

UNIVERSITÉ ABDELHAMID IBN BADIS-MOSTAGANEM
FACULTÉ DES SCIENCES EXACTES ET L'INFORMATIQUE
DÉPARTEMENT DE MATHÉMATIQUES ET INFORMATIQUE



Spécialité : Mathématique

Option : Modélisation Contrôle et Optimisation.

Mémoire de Master intitulé

Analyse En Composante Principale

Présenté par

BEKAR Sara

Soutenu le 26/05/2015

Devant le jury

Mm	Diala	Président	M.A.A	Université de MOSTAGANEM.
Mr	Mechdene M	Examineur	M.A.A	Université de MOSTAGANEM.
Mr	Mohammedi M	Encadreur	M.A.A	Université de MOSTAGANEM.

Année Universitaire 2014-2015

Table des matières

Remerciments	i
Introduction	iii
1 Rappels	1
1.1 Les matrices	1
1.1.1 Généralités	1
1.2 Les espaces	3
1.2.1 Espace vectoriel	3
1.2.2 Espace métrique	4
1.2.3 Espace euclidien	4
1.3 Projecteurs	5
1.3.1 Projecteurs M -orthogonaux	5
1.3.2 Ecriture explicite d'un projecteur	6
1.4 Formule de dérivation vectorielle	6
1.5 Application à la maximisation du quotient de deux formes quadratiques	7
2 Présentation générale de l'ACP	9
2.1 La formulation mathématique de l'ACP	10
2.1.1 Tableau des données	10
2.1.2 Choix de distance	11

2.1.3	Choix de l'origine	12
2.2	Inertie et moment d'inertie	13
2.2.1	Inertie en un point	13
2.2.2	Moment d'inertie par rapport à un sous-espace	14
2.3	Étude des valeurs et vecteurs propres de $V \circ M$	19
2.3.1	Axes et plans principaux	20
2.3.2	Généralisation	24
3	Représentation et Interprétation des points	25
3.1	Représentation des individus	25
3.1.1	Qualité de la représentation des individus	26
3.1.2	Interprétation des nouveaux axes en fonction des individus	27
3.2	Représentation des variables	27
3.2.1	Qualité de la représentation des variables	30
3.2.2	Étude des liaisons entre les variables	30
3.3	Individus et variables supplémentaires	31
4	Application	32
4.1	Les données	32
4.2	Résultats préliminaires	33
4.3	Résultats généraux	33
4.4	Interprétation	34
4.5	Résultats sur les variables	35
4.6	Interprétation	35
4.7	Résultats sur les individus	36
4.8	Interprétation	37
	Conclusion	40
	Bibliographie	41

Remerciements

Je remercie tout d'abord mon bon Dieu le tout puissant de m'avoir donné le courage, la patience et la volonté durant toutes mes années d'études.

A l'occasion de notre fin d'étude, et la réalisation du mémoire nous tenant à remercier tous qui nous ont aidés de réaliser ce travail, et surtout nos professeurs de la branche et spécialement "Mr Besnasi".

Je tiens en deuxième lieu à exprimer mes plus vifs remerciements à mon encadreur "Mr Mohammedi". Qui m'a guidé dans ce travail, pour son aide continues ses encouragements sans cesse et ses conseils précieux.

Je voudrais remercier également les membres du jury, et tout les personnes qui ont participé de près ou de loin à la réalisation de ce travail et qui m'ont tant soutenu et encouragé.

Enfin un grand merci à ma famille BEKAR et surtout mon père et ma mère.

Dédicace

Je dédie ce modeste travail à

Mes parent, pour leur soutien, patiences et leurs sacrifices durant ma études et durant ce projet.

A tous mes enseignants, pour leur bieveillance et pour leur contribution à notre solide formation.

A ma famille et mes amies pour leurs conseils et leurs encouragements.

A tous ceux qui ont contribué de près ou de loin à la réalisation de ce travail.

INTRODUCTION

L'analyse statistique multidimensionnelle "ou les méthodes factorielles" est la partie de la statistique traitent des données quantitatives complexes, lorsque l'on ne fait pas d'hypothèses de lois. Ces méthodes qui sont apparues au début des années 30 ont été surtout développées par plusieurs auteurs dans les années 60, en particulier par "*Jean-Paul benzecri*"[2] qui a beaucoup exploité les aspects géométriques, les représentations graphiques et la contribution des points aux axes principaux.

L'analyse en composantes principales "ACP" est souvent considérée comme la méthode de base de l'analyse factorielle des données dont l'objectif est de déterminer des facteurs qui serviront à visualiser les observations, stockées dans un tableau de données ($p \times n$) de façon plus simple.

D'un point de vue mathématique, l'ACP correspond à l'approximation d'une matrice ($p \times n$) ou représentation graphique optimal de dimension $q < p$ "réduction de dimension", q étant souvent petit 2 ou 3, pour qu'on puisse voir les représentation graphiques. Cette réduction de dimension se traduit par la projection dans un sou-espace des variables F_q de dimension q le plus proche du nuage initial au sens du moment d'inertie.

Nous avons choisi de structurer notre manuscrit selon le plan décrit ci-dessous :

Le premier chapitre abord des notions de base concernant l'algèbre linéaire, et les projecteurs ainsi que la maximisation des formes quadratiques. Dans le chapitre 2, l'ACP est présentée d'une façon géométrique et matricielle et ainsi l'étude de la matrice de variance-covariance.

La représentation et l'interprétation des données ont été présentées dans le chapitre 3.

Finalement, un exemple d'application sur les notes de statistique est abordé dans le dernier chapitre.

Rappels

Dans ce chapitre on rappelons quelque notions qui peut être utilisé ultérieurement dans nos travail.

1.1 Les matrices

Définition 1.1.1 *Une matrice à m lignes et n colonnes est un tableau rectangulaire de $m \times n$ nombres, rangés ligne par ligne. Il y a m lignes, et dans chaque ligne n nombres, ces matrices sont à coefficients réels ou complexes.*

1.1.1 Généralités

Matrice inversible

Définition 1.1.2 *Une matrice carrée A d'ordre n est dite inversible ou régulière ou encore non singulière, si et seulement si son déterminant est non nul donc il existe une matrice B d'ordre n , appelée matrice inverse de A notée A^{-1} , telle que*

$$AB = BA = I_n. \tag{1.1.1}$$

où I_n désigne la matrice unité d'ordre n et la multiplication est la multiplication ordinaire des matrices, d'après le théorème du rang, chacune des deux conditions $AB = I_n$ ou $BA = I_n$ suffit. Dans ce cas, la matrice B est unique.

Matrice symétrique

Définition 1.1.3 Une matrice symétrique A est une matrice carrée qui est égale à sa propre transposée, c'est-à-dire telle que $a_{i,j} = a_{j,i}$ pour tous i et j compris entre 1 et n , où les $a_{i,j}$ sont les coefficients de la matrice A et n est son ordre, noté A' .

Matrice symétrique définie positive A

Définition 1.1.4 C'est une matrice symétrique dont les valeurs propres sont strictement positives et pour toute x non nulle à n éléments, on a

$$x'Ax > 0.$$

Matrice diagonale

Définition 1.1.5 Une matrice carrée $D = (d_{i,j})_{1 \leq i,j \leq n}$ est dite diagonale si

$$\forall (i,j) \in [1,n]^2; i \neq j \implies d_{i,j} = 0.$$

La trace de matrice

Définition 1.1.6 En appelle trace de la matrice carrée A est la somme des éléments diagonaux de A ; noté

$$\text{Trace}A = \sum_{i=1}^n a_{i,i}.$$

Valeur propre d'une matrice carrée A

Définition 1.1.7 scalaire λ est une valeur propre de A s'il existe un vecteur x non nul tel que $Ax = \lambda x$.

- L'ensemble des valeurs propres de A est donc l'ensemble des scalaires λ tels que $A - \lambda I_d$ n'est pas injectif.

Vecteur propre de A

Définition 1.1.8 Soit x un vecteur non nul, x est un vecteur propre de A s'il existe un scalaire λ tel que $Ax = \lambda x$. On dit que x est un vecteur propre associé à la valeur propre λ .

- Un vecteur propre ne peut pas être associé à deux valeurs propres différentes.
- Une famille de k vecteurs propres associés à k valeurs propres différentes constitue une famille libre.

