.} \times f_{.j} \quad \forall 1 \leq i \leq n_1, \quad \forall 1 \leq j \leq n_2;$$

Nous disons qu'il y a liaison (dépendance) entre ces deux variables, ou que ces deux variables sont liées si elles ne sont pas indépendantes.

Remarque 2.3.1. Dans le cas de l'indépendance tous les profils lignes et colonnes sont identiques.

3.4 la distance du Khi-deux

On est en présence de deux nuages de points pesants. Le nuage des (n_1) profils lignes, chaque profil ligne i est muni du poids $f_{i.}$. Le nuage n_2 des profils colonnes, chaque profil colonne j est muni de son poids $f_{.j}$. La question posée comment mesurer les distances Euclidiennes entre les profils ? Le type de distance utilisé est la distance du du Khi-deux χ^2 qui est une distance Euclidienne.

— Espace des profils lignes, $(\mathbb{R}^{n_2}, D_c^{-1})$: Soient $l_i, l_{i'}$ deux lignes de tableau L :

$$d_{\chi^2}^2(l_i, l_{i'}) = \sum_{j=1}^{n_2} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}} \right)^2$$

— Espace des profils lignes, $(\mathbb{R}^{n_1}, D_r^{-1})$: Soient $C_j, C_{j'}$ deux lignes de tableau C :

$$d_{\chi^2}^2(C_j, C_{j'}) = \sum_{i=1}^{n_1} \frac{1}{f_{i.}} \left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}} \right)^2$$

La raison essentielle pour choisir la distance du χ^2 est qu'elle vérifie la propriété suivante :

Propriété 2.4.1. équivalence distributionnelle : Lorsque l'on effectue dans le tableau des fréquences F , la somme de deux colonnes proportionnelles (ou de deux lignes), les distances du χ^2 entre profils lignes (ou colonnes) restent inchangées.

Lien avec le χ^2 de tableau de contingence :

$$\frac{d^2}{n} = \sum_{i=1}^{n_1} f_{i.} d_{\chi^2}^2(L_i, \bar{L}) = \sum_{j=1}^{n_2} f_{.j} d_{\chi^2}^2(C_j, \bar{C})$$

Le coefficient $\frac{d^2}{n}$ est égal à l'inertie du nuage des profils lignes (des profils colonnes).

Démonstration.

$$\frac{d^2}{n} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \frac{1}{f_{i.} f_{.j}} (f_{ij} - f_{i.} f_{.j})^2 = \sum_{i=1}^{n_1} f_{i.} \sum_{j=1}^{n_2} \frac{1}{f_{.j}} \left(\frac{f_{ij}}{f_{i.}} - f_{.j} \right)^2$$

□

La distance euclidienne entre les profils-lignes ou entre les profils-colonnes permet de bien modéliser les ressemblances entre deux modalités.

Résumé	Points lignes	Points colonnes
matrice	$L = D_1^{-1} F$	$L = F D_2^{-1}$
nuages	$\mathbb{N}_L = L_1, \dots, L_{n_1}$	$\mathbb{N}_C = C_1, \dots, C_{n_2}$
poids	D_r	D_c
espace	$(\mathbb{R}^{n_2}, D_c^{-1})$	$(\mathbb{R}^{n_1}, D_r^{-1})$
point moyen	$1'_{n_1} F$	$F 1'_{n_2}$
inertie	$\frac{d^2}{n}$	$\frac{d^2}{n}$

3.5 Les nuages des deux profils

3.5.1 Le nuage des profils-lignes

Lorsque nous nous intéressons aux modalités de la première variable, il faut considérer les données comme une juxtaposition de profils-lignes. Ainsi chaque profil-ligne i peut être représenté comme un point de l'espace $(\mathbb{R}^{n_2}, D_c^{-1})$ dont chacune des n_2 dimensions représente une modalité de la seconde variable. L'utilisation de la distance entre deux profils est celle χ^2 , elle revient à affecter le poids $\frac{1}{f_j}$ à la $j^{\text{ème}}$ dimension de \mathbb{R}^{n_2} . Du fait que la somme de chaque profil-ligne est égale à 1, le nuage N_{n-1} appartient à un hyperplan, noté H_I . Pour l'AFC les poids affectés à chaque point du nuage sont imposés et ne sont pas identiques. Le point i a pour poids la fréquence marginale f_i . Ce poids est naturel puisqu'il est proportionnel à l'effectif de la classe d'individus qu'il représente. La coordonnée du point i sur l'axe Δ est donnée par $\frac{f_{ij}}{f_i}$.

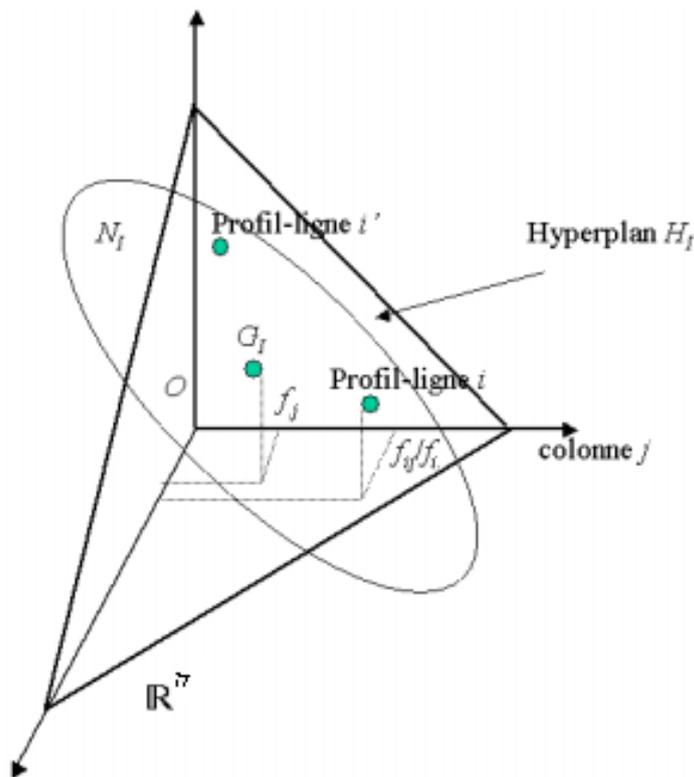


FIGURE 2.1 – Le nuage N_I des profils-lignes dans \mathbb{R}^{n_2}

Le point i a pour coordonnée sur l'axe Δ : $\frac{f_{ij}}{f_i}$; son poids est f_i ; la distance entre deux

profils est la distance du χ^2 , Le barycentre G_I du nuage N_I a pour coordonnée sur l'axe Δ la fréquence marginale $f_{.j}$; le nuage N_I appartient à un hyperplan noté H_I .

Le barycentre des points de N_I munis de ces poids, noté G_I , est la moyenne pondérée de tous les points sur tous les axes δ_j . La coordonnée de G_I sur l'axe δ_j est donc donnée par :

$$f_{.j} = \sum_i f_i \frac{f_{ij}}{f_i}$$

Le barycentre s'interprète comme un profil-moyen.

3.5.2 Le nuage des profils-colonnes

La construction du nuage des profils-colonnes est identique à celle du nuage des profils lignes du fait de la symétrie entre les lignes et les colonnes en AFC. Chaque profil-colonne j peut être représenté comme un point de l'espace \mathbb{R}^{n_1} dont chacune des n_1 dimensions représente une modalité de la première variable (i). Le point i a pour coordonnée sur l'axe Δ la proportion $\frac{f_{ij}}{f_{.j}}$, et le poids qui lui est associé est $f_{.j}$. Le nuage N_J appartient à un hyperplan noté H_J . De plus le barycentre des points de N_J munis de leur poids a pour coordonnée sur l'axe δ_i :

$$f_i = \sum_j f_{.j} \frac{f_{ij}}{f_{.j}}$$

Ce barycentre s'interprète également comme un profil moyen et joue le même rôle pour l'étude de la liaison entre les deux variables.

3.6 Schéma général de l'AFC

Nous résumons les principaux résultats de l'AFC dans un schéma général

Les deux nuages de points (dans l'espace des colonnes et dans l'espace des lignes) sont construits de manière analogue. Nous récapitulons ici les éléments de base de l'analyse qui vont permettre la construction des facteurs.

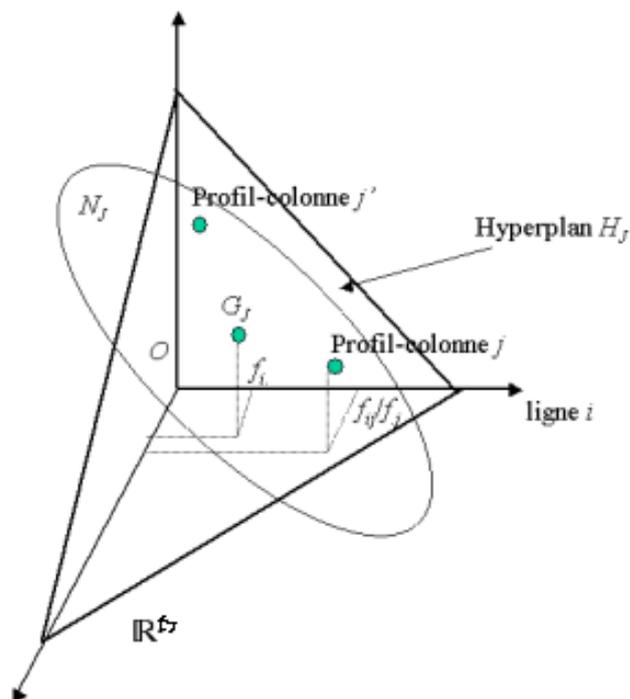


FIGURE 2.2 – Le nuage N_J des profils-colonnes dans l'espace \mathbb{R}^{n_1}

Nuage de n_1 points-lignes dans l'espace \mathbb{R}^{n_2}	Éléments de bases	Nuage de n_2 points-colonnes dans l'espace \mathbb{R}^{n_1}
$X = D_r^{-1}F$ n_1 coordonnées (point-ligne i) $\frac{f_{ij}}{f_i}; i = 1, \dots, n_1$	Analyse de tableau X	$X = D_c^{-1}F'$ n_2 coordonnées (point-colonne j) $\frac{f_{ij}}{f_j}; j = 1, \dots, n_2$
$M = D_c^{-1}$ $d^2(i, i') = \sum_{j=1}^{n_2} \frac{1}{f_j} \left(\frac{f_{ij}}{f_i} - \frac{f_{i'j}}{f_{i'}} \right)$	avec la metrique M	$M = D_r^{-1}$ $d^2(j, j') = \sum_{i=1}^{n_1} \frac{1}{f_i} \left(\frac{f_{ij}}{f_j} - \frac{f_{ij'}}{f_{j'}} \right)$

3.6.1 Critère à maximiser et matrice à diagonaliser

Nous voulons représenter graphiquement les proximités entre profils.

Nous avons les deux espaces, avec les centres de gravité des nuages. nous procédons l'analyse par rapport à l'origine ou par rapport aux centres de gravité, à condition de négliger dans le premier cas l'axe factoriel qui joint l'origine au centre de gravité (cet axe est associé à la valeur propre égale à 1, appelée valeur propre triviale).