1.2 Les espaces

1.2.1 Espace vectoriel

Définition 1.2.1 Un espace vectoriel E de \mathbb{R} est un ensemble muni de deux opérations "addition et la multiplication", c'est à dire qu'à tout couple $v, w \in E$ on peut associer $v + w \in E$ tel que les règles de calcul ordinaires dans \mathbb{R}^n aient lieu. A savoir

1. $(u + v) + w = u + (v + w)$; pour tous $u, v, w \in E$.
2. $u + v = v + u$; pour tous $v, u \in E$.
3. Il existe un élément noté 0_E tel que $u + 0_E = u$; pour tout $u \in E$.
4. Pour tout $u \in E$ existe un élément $v \in E$ tel que $u + v = 0_E$.
5. $1.v = v$; pour tous $v \in E$.
6. $(\alpha + \beta).v = \alpha v + \beta v$; pour tous $v \in E$ et $\alpha, \beta \in \mathbb{R}$ (resp \mathbb{C}).
7. $\alpha.(\beta.v) = (\alpha.\beta).v$; pour tous $v \in E$ et $\alpha, \beta \in \mathbb{R}$ (resp \mathbb{C}).
8. $\alpha(v + w) = \alpha.v + \alpha.w$.

Sous-espace vectoriel

Définition 1.2.2 Un sous-espace vectoriel F d'un espace vectoriel E est un sous-ensemble non-vide de E tel que

1. Si $v, u \in F$; alors $v + u \in F$.
2. Si $\beta \in \mathbb{R}$ (resp \mathbb{C}) ; alors $\beta.v \in F$.

- $\dim(E) \geq \dim(F)$.
- $\dim(F + G) = \dim(F) + \dim(G)$; pour tous E et F deux sous-espaces vectoriels de G , en somme directe.

L'intersection d'une famille non vide "finie ou infinie" de sous-espaces vectoriels est un sous-espace vectoriel.

Dual d'un espace vectoriel

Définition 1.2.3 On appelle le dual de E et que l'on note E^* , l'ensemble des formes linéaires est un espace vectoriel sur \mathbb{R}

1.2.2 Espace métrique

Définition 1.2.4 Un espace métrique est un ensemble au sein duquel une notion de distance entre les éléments de l'ensemble est définie.

- On appelle (E, d) un espace métrique si E est un ensemble et d une distance sur E .
- On appelle distance sur un ensemble E , une application d définie sur le produit $E_2 = E \times E$ et à valeurs dans l'ensemble \mathbb{R}_+ des réels positifs, telle que, pour tout x, y, z de E , on a

$$d(x, y) = d(y, x) \text{ (symétrie).}$$

$$d(x, y) = 0 \iff x = y \text{ (séparation).}$$

$$d(x, z) \leq d(x, y) + d(y, z) \text{ (inégalité triangulaire).}$$

1.2.3 Espace euclidien

Définition 1.2.5 Un espace euclidien permet de traiter les dimensions supérieures, il est défini par la donnée d'un espace vectoriel sur le corps des réels, de dimension finie, muni d'un produit scalaire, qui permet de "mesurer" distances et angles.

- Un produit scalaire sur un espace vectoriel réel E est une forme bilinéaire $\langle \cdot, \cdot \rangle$ "une forme bilinéaire sur E est une application de $E \times E$ dans \mathbb{R} , linéaire par rapport à chacune des deux variables".

- Symétrique

$$\forall x, y \in E; \langle x, y \rangle = \langle y, x \rangle .$$

- Définie

$$\forall x \in E; \langle x, x \rangle = 0 \Rightarrow x = 0.$$

- Positive

$$\forall x \in E; \langle x, x \rangle > 0.$$

- Un espace vectoriel euclidien est un espace vectoriel réel de dimension finie muni d'un produit scalaire.
- Un espace affine euclidien est un espace affine dont la direction est un espace vectoriel euclidien.
- La norme euclidienne, sur un espace vectoriel euclidien E , est la fonction de E dans \mathbb{R}_+ qui à un vecteur x associe la racine carrée du produit scalaire de x par lui-même

$$\|x\| = \sqrt{\langle x, x \rangle}. \quad (1.2.1)$$

- La distance euclidienne associée est entre deux vecteurs, la norme euclidienne de leur différence

$$d(x, y) = \|y - x\|. \quad (1.2.2)$$

1.3 Projecteurs

Définition 1.3.1 *Un projecteur P "ou une projection" est une application linéaire qu'on peut présenter de deux façons équivalentes :*

1. *Une projection linéaire associée à une décomposition de E comme somme de deux sous-espaces supplémentaires, c'est-à-dire qu'elle permet d'obtenir un des termes de la décomposition correspondante.*
2. *Une application linéaire idem potente, elle vérifie $P^2 = P$.*

1.3.1 Projecteurs M -orthogonaux

Etant donné un sous-espace W de E , P est la matrice de projection M -orthogonale sur W si $Py \in G$ et si $\langle Py, y - Py \rangle = 0$.

Ce qui revient à écrire que $Py \in G \forall y$, que $P^2 = P$ et que $P'M = MP$.

- Un projecteur M -orthogonal est une matrice idem potente et M -symétrique.
- $\text{Trace}P = \dim G$.

- Si P_1, P_2, \dots, P_k sont des projecteurs M -orthogonaux alors $P_1 + P_2 + \dots + P_k$ n'est un projecteur M -orthogonal que si, et seulement si, $P_i P_j = 0$ pour $i \neq j$, c'est à dire si les espaces d'arrivée de P_i sont M -orthogonaux.

1.3.2 Ecriture explicite d'un projecteur

Supposons W le sous-espace engendré par p vecteurs linéairement indépendants x^1, x^2, \dots, x^p et soit X la matrice (p, n) ayant les x^j pour vecteurs-lignes.

$y - Py$ doit être orthogonal à tout vecteur de W ; or tous les vecteurs de W sont de la forme Xu , en particulier $Py = Xb$.

Il faut donc $\langle Xu_i, y - Py \rangle_M = 0$; $i = 1, 2, \dots, n$ où les u_i forment une base de \mathbb{R}^p

$$u_i' X' M (y - Py) = 0, \quad \forall i, \quad (1.3.1)$$

donc $X' M y = X' M P y$; comme $Py = Xb$ et que $\dim W = p$, $X' M X$ est inversible, il vient

$$\begin{aligned} X' M y &= X' M X b; \quad b = (X' M X)^{-1} X' M y, \\ P y &= X b = X (X' M X)^{-1} X' M y, \\ P &= X (X' M X)^{-1} X' M. \end{aligned}$$

En particulier, le projecteur M -orthogonal sur un vecteur x s'écrit

$$P = x (x' M x)^{-1} x' M = \frac{x x' M}{x' M x}. \quad (1.3.2)$$

1.4 Formule de dérivation vectorielle

Soit g une application de l'espace vectoriel \mathbb{R}^p dans $\mathbb{R}/u \rightarrow g(u)$.

Par définition on a

$$\frac{dg}{du} = \begin{pmatrix} \frac{dg}{du_1} \\ \vdots \\ \frac{dg}{du_p} \end{pmatrix} \quad \text{où, } u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{pmatrix}.$$

Propriétés

- (i) : $\frac{d(a' u)}{du} = a$ si a est un vecteur constante.

En effet,

$$a'u = \sum_{i=1}^p a_i u_i,$$

donc,

$$\frac{d(a'u)}{du_i} = a_i.$$

(ii) : Soit A une matrice carrée $p \times p$, symétrique

$$\frac{d(u' Au)}{du} = Au + A'u.$$

Soit a_{ij} l'élément courant de A ; on a

$$u' Au = \sum_i^p \sum_j^p a_{ij} u_i u_j.$$

Cherchons $\frac{d(u' Au)}{du_1}$ les termes contenant u_1 au premier degré sont de deux sortes : ceux provenant de $u_i = u_1$ et ceux provenant de $u_j = u_1$ c'est à dire $\sum_{i \neq 1} a_{i1} u_i u_1$ et $\sum_{j \neq 1} a_{1j} u_1 u_j$ dont les dérivées sont $\sum_{i \neq 1} a_{i1} u_i$ et $\sum_{j \neq 1} a_{1j} u_j$ et il faut ajouter $a_{11} u_1$ à chacun d'eux.

On a donc,

$$\begin{aligned} \frac{d(u' Au)}{du} &= \begin{pmatrix} \sum_j^p a_{1j} u_j \\ \sum_j^p a_{2j} u_j \\ \vdots \\ \sum_j^p a_{pj} u_j \end{pmatrix} + \begin{pmatrix} \sum_j^p a_{i1} u_i \\ \sum_j^p a_{i2} u_i \\ \vdots \\ \sum_j^p a_{ip} u_i \end{pmatrix} \\ &= Au + A'u. \end{aligned}$$

Donc,

$$\frac{d(u' Au)}{du} = 2Au, \text{ d'après propriété (ii).} \tag{1.4.1}$$

1.5 Application à la maximisation du quotient de deux formes quadratiques

Soient A et B deux matrices symétriques de même taille. B sera supposée inversible.

Alors le rapport $\frac{u' Au}{u' Bu}$ est maximal pour u vecteur propre de $B^{-1}A$ associé à sa plus grande valeur propre λ_1 , λ_1 étant alors la valeur du maximum.