Plaçons-nous dans l'espace des colonnes \mathbb{R}^{n_2} et cherchons l'axe d'inertie maximum du

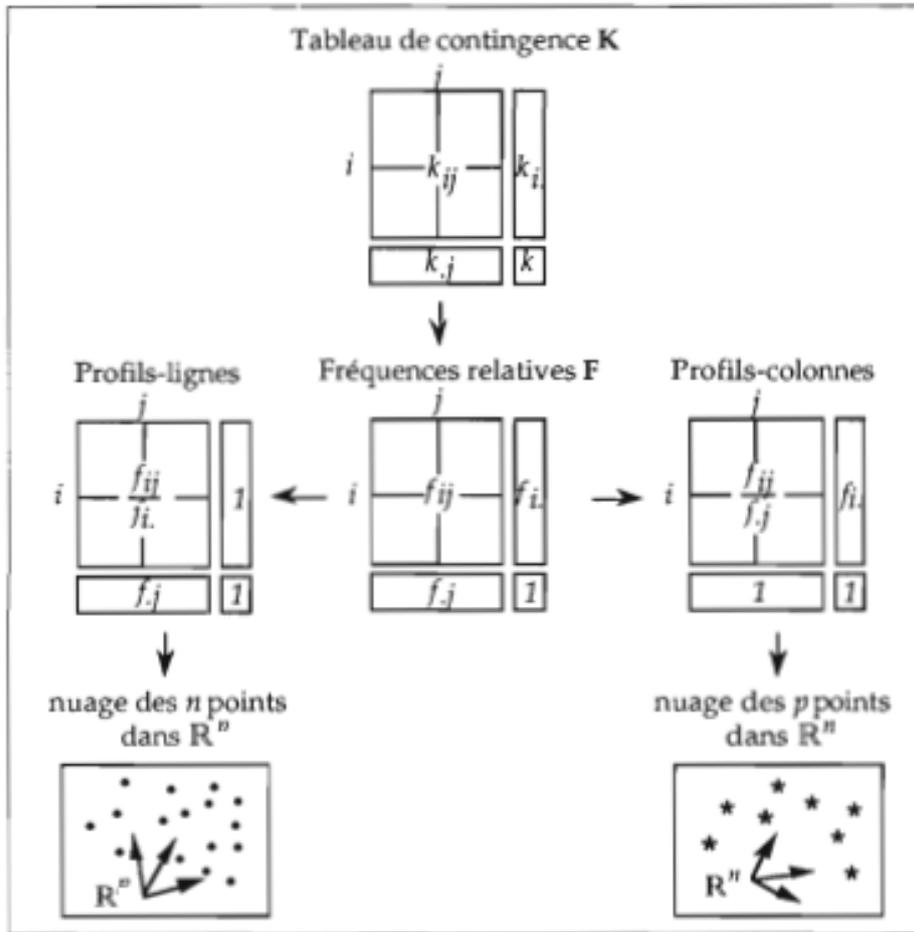


FIGURE 2.3 – Schéma général de l'afc

nuage des points-lignes passant par l'origine o et engendré par un vecteur-unitaire u pour la métrique $M = D_c^{-1}$. Ceci nous amène à maximiser la somme pondérée des carrés des projections sur l'axe c'est-à-dire :

$$\max_u = \left\{ \sum_i f_i \cdot d^2(i, O) \right\}$$

on rendre maximale la quantité : $u' D_c^{-1} F' D_r^{-1} F D_c^{-1} u$

avec la contrainte : $\mu' D_c^{-1} u = 1$ et u est un vecteur propre de la matrice :

$$S = F' D_r^{-1} F D_c^{-1}$$

associé à la plus grande valeur propre λ , différente de 1.

La matrice à diagonaliser est donc la matrice **S** de terme général :

$$s_{jj'} = \sum_i \frac{f_{ij}f_{ij'}}{f_{i.}f_{.j'}}$$

De la même façon, on doit rendre maximum dans \mathbb{R}^{n_1} , la quantité :

$$v' D_r^{-1} F D_c^{-1} F' D_r^{-1} v$$

avec la contrainte : $v' D_r^{-1} v = 1$. et **v** est vecteur propre de la matrice :

$$T = F D_c^{-1} F' D_r^{-1}$$

3.6.2 Les axes factoriels et facteurs

Nous supposons ici que $n_2 < n_1$ correspond à la plus petite dimension du tableau de données.

Après avoir écarté la valeur propre triviale égale à 1 et le vecteur propre associé, nous retenons, de la diagonalisation de la matrice, les $n_2 - l$ valeurs propres non nulles et les vecteurs propres associés. Nous obtenons ainsi au plus $n_2 - l$ axes factoriels.

Dans \mathbb{R}^{n_2}	Éléments de construction	Dans \mathbb{R}^{n_1}
$S = F' D_r^{-1} F D_c^{-1}$	Matrice à diagonaliser	$T = F D_c^{-1} F' D_r^{-1}$
$S u_\alpha = \lambda_\alpha u_\alpha$	Axe factoriel	$S v_\alpha = \lambda_\alpha v_\alpha$
$\psi_\alpha = D_r^{-1} F D_c^{-1} u_\alpha$ $\psi_{\alpha i} = \sum_{j=1}^{n_2} \frac{f_{ij}}{f_{i.}f_{.j}} u_{\alpha j}$	Coordonnées factorielles	$\phi_\alpha = D_c^{-1} F' D_r^{-1} v_\alpha$ $\phi_{\alpha j} = \sum_{i=1}^{n_1} \frac{f_{ij}}{f_{i.}f_{.j}} v_{\alpha i}$

Les coordonnées factorielles sont centrées : $\sum_{i=1}^{n_1} f_{i.} \psi_{\alpha i} = \sum_{j=1}^{n_2} f_{.j} \phi_{\alpha j} = 0$.

et de variance égale à λ_α : $\sum_{i=1}^{n_1} f_{i.} \psi_{\alpha i}^2 = \sum_{j=1}^{n_2} f_{.j} \phi_{\alpha j}^2 = \lambda_\alpha$

3.6.3 Relation entre les deux espaces

L'analyse générale a montré que les matrices **S** et **T** ont les mêmes valeurs propres non nulles λ_α et qu'entre le vecteur propre unitaire u_α de **S** associé à λ_α et le vecteur propre unitaire v_α de **T** relatif à la même valeur propre, il existe les relations dites de transition :

$$\begin{cases} v_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} F D_c^{-1} u_\alpha, \\ u_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} F' D_r^{-1} v_\alpha, \end{cases}$$