Preuve. Un extremum de $\frac{u' Au}{u' Bu}$ s'obtient en annulant sa dérivée qui vaut

$$\frac{(u' Bu)(2Au) - (u' Au)(2Bu)}{(u' Bu)^2},$$

soit

$$\begin{aligned}(u' Bu) Au &= (u' Au) Bu \\ B^{-1}Au &= \left(\frac{u' Au}{u' Bu}\right) u\end{aligned}$$

u est donc vecteur propre de $B^{-1}A$ associé à la valeur propre $\left(\frac{u' Au}{u' Bu}\right)$. Le maximum est donc atteint si cette valeur propre est maximale. \square

Présentation générale de l'ACP

Dans la plupart des situations, on dispose de plusieurs observations sur chaque individu constituant la population d'étude. On a donc à prendre en compte p variables par individu, p étant strictement supérieur à 1.

L'étude séparée de chacune de ces variables donne quelques informations mais est insuffisante car elle laisse de côté les liaisons entre elles, ce qui est pourtant souvent ce que l'on veut étudier.

L'analyse en Composantes Principales est une bonne méthode pour étudier les données multidimensionnelles, lorsque toutes les variables observées sont de type numérique, de préférence dans les mêmes unités, et que l'on veut voir s'il y a des liens entre ces variables.

Dans la littérature, on trouve deux approches différentes de l'ACP :

1. Elle peut être présentée comme la recherche d'un ensemble réduit de variables non-corrélées, combinaisons linéaires des variables initiales résumant avec précision les données.
2. Une autre représentation repose sur la représentation des données initiales à l'aide de nuage de points dans un espace géométrique.

2.1 La formulation mathématique de l'ACP

2.1.1 Tableau des données

La plupart de temps les données se trouvent rassemblées en une matrice X à p variables "lignes" et n individus "colonnes" supposé centrés de la forme suivante :

$$X = X(p \times n) = (x_1, x_2, \dots, x_n) = \begin{pmatrix} x_1^1 & \cdots & \cdots & \cdots & x_n^1 \\ \vdots & & & & \vdots \\ x_1^j & \cdots & \cdots & \cdots & x_n^j \\ \vdots & & & & \vdots \\ x_1^p & \cdots & \cdots & \cdots & x_n^p \end{pmatrix}.$$

Avec $X = (x_i^j)$; $(1 \leq i \leq n; 1 \leq j \leq p)$; $\forall (i, j), x_i^j \in \mathbb{R}$.

Par convention, on placera en exposant ce qui se rapporte aux variables, la première variable x_1 et en indice ce qui se rapporte au premier individu x_1 .

x_1^1 désigne la valeur numérique prise par la première variable, pour le premier individu.

A ce tableau sont donc associés deux espaces vectoriels, F espaces des variables isomorphe à \mathbb{R}^n et E espace des individus isomorphe à \mathbb{R}^p ; aux variables sont associés des axes de \mathbb{R}^p , aux individus des axes de \mathbb{R}^n .

Pour la suite, on considère le tableau X comme un tableau de variables centrées, on considère maintenant $N = D_p$ comme le nuage de points formé par l'ensemble des individus x_i , dans l'espace p dimension des variables et F muni la métrique $M = I_p$, donc le but est d'obtenir une représentation la plus fidèle possible du nuage N en le projetant sur un espace de faible dimension.

Pour cela, on cherche à minimiser les "Écarts" entre les points de N et leurs projections.

Avec ; D_p est la métrique euclidienne des poids "matrice diagonale des poids", s'écrit sous la forme

$$D_p = \begin{pmatrix} p_1 & & & 0 \\ & p_2 & & \\ & & \cdots & \\ 0 & & & p_n \end{pmatrix},$$

et

$$I_p = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \cdots & \\ 0 & & & 1 \end{pmatrix}.$$

2.1.2 Choix de distance

Pour faire une représentation géométrique, il faut choisir une distance entre deux points de l'espace. La distance utilisée par l'ACP de manière naturelle pour mesurer les proximités entre variables est celle définie par la matrice D_p qui la métrique de la covariance quand les variables sont centrées

$$\begin{aligned} COV(x_i, x_j) &= x_i' D_p x_j, \\ \|x_i\|^2 &= x_i' D_p x_i, \end{aligned}$$

où $\|x_i\|$ est l'écart type de la variable x_i .

L'angle formé par deux variables centrées a pour cosinus le coefficient de corrélation empirique. Pour mesurer la distance entre individus, on prendra une matrice M , symétrique définie positive tel que

$$d^2(x_i, x_j) = (x_i - x_j)' M (x_i - x_j).$$

Les métrique les plus couramment utilisées sont

- $M = I$.
- $M = V^{-1}$ métrique de Mahalanobis.
- $M = D_{1/\sigma^2}$ avec D_{1/σ^2} désigne la matrice diagonale des inverses des variances des p variables.

Le choix se fait en fonction des données et des problèmes à résoudre.

- Si E est muni de la métrique M et F de la métrique D_p , alors E^* muni la métrique M^{-1} .

Preuve. Considérons un élément normé à 1 a de $\mathbb{R}^p = E$ espace des individus qui engendre donc un axe de E

représentant une variable. A cet axe associons la forme linéaire $u \in E^*$ telle que $u = Ma$; autrement dit u

est la forme linéaire "coordonnée M -orthogonale sur l'axe défini par a ".

La matrice M met donc en relation E avec E^* .

Soit maintenant deux éléments a_1 et a_2 de E , leur produit scalaire est $\langle a_1, a_2 \rangle_M = a_1' M a_2$.

Soient u_1 et u_2 les formes linéaires associées, on a

$$u_1 = M a_1, \quad u_2 = M a_2 \quad \text{et} \quad M^{-1} u_1 = a_1, \quad M^{-1} u_2 = a_2.$$

On en déduit

$$a'_1 M a_2 = u'_1 M^{-1} M M^{-1} u_2 = u'_1 M^{-1} u_2,$$

en particulier si $a' M a = 1$, $u' M^{-1} u = 1$.

Il faut donc que E^* soit muni de la métrique M^{-1} . □

2.1.3 Choix de l'origine

Le point "o" correspondant au vecteur de coordonnées toutes nulles n'est pas forcément une origine satisfaisante, car si les coordonnées des points du nuage des individus sont grandes, le nuage est éloigné de cette origine. Il apparaît plus juducious de choisir une origine liée au nuage lui-meme : le centre de gravité du nuage.

Pour définir ce centre de gravité, il faut choisir un système de pondération des individus "unités".

$\forall i = 1, \dots, n$; $p_i \succ 1$, poids de l'unité x_i tel que : $\sum_{i=1}^n p_i = 1$. Par définition le centre de gravité g est défini par

$$g = \sum_{i=1}^n x_i p_i.$$

Pour l'ACP on choisit de donner le même poids $\frac{1}{n}$ à tous les individus, le centre de gravité g du nuage des individus est alors le point dont les coordonnées sont les valeurs moyennes des variables.

$$g = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^1 \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_i^j \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n x_i^p \end{pmatrix} = \begin{pmatrix} x_0^1 \\ x_0^j \\ x_0^p \end{pmatrix}.$$

Prendre g comme origine, revient alors sur le tableau des données centrées

$$X_c = \begin{pmatrix} x_1^1 - x_0^1 & \dots & x_1^j - x_0^j & \dots & x_1^p - x_0^p \\ \vdots & & \vdots & & \vdots \\ x_1^j - x_0^1 & \dots & x_1^j - x_0^j & \dots & x_1^j - x_0^p \\ \vdots & & \vdots & & \vdots \\ x_1^p - x_0^1 & \dots & x_1^j - x_0^j & \dots & x_1^p - x_0^p \end{pmatrix}.$$

Remarque 2.1.1 *une variable centrée est une variable dont la moyenne est nulle.*

2.2 Inertie et moment d'inertie

L'espace vectoriel des individus E a son origine "0" placé au centre de gravité du nuage g

$$g = \sum_{i=1}^n p_i x_i = 0. \quad (2.2.1)$$

A la forme quadratique d'inertie V est alors associée l'application linéaire $V = X \circ Dp \circ X'$ est associée une forme une forme définie positive V sur E^* , la forme de variance

$$V = \text{var}(x) = X \circ Dp \circ X' = \sum_{i=1}^n p_i x_i x_i'. \quad (2.2.2)$$

On appelle inertie du nuage de points la somme pondérée des carrés des distances des points au centre de gravité

$$I_g = \sum_{i=1}^n p_i \|x_i - g\|^2. \quad (2.2.3)$$

2.2.1 Inertie en un point

$I_a = \sum_{i=1}^n p_i \|x_i - a\|^2$ est l'inertie du nuage N au point $a \in E$.