La comparaison de ces relations avec les expressions des coordonnées factorielles :

$$\begin{cases} \psi_\alpha = D_r^{-1} F D_c^{-1} u_\alpha, \\ \phi_\alpha = D_c^{-1} F' D_r^{-1} v_\alpha \end{cases}$$

montre que celles-ci sont liées aux composantes des axes de l'autre espace par les formules :

$$\begin{cases} \psi_\alpha = \sqrt{\lambda_\alpha} D_r^{-1} v_\alpha, \\ \phi_\alpha = \sqrt{\lambda_\alpha} D_c^{-1} u_\alpha \end{cases}$$

C'est à dire explicitement

$$\begin{cases} \psi_\alpha = \frac{\sqrt{\lambda_\alpha}}{f_{.i}} v_{\alpha i}, \\ \phi_\alpha = \frac{\sqrt{\lambda_\alpha}}{f_{.j}} u_{\alpha j} \end{cases}$$

Ces Relations de transition (ou quasi-barycentriques) conduisent aux relations fondamentales existant entre les coordonnées des points lignes et des points-colonnes sur l'axe α , les relations quasi-barycentriques :

$$\begin{cases} \psi_{\alpha i} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j=1}^{n_2} \frac{f_{ij}}{f_{.i}} \phi_{\alpha j}, \\ \phi_{\alpha j} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^{n_1} \frac{f_{ij}}{f_{.j}} \psi_{\alpha i} \end{cases}$$

La matrice de terme général $\frac{f_{ij}}{f_{.i}}$ permettant de calculer les coordonnées d'un point i à partir de tous les points j n'est autre que le tableau des profils-lignes.

Remarque 2.6.1. Toutes les valeurs propres sont nécessairement inférieures ou égales à 1 (≤ 1).

En effet puisque :

$$\sqrt{\lambda_\alpha} \psi_{\alpha i} = \sum_{j=1}^{n_2} \frac{f_{ij}}{f_{.i}} \phi_{\alpha j}$$

$$\text{on a : } \min_j \{ \phi_{\alpha j} \} \leq \sqrt{\lambda_\alpha} \psi_{\alpha i} \leq \max_j \{ \phi_{\alpha j} \}$$

d'où :

$$\max_i \{ \sqrt{\lambda_\alpha} \psi_{\alpha i} \} \leq \max_j \{ \phi_{\alpha j} \}$$

De la même manière, on a :

$$\max_j \{\sqrt{\lambda_\alpha} \phi_{\alpha j}\} \leq \max_i \{\psi_{\alpha i}\}$$

comme $\lambda_\alpha \geq 0$:

$$\max_j \{\sqrt{\lambda_\alpha} \phi_{\alpha j}\} \leq \max_j \{\phi_{\alpha j}\}$$

finalement : $\lambda_\alpha \leq 1$.

3.6.4 Représentation simultanée

Les relations de transition où quasi-barycentriques justifient la représentation simultanée des lignes et des colonnes. La figure ci-dessous illustre schématiquement le processus de l'analyse des correspondances.

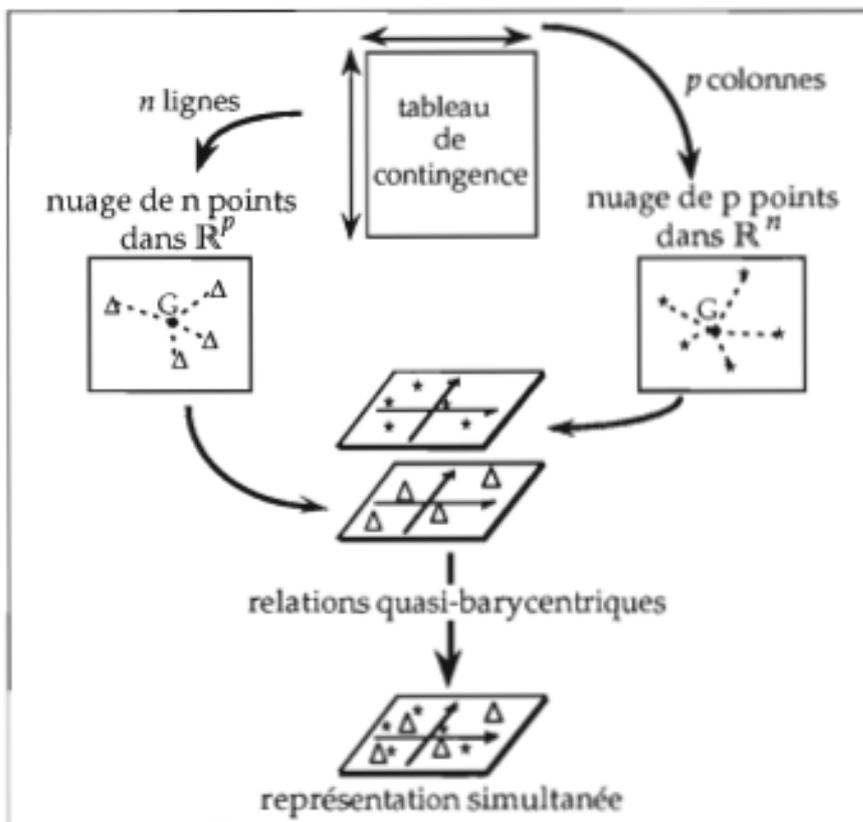


FIGURE 2.4 – Schéma de la représentation simultanée

3.7 Règles d'interprétation

Les nuages de points-lignes et de points-colonnes vont être représentés dans les plans de projection formés par les premiers axes factoriels pris deux à deux. La lecture des graphiques nécessite cependant des règles d'interprétation, en particulier pour apprécier les proximités, identifier les éléments responsables de la formation des facteurs et ceux qui en sont des caractéristiques.

Ces règles s'appuient sur le bilan de l'opération de réduction que constitue la séquence des valeurs propres et des pourcentages d'inertie, ainsi que sur un ensemble de coefficients classiques : les contributions absolues et les cosinus carrés .

3.7.1 Inertie et test d'indépendance

L'inertie totale I du nuage de points par rapport au centre de gravité s'écrit par définition :

$$I = \sum_{i=1}^{n_1} f_i \cdot d^2(i, G) = \sum_{j=1}^{n_2} f_j \cdot d^2(j, G) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(\frac{f_{ij} - f_i \cdot f_j}{f_i \cdot f_j} \right)$$

L'inertie s'exprime également par : $I = \sum_{\alpha=1}^{n_2-1} \lambda_{\alpha}$.

La somme des valeurs propres non triviales d'une analyse des correspondances a donc une interprétation statistique simple.

On pourra rejeter l'hypothèse nulle d'indépendance des variables en lignes et en colonnes si la valeur observée χ^2 dépasse la valeur χ_0^2 qui a une probabilité d'être dépassée inférieure à un seuil fixé au préalable.

3.7.2 contributions et cosinus

Deux séries de coefficients apportent une information supplémentaire par rapport aux coordonnées factorielles :

- les *contributions*, parfois appelées contributions absolues, qui expriment la part prise par une modalité de la variable dans l'inertie (ou variance) "expliquée" par un facteur ;

-
- les *cosinus carrés*, parfois appelés *contributions relatives* ou *qualité de représentation*, qui expriment la part prise par un facteur dans la dispersion d'une modalité de la variable.

Contributions

L'origine étant prise au centre de gravité, les coordonnées factorielles sont centrées et la variance vaut λ_α . Ainsi le quotient :

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

mesure la part de l'élément i dans la variance prise en compte sur l'axe α .

Ce quotient est appelé contribution de l'élément i à l'axe α et permet de savoir dans quelle proportion un point i contribue à l'inertie λ_α du nuage projeté sur l'axe α . Pour tout axe on a :

$$\sum_{i=1}^{n_1} Ctr_\alpha(i) = 1$$

De la même façon on définit la contribution de l'élément j à l'axe α par :

$$Ctr_\alpha(j) = \frac{f_j \phi_{\alpha j}^2}{\lambda_\alpha}.$$

avec la relation : $\sum_{j=1}^{n_2} Ctr_\alpha(j) = 1$

Cosinus carrés

On cherche à apprécier si un point est bien représenté sur un sous-espace factoriel.

La "qualité" de la représentation du point i sur l'axe α peut être évaluée par le *cosinus* de l'angle entre l'axe et le vecteur joignant le centre de gravité du nuage au point i :

$$Cos_\alpha^2(i) = \frac{d_\alpha^2(i, G)}{d^2(i, G)} = \frac{\psi_{\alpha i}^2}{d^2(i, G)}.$$

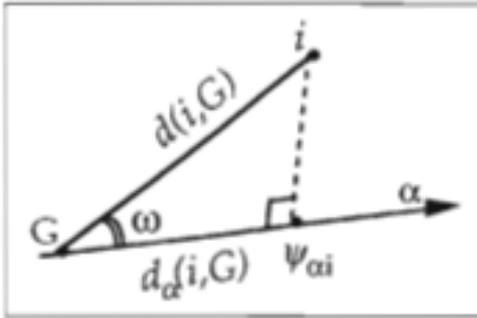


FIGURE 2.5 – Projection du point i sur l'axe α

Cette quantité, appelée *cosinus carré*, représente la part de la distance au centre prise en compte dans la direction α . On l'appelle aussi *la contribution relative* du facteur à la position du point i .

Plus le cosinus carré est proche de 1, plus la position du point observé en projection est proche de la position réelle du point dans l'espace .

Notons que pour tout i : $\sum_{\alpha} \text{Cos}_{\alpha}^2(i) = 1$

La même procédure pour les points colonnes On mesure la contribution relative du facteur α à la position du point j par le cosinus carré de j :

$$\text{Cos}_{\alpha}^2(j) = \frac{d_{\alpha}^2(j, G)}{d^2(j, G)} = \frac{\phi_{\alpha j}^2}{d^2(j, G)}$$

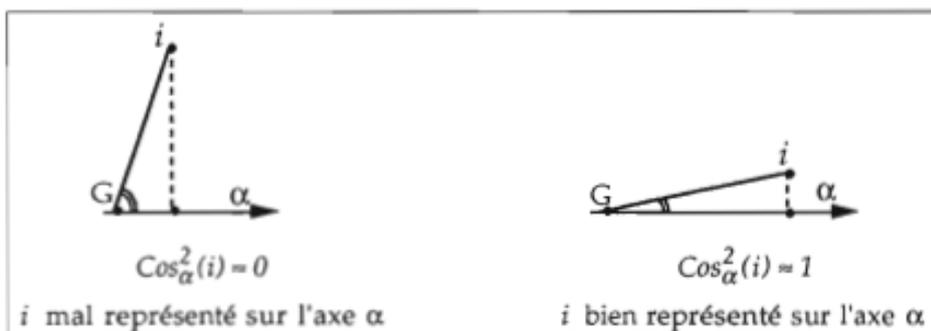


FIGURE 2.6 – Qualité de représentation d'un point i sur l'axe α

et l'on a également pour tout j : $\sum_{\alpha} \text{Cos}_{\alpha}^2(j) = 1$

3.8 Conclusion

L'ACP et l'AFC sont différentes en plusieurs points, elles fournissent des éclairages complémentaires. L'AFC est une méthode puissante pour synthétiser et résumer de vastes tableaux de contingence. En pratique elle est appliquée à beaucoup d'autres tableaux, notamment les tableaux individus-variables. Les individus sont alors considérés comme une variable. Dans le cas de tableaux de contingence, le principal objectif de cette analyse est de dégager les liaisons entre deux variables. L'analyse des correspondances multiples que nous exposons dans le chapitre suivant permet l'étude des liaisons entre plus de deux variables.