Cette inertie est minimum quand a coïncide avec le centre de gravité $g = 0$ du nuage N .

$$I_a = I_g + \|a\|^2. \quad (\text{Huyghens}), \quad (2.2.4)$$

le nuage N est concentré en g si

$$I_g = \sum_{i=1}^n p_i \|x_i\|^2 = 0. \quad (2.2.5)$$

I_g mesure la dispersion autour du point g , de même que la variance mesure sur un axe la dispersion autour de la moyenne. Si ce moment d'inertie est grand, cela signifie que le nuage est très dispersé.

Une autre expression de l'inertie en g .

Le calcul qui suit utilise les deux propriétés matricielles suivantes :

on a

$$\text{Trace}(AB) = \text{Trace}(BA)$$

$$\text{Trace}(A+B) = \text{Trace}(A) + \text{Trace}(B)$$

$$I_g = \sum_{i=1}^n p_i M(x_i, x_i) = \sum_{i=1}^n p_i x_i' M x_i.$$

$$x_i' M x_i = \text{Trace}(x_i' M x_i) = \text{Trace}(x_i x_i' M).$$

On en déduit

$$I_g = \text{Trace} \left[\left(\sum_{i=1}^n p_i x_i x_i' \right) M \right], \quad (2.2.6)$$

d'où,

$$I_g = \text{Trace}(VM) \text{ d'après (2.2.1)}. \quad (2.2.7)$$

I_g et $\text{Trace}(VM)$ généralisent donc les notions de moyenne et de variance.

2.2.2 Moment d'inertie par rapport à un sous-espace

Moment d'inertie par rapport à un sous-espace vectoriel W est un sous-espace vectoriel de E

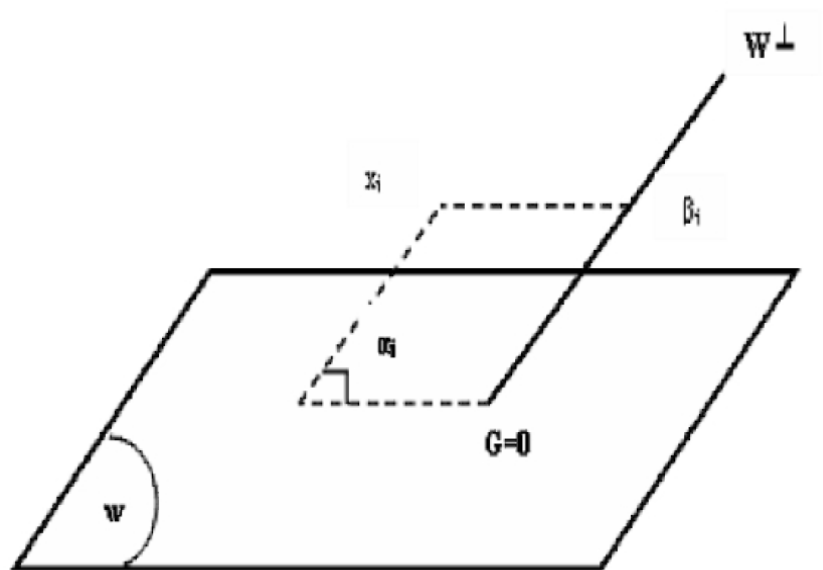


Figure (1)

Considérons la décomposition en somme directe

$E = W \oplus W^\perp$, ou W^\perp est supplémentaire M -orthogonal de W dans E

$$x_i = \alpha_i + \beta_i, \quad \alpha_i \in W, \quad \beta_i \in W^\perp, \quad i = 1, \dots, n.$$

α_i est la projection M -orthogonal de x_i sur W .

$\|x_i - \alpha_i\| = \|\beta_i\|$ mesure la proximité du points x_i et du sous-espace vectoriel W .

Pour mesure la proximité du nuage N et du sous-espace vectoriel W on utilise la quantité

$$I_W = \sum_{i=1}^n p_i \|\beta_i\|^2 = \sum_{i=1}^n p_i \|x_i - \alpha_i\|^2, \quad (2.2.8)$$

appelée moment d'inertie de N par rapport au sous-espace vectoriel W de même

$$I_{W^\perp} = \sum_{i=1}^n p_i \|\alpha_i\|^2 = \sum_{i=1}^n p_i \|x_i - \beta_i\|^2. \quad (2.2.9)$$

$$N \subset W \iff I_W = 0.$$

D'après le théorème de Pythagore

$$\begin{aligned} I_g &= \sum_{i=1}^n p_i \|x_i\|^2 = \sum_{i=1}^n p_i \|\alpha_i\|^2 + \sum_{i=1}^n p_i \|\beta_i\|^2, \\ &= \sum_{i=1}^n p_i \|\alpha_i\|^2 + \sum_{i=1}^n p_i \|x_i - \alpha_i\|^2, \end{aligned}$$

d'où,

$$I_g = I_W + I_{W^\perp}. \quad (2.2.10)$$

Moment d'inertie par rapport à un sous-espace affine

W_1 est un sous espace vectoriel affine de E parallèle au sous espace vectoriel W .

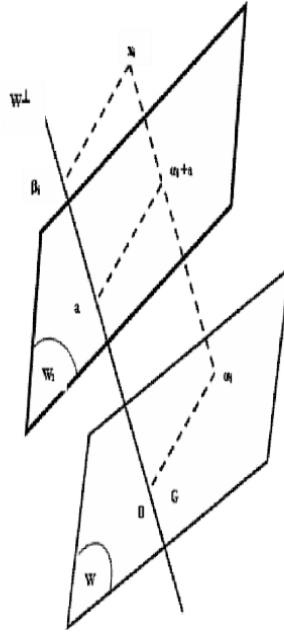


Figure (2)

$$W_1 \cap W^\perp = \{a\}$$

$W_1 = \{Y/Y = X + a; X \in W\}$ de façon condensée

$$W_1 = W + a.$$

$$x_i = \alpha_i + \beta_i = (\alpha_i + a) + (\beta_i - a)$$

$$\alpha_i \in W; \alpha_i + a \in W_1,$$

$$\beta_i \in W^\perp, \beta_i - a \in W^\perp.$$

Le moment d'inertie par rapport à W_1

$$I_{W_1} = \sum_{i=1}^n p_i \|\beta_i - a\|^2,$$

mesure la proximité du nuage N et du sous-espace affine W_1

$$I_{W_1} = \sum_{i=1}^n p_i \|\beta_i\|^2 + \sum_{i=1}^n p_i \|a\|^2 - 2 \sum_{i=1}^n p_i M(\beta_i, a),$$

$$\sum_{i=1}^n p_i = 1 \implies \sum_{i=1}^n p_i \|a\|^2 = \sum_{i=1}^n \|a\|^2,$$

d'après (2.2.1), on a

$$g = \sum_{i=1}^n p_i x_i = 0 \implies \sum_{i=1}^n p_i \alpha_i + \sum_{i=1}^n p_i \beta_i = 0.$$

$$\implies \begin{cases} \sum_{i=1}^n p_i \alpha_i = 0 \\ \sum_{i=1}^n p_i \beta_i = 0 \end{cases} \wedge ,$$

l'égalité

$$\sum_{i=1}^n p_i M(\beta_i, a) = M\left(\sum_{i=1}^n p_i \beta_i, a\right) = 0,$$

permet d'écrire $I_{W_1} = I_W + \|a\|^2$ (theoreme de Hyughens)

$$I_{W_1} \geq I_W.$$

$\|a\|^2$ est la distance entre les sous-espaces W et W_1 .

Théorème 2.2.1 *Tout sous-espace vectoriel W_{k+1} de dimension $(k+1)$, tel que $I_{W_{k+1}}$ soit minimum, contient un sous-espace vectoriel W_k de dimension k tel que I_{W_k} soit minimum.*

Preuve. $\begin{cases} \dim(W_k^\perp) = n - k \\ \dim(W_{k+1}) = k + 1 \end{cases} \implies \dim(W_k^\perp) + \dim(W_{k+1}) = n + 1 > n,$

d'où,

$$\dim(W_k^\perp \cap W_{k+1}) \geq 1,$$

il existe donc v tel que $v \neq 0$, $v \in W_k^\perp \cap W_{k+1}$.

Δ_v : désigne la droite engendrée par le vecteur v .

$$W_{k+1} = \Delta_v + R,$$

où R est le sous-espace vectoriel de dimension k supplémentaire M -orthogonale de Δ_v dans W_{k+1} .