3.9 Exemple

Une étude a été menée sur 592 femmes dans le but d'analyser une éventuelle liaison entre les variables "couleur des yeux" et "couleur des cheveux". Dans la suite, on désignera par X la variable "couleur des yeux" et par Y la variable "couleur des cheveux". Les résultats de l'étude sont contenus dans le tableau suivant :

X	Brun	Châtain	Roux	Blond	
Marron	68	119	26	7	
Noisette	15	54	14	10	
Vert	5	29	14	16	
Bleu	20	84	17	94	
					592

1. Tableau des fréquences exprimées en pourcentage, et distributions marginales : $f_{ij} = \frac{n_{ij}}{n} = \frac{n_{ij}}{592}$

Les profils-lignes sont les distributions conditionnelles de Y sachant X .

Les profils-colonnes sont les distributions conditionnelles de X sachant Y . On obtient :

	Brun	Châtain	Roux	Blond	$f_{i.}$	%	Brun	Châtain	Roux	Blond	
Marron	11.49	20.11	4.39	1.18	37.17	Marron	30.91	54.11	11.81	3.17	100
Noisette	2.53	9.12	2.36	1.69	15.70	Noisette	16.11	58.10	15.03	10.76	100
Vert	0.84	4.90	2.36	2.70	10.80	Vert	7.78	45.37	21.85	25.00	100
Bleu	3.38	14.19	2.87	15.89	36.33	Bleu	9.30	39.06	7.90	43.74	100
$f_{.j}$	18.24	48.32	11.98	21.46	100	$f_{.j}$	18.24	48.32	11.98	21.46	100

%	Brun	Châtain	Roux	Blond	f_i
Marron	62.99	41.62	36.64	5.50	37.17
Noisette	13.87	18.87	19.70	7.88	15.70
Vert	4.60	10.14	19.70	12.58	10.80
Bleu	18.54	29.37	23.96	74.04	36.33
	100	100	100	100	100

Chaque profil-ligne ayant quatre coordonnées, les profils-lignes seront représentés dans l'espace \mathbb{R}^4 . La somme des coordonnées étant constante et égale à 1, seules trois coordonnées sont suffisantes pour déterminer la position d'un profil. Par conséquent, les profils-lignes vont appartenir à un hyperplan, c'est-à-dire un sous-espace de dimension 3.

Chaque profil-colonne ayant quatre coordonnées, les profils-colonnes seront représentés dans l'espace \mathbb{R}^4 . La somme des coordonnées étant constante et égale à 1, seules trois coordonnées sont suffisantes pour déterminer la position d'un profil. Par conséquent, les profils-colonnes vont appartenir à un hyperplan, c'est-à-dire un sous-espace de dimension 3.

2. La matrice des pondérations des profils-lignes et La matrice des pondérations des profils-colonnes :

$$D_r = D_n = \begin{pmatrix} 0,3717 & 0 & 0 & 0 \\ 0 & 0,1570 & 0 & 0 \\ 0 & 0 & 0,1080 & 0 \\ 0 & 0 & 0 & 0,3633 \end{pmatrix} \quad D_p = \begin{pmatrix} 0,1824 & 0 & 0 & 0 \\ 0 & 0,4832 & 0 & 0 \\ 0 & 0 & 0,1198 & 0 \\ 0 & 0 & 0 & 0,2146 \end{pmatrix}$$

A partir de ces matrices, et des relations $M_p = D_p^{-1}$ et $M_n = D_n^{-1}$,

3. La matrice S à diagonaliser est $S = F_\ell^t D_n F_\ell D_p^{-1}$. Sachant que l'on a $F_\ell = D_n^{-1} F$, on a aussi $S = F^\dagger D_n^{-1} F D_p^{-1}$. Dans les deux cas, on aboutit à

$$S = \begin{pmatrix} 0.2378 & 0.1942 & 0.1825 & 0.1083 \\ 0.5146 & 0.4956 & 0.4956 & 0.4217 \\ 0.1199 & 0.1228 & 0.1348 & 0.1043 \\ 0.1274 & 0.1873 & 0.1868 & 0.3655 \end{pmatrix}$$

4. La matrice T à diagonaliser pour l'ajustement des profils-colonnes est $T = F_c D_p F_c^t D_n^{-1}$. Sachant que l'on a $F_c = F D_n^{-1}$, on a aussi $T = F D_n^{-1} F^t D_n^{-1}$. Dans les deux cas, on aboutit à :

$$T = \begin{pmatrix} 0.4649 & 0.4042 & 0.3316 & 0.2741 \\ 0.1707 & 0.1700 & 0.1591 & 0.1366 \\ 0.0963 & 0.1094 & 0.1240 & 0.1144 \\ 0.2679 & 0.3161 & 0.3851 & 0.4747 \end{pmatrix}$$

-
5. L'opération de diagonalisation de la matrice S par un logiciel de mathématiques a permis d'obtenir les valeurs propres suivantes : $\lambda_0 = 1, \lambda_1 = 0.20877, \lambda_2 = 0.02223, \lambda_3 = 0.00260$. Que pouvez-vous dire sur la valeur propre λ_0 ? Sur quelles valeurs propres allez-vous concentrer votre analyse?

D'après le cours, on sait que la première valeur propre est toujours égale à l'unité. Cette valeur propre est associée à un vecteur propre engendrant un sous-espace vectoriel de dimension un dit trivial. En pratique, on supprime cette valeur propre qui n'apporte pas d'information. On concentre alors l'analyse sur les trois dernières valeurs propres.

6. Déterminer l'inertie associée à chacun des axes factoriels non-triviaux. En déduire le pourcentage d'inertie associé à chacun des axes, puis l'inertie cumulée.

on a $I_T = \sum_{\alpha=1}^3 = \lambda_1 + \lambda_2 + \lambda_3 = 0.23360$.