$U = \Delta_v + W_k$ est un sous-espace vectoriel de E de dimension $k+1$.

d'après le théorème $\begin{cases} I_{W_{k+1}^\perp} = I_{\Delta_v^\perp} + I_{R^\perp}, \\ I_{U^\perp} = I_{\Delta_v^\perp} + I_{W_k^\perp}, \end{cases} \Leftrightarrow \begin{cases} I_{W_{k+1}} = I_g - I_{\Delta_v^\perp} - I_{R^\perp}, \\ I_U = I_g - I_{\Delta_v^\perp} - I_{W_k^\perp}, \end{cases}$

I_{W_k} minimum $\Leftrightarrow I_{W_k^\perp}$ maximum $\implies I_{W_k^\perp} \succeq I_{R^\perp}$

d'où, $I_{W_{k+1}} \succeq I_U,$

et

$$I_{W_{k+1}} = I_U \Leftrightarrow I_{W_k} = I_R.$$

Les deux égalité $U = \Delta_v + W_k$ et $I_{W_{k+1}} = I_U$, montrent de plus que qu'à partir de tout espace W_k de dimension k à moment d'inertie minimum on peut construire, en lui adjoignant une droite M -orthogonal, un sous-espace U de dimension $(k + 1)$ à moment d'inertie minimum. Donc la recherche d'un sous-espace vectoriel de dimension k "le plus proche possible de N " revient aux recherches succesives :

1. Recherche de la droite Δ_{u_1} , la plus proche de N .
2. Recherche de la droite Δ_{u_2} , M -orthogonal à Δ_{u_1} , la plus proche de N . □

Moment d'inertie par rapport a une droite Δ_u et par rapport à l'hyperplan M -orthogonal Δ_u^\perp .

Il est naturel, dans la recherche de l'axe Δ_{u_1} le plus proche de N , de s'interroger sur l'expression du moment d'enertie I_{Δ_u} du nuage N par rapport à une droite quelconque Δ_u .

Δ_u : désigne la droite engendrée par le vecteur normé u .

$\Delta_u = (u)$; $\|u\| = \sqrt{M(u)} = 1$ " M est considéré comme une fonction quadratique"

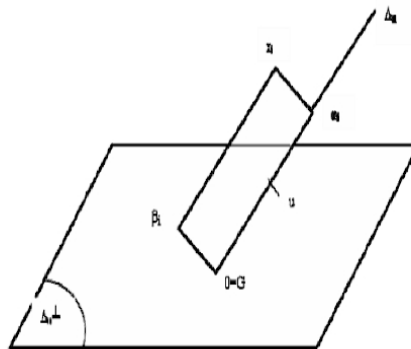


Figure (3)

$$E = \Delta_u + \Delta_u^\perp ; x_i = \alpha_i + \beta_i,$$

$$\alpha_i \in \Delta_u,$$

$$\beta_i \in \Delta_u^\perp,$$

$$I_{\Delta_u} = \sum_{i=1}^n p_i \|\beta_i\|^2,$$

et

$$I_{\Delta_u^\perp} = \sum_{i=1}^n p_i \|\alpha_i\|^2,$$

sont les moments d'inertie du nuage N respectivement par rapport à la droite Δ_u et à l'hyperplan supplémentaire M -orthogonal sur Δ_u^\perp .

On a vu que

$$I_g = \text{Trace}(VM) = I_{\Delta_u^\perp} + I_{\Delta_u}. \quad (2.2.11)$$

$\alpha_i = c_i u_i$, u étant normé.

$c_i = M(x_i, u) = V(x_i)$ où, $V = Mu \in E^*$ est la forme linéaire qui à tout x de E associe la coordonnée de sa projection M -orthogonal sur Δ_u , notons

$$c = X'v = X' \circ Mu \in F,$$

le vecteur de F dont les coordonnées dans la base canonique $\{f_i/i = 1, \dots, n\}$ sont les

$$c_i : c = X'v = \sum_{i=1}^n c_i f_i.$$

À l'axe $\Delta_u \subset E$ on a associé

1. La forme linéaire $v \in E^*$.
2. Le vecteur "caractère" : $c \in F$,

$$I_{\Delta_u^\perp} = \sum_{i=1}^n p_i \|\alpha_i\|^2 = \sum_{i=1}^n p_i c_i^2.$$

2.3 Étude des valeurs et vecteurs propres de $V \circ M$

$$E \xrightarrow{M} E^* \xrightarrow{V} E$$

Propriétés

- Si V est inversible $\Rightarrow V \circ M$ est V^{-1} -symétrique.
 - $V \circ M$ est M -symétrique possède alors des vecteurs propres M -orthogonale deux à deux.
- D'où, le théorème suivant :

Théorème 2.3.1 *Le sous-espace W_k de dimension k est engendré par les k vecteurs propres de $V \circ M$ associés aux k plus grand valeurs propres.*

(i) : Les valeurs propres de $V \circ M$ sont toutes réelles on les note

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p,$$

$$E = E_1 \oplus E_2 \oplus \dots \oplus E_k = \bigoplus \{E_i / i = 1, \dots, k\}.$$

Où $\{E_i / i = 1, \dots, k\}$ désigne l'ensemble de sous-espaces propres associés aux k valeurs propres différents.

(ii) $i \neq j \implies E_i \perp_M E_j$.

Remarque 2.3.1 $\dim(E) = \sum_{i=1}^k \dim(E_i) = p$; $E_i \cap E_j = \{0\}$ si $i \neq j$.

2.3.1 Axes et plans principaux

Tout point $u \in E$ s'écrit de façon unique $u = \sum_{i=1}^p \alpha_i u_i / \{u_i / i = 1, \dots, p\}$.

Nous devons chercher la droite de $E = \mathbb{R}^p$ passant par g maximisant l'inertie du nuage projeté sur cette droite.

Soit u un vecteur porté par cette droite, on sait que le projecteur M -orthogonal sur la droite est alors,

$$P = u \left(u' M u \right)^{-1} u' M.$$

L'inertie du nuage projeté sur cette droite vaut, d'après ce qui précède

$$\begin{aligned} \text{Trace } V \circ M \circ P &= \text{Trace } V \circ M u \left(u' M u \right)^{-1} u' M. \\ &= \frac{1}{u' M u} \text{Trace } V \circ M u u' M. \\ &= \frac{\text{Trace } u' M \circ V \circ M u}{u' M u}. \end{aligned}$$

L'application $M \circ V \circ M$ est appelée application d'inertie du nuage, elle définit la forme quadratique d'inertie qui, à tout vecteur de M -norme 1, associe l'inertie projetée sur l'axe défini par u . L'application d'inertie ne se confond avec l'application de variance-covariance que si $M = I$.

La matrice M étant régulière, on sait que la matrice maximum du quotient $\frac{u' M \circ V \circ M u}{u' M u}$ est atteint si u est vecteur propre de $M^{-1} \circ (M \circ V \circ M)$ associé à sa plus grande valeur propre.

Soit

$$V \circ Mu = \lambda u, \quad (2.3.1)$$

alors, u désigne la base M -orthonormée des vecteurs propres de $V \circ M$.

La matrice $V \circ M$ étant M -symétrique possède alors des vecteurs propres M -orthogonal deux à deux.

Le premier axe principal

Le premier axe principal est l'axe Δ_u engendré par le vecteur normé u qui rend maximum $I_{\Delta_u^\perp}$.

Or,

$$\begin{aligned} \|u\|^2 &= M(u, u) = M\left(\sum_{i=1}^p \alpha_i u_i, \sum_{i=1}^p \alpha_i u_i\right), \\ I_{\Delta_u^\perp} &= M(V \circ Mu, u) = M\left(V \circ M \sum_{i=1}^p \alpha_i u_i, \sum_{i=1}^p \alpha_i u_i\right). \end{aligned}$$

En exploitant la bilinéarité de la forme M et le fait que la base des vecteurs propres est M -orthonormée, on déduit

$$\begin{cases} \|u\|^2 = \sum_{i=1}^p \alpha_i^2 = 1, \\ I_{\Delta_u^\perp} = \sum_{i=1}^p \alpha_i^2 \lambda_i, \end{cases}$$

il s'agit donc de rendre maximum $\sum_{i=1}^p \alpha_i^2 \lambda_i$ sous la contrainte $\sum_{i=1}^p \alpha_i^2 = 1$

$$\begin{cases} \sum_{i=1}^p \alpha_i^2 = 1, \\ \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0, \end{cases} \implies \sum_{i=1}^p \alpha_i^2 \lambda_i \leq \lambda_1,$$

$I_{\Delta_u^\perp}$ est donc maximum sous la contrainte $\|u\| = 1$, si $\alpha_1 = 1$ et $\alpha_2 = \alpha_3 = \dots = \alpha_p = 0$ c'est-à-dire si u coïncide avec l'un des vecteurs propres normés de $V \circ M$ de valeur propre associée λ_1 donc, le premier axe principal est l'axe Δ_{u_1} engendré par u_1 , vecteur propre normé de $V \circ M$ associé à la plus grande valeur propre λ_1

$$V \circ Mu_1 = \lambda_1 u_1, \quad (2.3.2)$$

alors,

$$I_{\Delta_{u_1}^\perp} = \lambda_1.$$

Part d'inertie expliquée par Δ_{u_1}

$$I_{\Delta_{u_1}} = \text{Trace}(VM) - \lambda_1,$$

donc,

$$\begin{aligned} N \subset \Delta_{u_1} &\iff I_{\Delta_{u_1}} = 0 \iff \text{Trace}(VM) = \lambda_1 \\ &\iff \frac{\lambda_1}{\text{Trace}(VM)} = 1 \\ N \not\subset \Delta_{u_1} &\iff I_{\Delta_{u_1}} > 0 \iff \text{Trace}(VM) > \lambda_1 \\ &\iff \frac{\lambda_1}{\text{Trace}(VM)} < 1 \end{aligned}$$

L'indice $\frac{\lambda_1}{\text{Trace}(VM)}$ est la part d'inertie expliquée par le premier axe principale Δ_{u_1} .

$$\begin{array}{ccc} & X & \\ \begin{array}{c} E=R^p \\ M \downarrow \uparrow V \end{array} & \xleftarrow{\quad} & \begin{array}{c} F^* \\ D_p \uparrow \downarrow W \end{array} \\ E^* & \xrightarrow{\quad X' \quad} & F=R^n \end{array}$$

A l'axe principale Δ_{u_1} symbolisant dans E un nouveau caractère, est associé :

– Dans E^* : le premier facteur principal "forme linéaire principale"

$$V_1 = Mu_1. \tag{2.3.3}$$

– Dans F : la première composante principale

$$c^1 = X'V_1 = X' \circ Mu_1, \tag{2.3.4}$$

en effet, $M \circ VV_1 = \lambda_1 V_1$ s'écrit

$$\begin{aligned} M \circ X \circ D_P \circ X'V_1 &= \lambda_1 V_1, \\ M \circ X \circ D_P c^1 &= \lambda_1 V_1, \\ X' \circ M \circ X \circ D_P c^1 &= \lambda_1 X'V_1 = \lambda_1 c^1. \end{aligned}$$

$c_i^1 = M(u_1; x_i) = V_1(x_i)$ est la valeur prise par la première composante principale pour l'individu "i".

Remarque 2.3.2 $\sum_{i=1}^n p_i c_i^1 = D_p(j, c^1) = 0$ où, $j = \begin{pmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} \in F$, la première composante

principale est "centrée".

$\sum_{i=1}^n p_i (c_i^1)^2 = D_p(c^1) = V(v_1) = M \circ V \circ M(u_1) = \lambda_1$, la première composante principale est de variance λ_1 .

Plan principal

Le plan principal W_2 est le sous-espace vectoriel de dimension 2 à moment d'inertie minimum.

- Le plan principal peut être considéré comme engendré par l'axe principal Δ_{u_1} , et l'axe Δ_u , engendré par le vecteur normé u M -orthogonal à u_1 , à moment d'inertie minimum.

Sous-espace principal de dimension k

Le sous-espace principal W de dimension k , sous-espace de E "le plus proche" du nuage N au sens du moment d'inertie, est le sous-espace vectoriel engendré par les k premiers axes principaux $\Delta_{u_1}, \Delta_{u_2}, \dots, \Delta_{u_k}$.

L'axe principal Δ_{u_i} est l'axe engendré par le vecteur propre normé u_i de $V \circ M$ de valeur propre λ_i .

$$\begin{aligned} V \circ M u_i &= \lambda_i u_i, \quad i = 1, \dots, k \\ I_{\Delta_{u_i}^\perp} &= \lambda_i, \\ I_{W^\perp} &= \sum_{i=1}^k I_{\Delta_{u_i}^\perp} = \sum_{i=1}^k \lambda_i, \\ I_W &= \text{Trace}(VM) - \sum_{i=1}^k \lambda_i. \end{aligned}$$

$\frac{\sum_{i=1}^k \lambda_i}{\text{Trace}(VM)}$ c'est la part d'inertie expliquée par le sous-espace W

$$N \subset W \Leftrightarrow I_W = 0 \Leftrightarrow \frac{\sum_{i=1}^k \lambda_i}{\text{Trace}(VM)} = 1.$$

2.3.2 Généralisation

Axes principaux Δ_{u_i} correspondent

$$V \circ M u_i = \lambda_i u_i, \quad \begin{cases} M(u_i, u_i) = 1, \\ M(u_i, u_j) = 0 \text{ si } i \neq j \end{cases}$$

- Dans E^* : les facteurs principaux $V_i = M u_i$. Les facteurs principaux sont vecteurs propres de $M \circ V$ de valeur propre λ_i .
- Dans F : les composantes principales $c_i = X' V_i = X' \circ M u_i$.

Les composantes principales sont vecteurs propres de $X' \circ M \circ X \circ D_P$ de valeur propre λ_i .

Représentation et Interprétation des points

On représentons et interprétons les points dans ce chapitre

3.1 Représentation des individus

Pour faire la représentation des individus dans les plans définis par les nouveaux axes, il suffit de calculer les coordonnées des individus dans les nouveaux axes.

Pour obtenir y_{ik} , coordonnée de l'unité x_i sur l'axe Δ_{u_k} , on projette orthogonalement le vecteur X_i sur l'axe Δ_u

$$y_{ik} = \langle X_i, u_k \rangle_M,$$

si le sous-espace propre de $M \circ V$ de valeur propre $\lambda = 0$ est de dimension $(p - q)$, on a

$$\text{Trace}(VM) = \sum_{i=1}^q \lambda_i.$$

Le nuage de points N est alors situé dans le sous-espace W de dimension q engendré par les q premiers axes principaux : W est appelé support du nuage N .

- Le signe des coordonnées des individus sur un axe n'a donc pas de signification.
- En revanche, la comparaison des signes peut s'interpréter, si deux individus $x_i, x_{i'}$ ont sur un axe Δ_u , le premier une coordonnée positive et le second une coordonnée négative, cela signifie qu'ils s'opposent sur cet axe.

-La qualité des points

$\frac{\lambda_1 + \lambda_2}{\text{Trace}(VM)}$ mesure la qualité globale de la représentation du nuage N dans le plan principale P .

3.1.1 Qualité de la représentation des individus

Lorsque des points de projections des individus sont éloignés sur un axe (ou sur un plan), on peut assurer que les points représentant ces individus sont éloignés dans l'espace.

En revanche, deux individus dont les projections sont proches sur un axe (ou sur un plan), peuvent ne pas être proches dans l'espace.

Pour interpréter correctement la proximité des projections de deux individus sur un plan, il faut donc s'assurer que ces individus sont bien représentés dans le plan.

Un individu x_i soit bien représenté sur un axe (ou un plan ou le sous-espace), il faut que l'angle entre le vecteur X_i et l'axe (ou sur un plan, ou un sous-espace), soit petit. On calcule donc le cosinus de cet angle

$$\cos \alpha = \frac{\|\alpha_i\|_M}{\|X_i\|_M}.$$

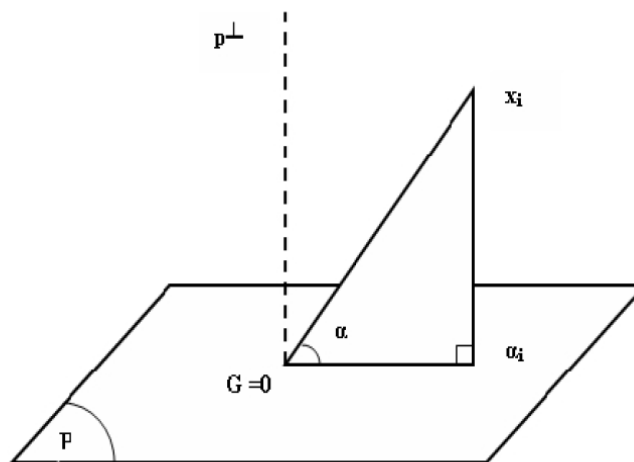


Figure (4)

ou plutôt le carré de ce cosinus. En utilisant le théorème de Pythagore, on peut montrer que le carré du cosinus de l'angle d'un vecteur avec un plan engendré par deux vecteurs orthogonaux, est égale à la somme des carrés des cosinus des angles du vecteur avec chacun des deux vecteurs qui engendrent le plan.

Si le carré du cosinus de l'angle entre X_i et l'axe (ou plan, ou le sous-espace) est proche de 1, alors on pourra dire que l'individu x_i est bien représenté par sa projection sur l'axe (ou plan, ou le sous-espace) et si deux individus sont bien représentés en projection sur un axe (ou un plan, ou un sous-espace) et ont des projections proches dans l'espace.