	λ	%
λ_1	0.20877	89.37%
λ_2	0.02223	9.52%
λ_3	0.00260	1.11%

Sous la contrainte de devoir conserver au moins 90% d'information, on est amené à garder les deux premiers axes factoriels.

7. Les deux premiers vecteurs propres issus de la diagonalisation de la matrice S sont respectivement

$$u_1 = (-0.201436, -0.156799, -0.033982, 0.392135) \quad u_2 = (-0.262634, 0.105975, 0.256799, -0.100131)$$

rappeler la formule permettant d'obtenir les coordonnées des profils-lignes dans la nouvelle base, et calculer les composantes Ψ_α .

$$\Psi_\alpha = F_\ell D_p^{-1} u_\alpha$$

$$\Psi_1 = \begin{pmatrix} -0.4925 \\ -0.2125 \\ 0.1617 \\ 0.5474 \end{pmatrix} \quad \text{et} \quad \Psi_2 = \begin{pmatrix} -0.0883 \\ 0.1673 \\ 0.3391 \\ -0.0831 \end{pmatrix}$$

8. Rappeler la formule permettant d'obtenir la contribution d'un profil-ligne à la formation d'un axe. et faire les calculs :

$$Ctr_{\alpha}(i) = f_i \frac{\Psi_{\alpha}^2(i)}{\lambda_{\alpha}}$$

Profils-lignes	$Ctr_1(i)$	$Ctr_2(i)$
1	43.18	13.04
2	3.40	19.79
3	1.35	55.83
4	52.14	11.26

9. Indiquer la formule quasi-barycentrique permettant de calculer les coordonnées des profils-colonnes à partir des coordonnées des profils-lignes.

$$\sqrt{\lambda_{\alpha}}\varphi_{\alpha}(j) = \sum_{i=1}^4 f_{i/j}\Psi_{\alpha}(i), \quad \forall j = 1, \dots, 4$$

On obtient alors

$$\varphi_1 = \begin{pmatrix} -0.5046 \\ -0.1483 \\ -0.1296 \\ 0.8349 \end{pmatrix} \quad \text{et} \quad \varphi_2 = \begin{pmatrix} -0.2148 \\ 0.0327 \\ 0.3196 \\ -0.0696 \end{pmatrix}$$

10. Rappeler la formule permettant d'obtenir la contribution d'un profil-colonne à la formation d'un axe et faire les calculs :

$$Cr_{\alpha}(j) = f_{.j} \frac{\varphi_{\alpha}^2(j)}{\lambda_{\alpha}}$$

Profils-colonnes	$Ctr_1(j)$	$Ctr_2(j)$
1	22.25	37.88
2	5.09	2.32
3	0.96	55.13
4	71.7	4.67

On obtient alors pour les profils-colonnes

Interprétation : La représentation simultanée sur le premier plan factoriel (2.7) montre par exemple que les femmes aux yeux bleus et aux yeux marrons sont éloignées. En confirmation avec le tableau, nous remarquons que les femmes aux yeux bleus auront tendance à avoir les cheveux blonds, ainsi que pour celles aux yeux marrons qui seront davantage brunes. Les femmes aux cheveux roux auront plutôt les yeux verts ou noisettes. La modalité des cheveux châains est proche de l'origine, elle représente donc un profil moyen et n'est rattachée à aucune couleur de cheveux.

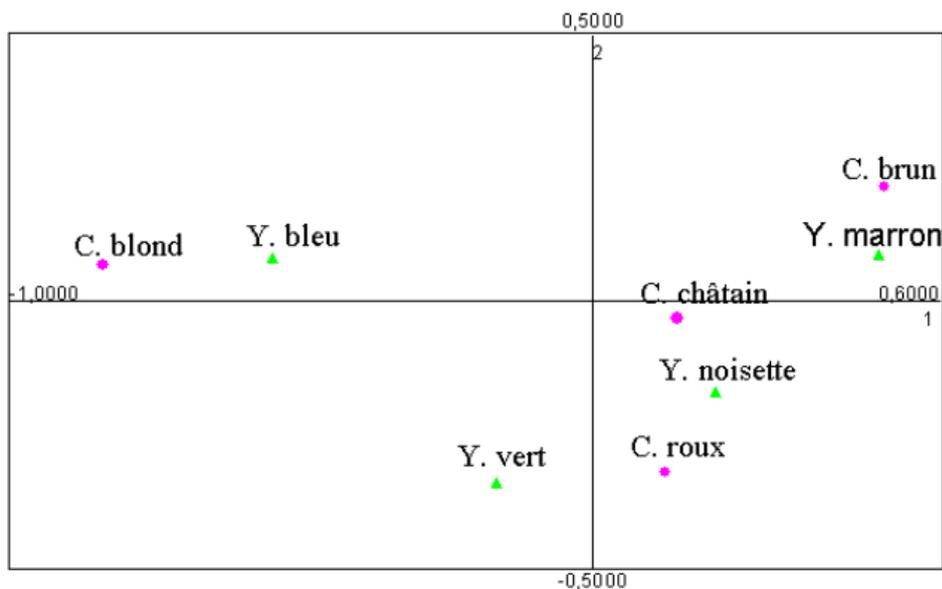


FIGURE 2.7 – Représentation simultanée dans le premier plan

3.10 Exercice

Exercice 3.10.1. Au cours d'un enquête sur un échantillon de taille 60, on a obtenu le tableau de contin-

gence suivant :

	<i>a</i>	<i>b</i>
1	10	10
2	5	15
3	15	5

Réaliser une (AFC) sur ces données, en répondant aux questions suivantes :

- 1) Donner le tableau des fréquences conjointes et marginales, associé au tableau précédent. (Conseil : Utiliser des fraction au lieu des nombres décimales)
- 2) Déterminer la matrice des profils lignes et profils colonnes ?
- 3) Déterminer la distance de χ^2 entre les deux dernières lignes ?