Le carré du cosinus de l'angle entre X_i et un axe Δ_u de vecteur directeur unitaire a_k est égale à

$$\cos^2 \alpha^2 = \frac{\|\alpha_i\|_M^2}{\|X_i\|_M^2} = \frac{\langle X_i, G_{a_k} \rangle_M}{\|X_i\|_M^2} = \frac{c_{1i}^2}{\|X_i\|_M^2}.$$

Remarque 3.1.1 *Si un individu est très proche du centre de gravité dans l'espace, c.à.d si $\|X_i\|^2$ est très petite, les points représentant cet individu sur un axe (ou un plan, ou un sous-espace) sera bien représenté.*

3.1.2 Interprétation des nouveaux axes en fonction des individus

Lorsqu'on calcule l'inertie I_{Δ_u} portée par l'axe Δ_u , on peut voir quelle est la part de cette inertie due à un individu x_i particulier.

De plus, pour faciliter l'interprétation des axes, on introduit la notion de contribution à l'axe. on définit la contribution de l'individu i à la composante c_k par le quotient

$$\frac{p_i (c_i^k)^2}{\lambda_k}.$$

Où c_i^k est la $i^{\text{ème}}$ coordonnée de la composante principale c^k .

Contribution relative d'un individu à un axe

On peut aussi, pour un individu particulier x_i , donner sa contribution relative à l'inertie portée par cet axe

$$cr(x_i, \Delta_u) = \frac{\frac{1}{n} \langle X_i, \overrightarrow{G_{a_k}} \rangle_M^2}{\lambda_k} = \frac{p_i \langle X_i, \overrightarrow{G_{a_k}} \rangle_M^2}{\lambda_k}.$$

3.2 Représentation des variables

On peut envisager le problème de la représentation des variables de façon complètement symétrique de celui des individus. Les raisonnements se font dans \mathbb{R}^p au lieu dans \mathbb{R}^n , mais dans l'ACP, au de la de la symétrie formelle entre les individus et les variables, on peut

utiliser la dissymétrie liée à la sémantique : les variables n'ont pas la même signification que les individus.

On peut alors faire le raisonnement suivant : on a représenté les individus dans l'espace des anciennes variables, et on a fait un changement de base dans cet espace.

Les nouveaux axes sont des combinaisons linéaires des anciens axes et peuvent donc être considérés comme de nouvelles variables combinaisons linéaires des anciennes, on appelle ces nouvelles variables des composantes principales. Il est alors intéressant de voir comment les anciennes variables sont liées aux nouvelles et pour cela on calcule les corrélations des anciennes variables avec les nouvelles.

La représentation des anciennes variables se fera en prenant comme coordonnées des anciennes variables leurs coefficients de corrélation avec les nouvelles variables, on obtient alors ce que l'on appelle le "cercle des corrélations", dénomination qui vient du fait qu'un coefficient de corrélation variant entre -1 et 1 , les représentations des variables de départ sont des points qui se trouvent à l'intérieur d'un cercle de rayon 1 si on fait la représentation sur un plan

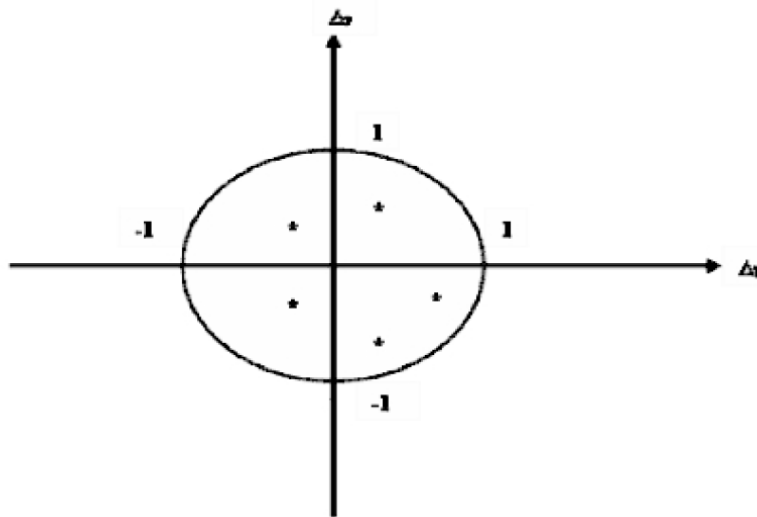


Figure (5)

L'ensemble des composantes principales $\{c^i / i = 1, \dots, q\}$ forme une base D_p -orthogonale du sous-espace vectoriel $X'(E^*)$ de F engendré par l'ensemble des caractères initiaux

$$D_p(c^i, c^j) = 0, i \neq j,$$

$$D_p(c^i) = D_p(c^i, c^i) = \lambda_i.$$

Tout caractère x_i peut être repère par ses coordonnées

$$\begin{aligned} \frac{D_p(x^i, c^j)}{\sqrt{\lambda_i}} &= \frac{Cov(x^i, c^j)}{\sqrt{\lambda_i}}, \\ &= r(x^i, c^j), j = 1, \dots, q \text{ (le coefficient de corrélation),} \end{aligned}$$

dans la base des $\frac{1}{\sqrt{\lambda_i}}c^j$.

De plus, le carée du coefficient de corrélation multiple entre le caractère x^i et les deux caractères c^1 et c^2 est donné par la relation suivante :

$$r^2(x^i, \{c^1, c^2\}) = r^2(x^i, c^1) + r^2(x^i, c^2).$$

A la représentation des points individus dans le plan principal P est associée la représentation des points caractères dans le plan (c^1, c^2) .

Le caractère x^i intervient d'autant plus dans la description du nuage N : Par sa projection dans le plan principal que

$$\|x_i\|^2 = \sum_{j=1}^q \frac{D_p^2(x^i, c^j)}{\lambda_j},$$

par exemple i désigne la projection D_p -orthogonale du caractère x^i sur le plan $(c^1; c^2)$

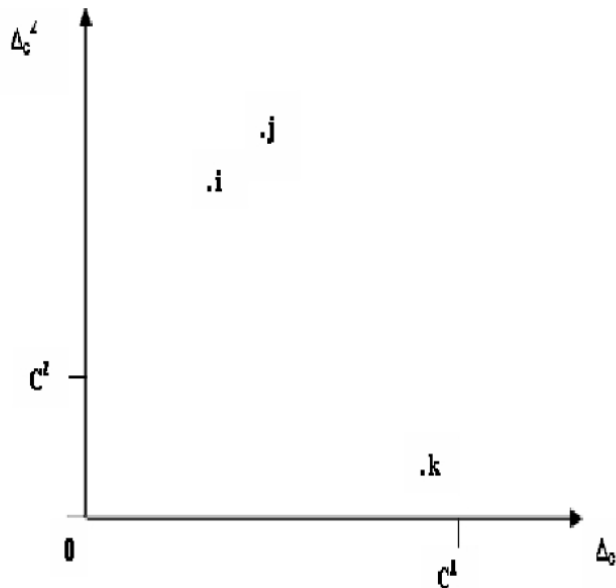


Figure (6)

L'interprétation des composantes principales c^1 et c^2 se fait au vu des proximités entre les projections des caractères x^i et les composantes c^1 et c^2 , dans le plan $(c^1; c^2)$ où les axes D_p -orthogonaux c^1 et c^2 sont représentés orthogonaux.

3.2.1 Qualité de la représentation des variables

Pour les mêmes raisons qui ont poussé à se préoccuper de la qualité de la représentation des individus, il faut se préoccuper de la représentation des variables sur un axe, un plan ou un sous-espace.

Une variable sera d'autant mieux représentée sur un axe que sa corrélation avec la composante principale correspondante est en valeur absolue proche de 1.

Une variable sera bien représentée sur un plan si elle est proche du bord du cercle des corrélations, car cela signifie que le cosinus de l'angle du vecteur joignant l'origine au point représentant la variable avec le plan est, en valeur absolue, proche de 1,*etc.*

3.2.2 Étude des liaisons entre les variables

Sur le graphique du cercle des corrélations, on peut aussi interpréter les positions des anciennes variables les unes par rapport aux autres en termes de corrélations.

Deux points très proches du cercle des corrélations, donc bien représentées dans le plan, seront très corrélés positivement entre elles, si elles sont proches du cercle, mais dans des positions symétriques par rapport à l'origine, elles seront très corrélées négativement.

Deux variables proches du cercle des corrélations et dont les vecteurs qui les joignent à l'origine forment un angle droit, ne seront pas corrélées entre elles. Il faut, pour interpréter correctement ces graphiques des cercles de corrélation, se souvenir qu'un coefficient de corrélation est une mesure de liaison linéaire entre deux variables, et qu'il peut arriver que deux variables très fortement liées aient un coefficient de corrélation nul ou très faible, si leur liaison n'est pas linéaire.

Remarque 3.2.1 *Expression des corrélations entre les caractères x_i et les composantes principales*

$$r(x^i, c^j) = \frac{D_p(x^i, c^j)}{\sqrt{\lambda_i}}.$$

Matriciellement, ceci s'écrit

$$r(x^i, c^j) = \frac{(X^i)' D_p c^j}{\sqrt{\lambda_i}} = \frac{(X^i)' D_p X' M u_j}{\sqrt{\lambda_i}}.$$

3.3 Individus et variables supplémentaires

Il arrive que l'on veuille faire apparaître dans les représentations graphiques certains individus sans qu'ils interviennent dans la détermination des axes.