Exercice 3.10.2. On considère le tableau *K* suivant où *a* est un entier non nul :

<i>I/J</i>	<i>j</i> ₁	<i>j</i> ₂	<i>j</i> ₃	<i>j</i> ₄	<i>j</i> ₅
<i>i</i> ₁	<i>a</i>	<i>a</i>	<i>a</i>	0	0
<i>i</i> ₂	<i>a</i>	<i>a</i>	0	<i>a</i>	0
<i>i</i> ₃	0	<i>a</i>	0	<i>a</i>	<i>a</i>

On pose $I = \{i_1, i_2, i_3\}$ et $J = \{j_1, j_2, j_3, j_4, j_5\}$:

On effectue l'analyse factorielle des correspondances (AFC) de *K*.

1. Déterminer les centres de gravité des nuages $N(I)$ et $N(J)$.
2. Déterminer la matrice des profils colonnes F_1 ainsi que la matrice des profils lignes F_2 de K .
3. calculer le produit $F_1 F_2$
4. Quel est l'influence du réel a sur l'AFC de ce tableau ?
5. Quel est l'axe factoriel trivial, à quelle valeur propre est-il associé ?
6. Quelle est l'inertie du nuage $N(J)$?
7. On pose

$$w_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \text{ Montrer que } w_1 \text{ et } w_2 \text{ sont des vecteurs propres de } F_1 F_2, \text{ en}$$

deduire les axes factoriels t' non triviaux u_1 et u_2 ainsi que les valeurs propres associées. On choisira u_1 de manière que la première coordonnée soit positive, de même pour u_2 .

8. On note $\varphi_\alpha(i)$ l'abscisse de la projection du profil de la ligne i sur le α ème axe factoriel. Remplir le tableau suivant avec la contrainte $\varphi_\alpha(i) \geq 0$

I/J	φ_1	φ_2	φ_3
i_1			
i_2			
i_3			

- 9.
10. Représenter les deux nuages $N(I)$ et $N(J)$ simultanément dans le plan factoriel 1-2.
11. Calculer la contribution de i_1 à chacun des axes factoriels non triviaux ainsi que la qualité de représentation de i_1 dans le plan factoriel 1-2 c'est-à-dire $COR_1(i_1) + COR_2(i_1)$.

Solution :

1. $f_I = \left(1/3, 1/3, 1/3 \right)$ et $f_J = \frac{1}{9} \left(2, 3, 1, 2, 1 \right)$.

2. $F_1 = \frac{1}{6} \begin{pmatrix} 3 & 2 & 6 & 0 & 0 \\ 3 & 2 & 0 & 3 & 0 \\ 0 & 2 & 0 & 3 & 6 \end{pmatrix}$ et $F_2 = \frac{1}{3} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$

$$3. F_1 F_2 = \frac{1}{18} \begin{pmatrix} 11 & 5 & 2 \\ 5 & 8 & 5 \\ 2 & 5 & 11 \end{pmatrix}$$

4. **Aucune** puisque les individus des nuages ne dependent pas de a ainsi que les poids et les métriques.

5. f_I est le vecteur propre de $F_1 F_2$ associé a la valeur propre triviale 1.

6. L'inertie du nuage $N(J)$ est la trace de $F_1 F_2$ moins 1 donc $30/18 - 1 = 2/3$.

7. On vérifie que : $F_1 F_2 w_1 = \frac{1}{2} w_1$ et $F_1 F_2 w_2 = \frac{1}{6} w_2$.

Or la norme de w_1 pour la metrique $D_1^{-1} = \frac{1}{f_I} = 3I_3$ donc le premier axe factoriel u_1 associé a la valeur propre $\lambda_1 = 1/2$ est $u_1 = \frac{1}{\sqrt{6}} w_1$ et le deuxième axe factoriel u_2 associé a la valeur propre $\lambda_2 = 1/6$ est $u_2 = \frac{1}{3\sqrt{2}} w_2$.

8. on a $\varphi_1 = \frac{\sqrt{3}}{2} w_1$ et $\varphi_2 = \frac{\sqrt{3}}{6} w_2$

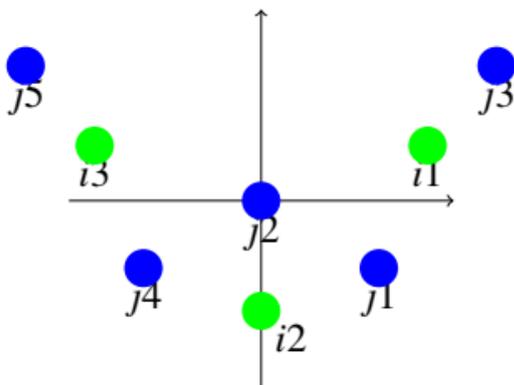
I/J	φ_1	φ_2	φ_3
i_1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{6}$	0
i_2	0	$-\frac{2\sqrt{3}}{6}$	0
i_3	$-\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{6}$	0

9.

I/J	j_1	j_2	j_3	j_4	j_5
ψ_1	$\frac{\sqrt{6}}{4}$	0	$\frac{2\sqrt{6}}{4}$	$-\frac{\sqrt{6}}{4}$	$-\frac{2\sqrt{6}}{4}$
ψ_2	$-\frac{\sqrt{2}}{4}$	0	$\frac{2\sqrt{2}}{4}$	$-\frac{\sqrt{2}}{4}$	$\frac{2\sqrt{2}}{4}$

 Avec les formules de transition on a

$$\psi_1 = \sqrt{2} F_1' \varphi_1 \text{ et } \psi_2 = \sqrt{6} F_1' \varphi_2$$



10.

11.

$$CTR_1(i_1) = \frac{1}{3} \times \frac{3/4}{1/2} = 1/2 \text{ et } CTR_2(i_1) = \frac{1}{3} \times \frac{3/36}{1/6} = 1/6 \quad (2.1)$$

Comme il n'y a que deux axes non triviaux, la qualité de représentation de i_1 dans le plan factoriel 1-2 est 1.