Cela peut être le cas de nouveaux individus que l'on veut simplement positionner par rapport aux autres sans que les positions de ceux-ci soient influencées par les nouveaux. On dit d'eux qu'ils sont des individus supplémentaires.

Il en est de même pour les variables. On peut, par exemple, vouloir représenter une variable qui dépend de façon synthétique des p variables choisies pour faire l'ACP, afin de mieux comprendre comment cette variable est liée aux anciennes, mais on ne souhaite pas qu'elle intervienne dans la confection des axes car ses liaisons avec les p variables de départ fausseraient la représentation si elle faisait partie intégrante de l'ACP.

Elles sont appelées variables supplémentaires. Pour représenter un individu supplémentaire, il suffit d'exprimer les coordonnées de cet individu dans la nouvelle base des axes principaux. Pour une variable supplémentaire, il suffit de calculer ses coefficients de corrélation empiriques avec les composantes principales.

Application

Présentation élémentaire de l'ACP

4.1 Les données

Considérons les notes (de 0 à 20) obtenus par 9 élèves (math, physique, français, anglais)

\	<i>math</i>	<i>phys</i>	<i>fran</i>	<i>angl</i>
<i>Jean</i>	06.0	06.0	05.0	05.5
<i>Alan</i>	08.0	08.0	08.0	08.0
<i>Anni</i>	06.0	07.0	11.0	09.5
<i>Moni</i>	14.5	14.5	15.5	15.0
<i>Didi</i>	14.0	14.0	12.0	12.0
<i>Andr</i>	11.0	10.0	05.5	07.0
<i>Pier</i>	05.5	07.0	14.0	11.5
<i>Brig</i>	13.0	12.5	08.5	09.5
<i>Evel</i>	09.0	09.5	12.5	12

Nous savons comment analyser séparément chacune de ces 4 variables, soit en faisant un graphique, soit en calculant des résumés numériques nous savons également qu'on peut regarder les liaisons entre 2 variables soit en faisant un graphique du type nuage de points, soit en calculant leur coefficient de corrélation linéaire, voire en réalisant la régression de l'une sur l'autre.

Mais comment faire une étude simultanée des 4 variables, ne serait-ce qu'en réalisant un graphique ?.

La difficulté vient de ce que les individus "élèves" ne sont plus représentés dans un plan "2 dimension", mais dans un espace de dimension 4.

L'objectif de l'ACP est de revenir à un espace de dimension réduite (par exemple ici 2) en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent des données initiales.

4.2 Résultats préliminaires

Tout logiciel fournit la moyenne, le minimum et le maximum de chaque variable. Il s'agit donc, pour l'instant, d'études univariées.

<i>variable</i>	<i>moyenne</i>	<i>min</i>	<i>max</i>	<i>ecart</i>
<i>math</i>	09.67	05.5	14.5	03.37
<i>phys</i>	09.83	06.0	14.5	02.99
<i>fran</i>	10.22	05	15.5	03.47
<i>angl</i>	10.06	05.5	15.0	0.81

Notons au passage la grande homogénéité des 4 variables considérées : même ordre de grandeur pour les moyennes, les écarts-types, les minima et les maxima.

Le tableau suivant est la matrice des corrélations, elle donne les coefficients de corrélation linéaire des variables prises deux à deux.

C'est une succession d'analyses bi variées, constituant un premier pas vers l'analyse multi variée.

\	<i>math</i>	<i>phys</i>	<i>fran</i>	<i>angl</i>
<i>math</i>	1.00	0.98	0.23	0.51
<i>phys</i>	0.98	1.00	0.04	0.65
<i>fran</i>	0.98	0.40	1.00	0.95
<i>angl</i>	0.51	0.65	0.95	1.00

Remarquons que toutes les corrélations linéaires sont positives.

Certaines étant très fortes (0.98 et 0.95), d'autres moyennes (0,65 et 0,95).

D'autre enfin plutôt faibles (0.4 et 0.23).

4.3 Résultats généraux

Continuons l'analyse par celui de la matrice de variances-covariances, matrice de même nature que celle de corrélations, bien que moins "parlante".

La diagonale de cette matrice fournit les variances des 4 variables considérées

\	<i>math</i>	<i>phys</i>	<i>fran</i>	<i>angl</i>
<i>math</i>	11.39	9.92	2.66	04.82
<i>phys</i>	09.92	8.94	04.12	05.48
<i>fran</i>	2.66	4.12	12.06	09.29
<i>angl</i>	4.82	5.48	09.29	07.91

Les valeurs propres données ci-dessous sont celles de la matrice des variances-covariances

<i>facteur</i>	<i>val.pr</i>	<i>pct.var</i>	<i>pct.cum</i>
1	28.23	00.70	00.70
2	12.03	00.30	01.00
3	00.03	00.00	01.00
4	00.01	00.00	01.00
	40.30	01.00	

4.4 Interprétation

Chaque ligne du tableau ci-dessus correspond à une variable virtuelle (voilà les facteurs) dont la colonne *val.pr* (valeur propre) fournit la variance (en fait, chaque valeur propre représente la variance du facteur correspondant). La colonne *pct.var*, où pourcentage de variance, correspond au pourcentage de variance de chaque ligne par rapport au total. La colonne *pct.cum* Représente le cumul de ces pourcentages.

Additionnons maintenant les variances des 4 variables initiales (diagonale de la matrice des variances-covariances)

$11.39 + 8.94 + 12.06 + 7.91 = 40.30$. La dispersion totale des individus considérés, en dimension 4, est ainsi égale à 40.30.

Additionnons par ailleurs les 4 valeurs propres obtenues : $28.23 + 12.03 + 0.03 + 0.01 = 40.30$.

Le nuage de points en dimension 4 est toujours le même et sa dispersion globale n'a pas changé.

Il s'agit d'un simple changement de base dans un espace vectoriel. C'est la répartition de cette dispersion, selon les nouvelles variables que sont les facteurs, ou composantes principales, qui se trouvent modifiée.

Les 2 premiers facteurs restituent à eux seuls la quasi-totalité de la dispersion du nuage, ce qui permet de négliger les 2 autres.

Par conséquent, les graphiques en dimension 2 présentés ci-dessous résument presque parfaitement la configuration réelle des données qui se trouvent en dimension 4. L'objectif (résumé pertinent des données en petite dimension) est donc atteint.

4.5 Résultats sur les variables

Le résultat fondamental concernant les variables est le tableau des corrélations variables facteurs. Il s'agit des coefficients de corrélation linéaire entre les variables initiales et les facteurs.

Ce sont ces corrélations qui vont permettre de donner un sens aux facteurs (de les interpréter)

<i>facteur</i>	V_1	V_2	V_3	V_4
<i>math</i>	0.60	0.60	0.01	-0.02
<i>phys</i>	0.60	0.40	-0.03	0.02
<i>fran</i>	0.50	-0.66	-0.02	-0.01
<i>angl</i>	0.47	-0.33	0.05	0.01

Les deux premières colonnes de ce tableau permettent, tout d'abord, de réaliser le graphique des variables donné ci-dessous. Mais, ces deux colonnes permettent également de donner une signification aux facteurs (donc aux axes des graphiques).

On notera que les deux dernières colonnes ne seront pas utilisées puisqu'on ne retient que deux dimensions pour interpréter l'analyse.

4.6 Interprétation

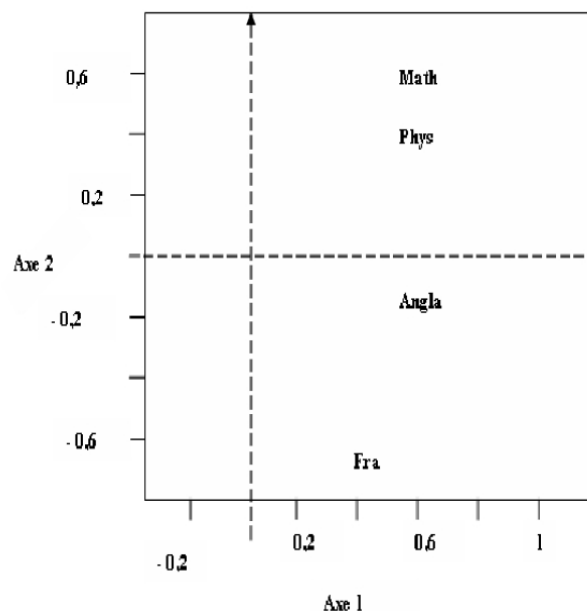
Ainsi, on voit que le premier facteur est corrélé positivement, et assez fortement, avec chacune des 4 variables initiales.

Plus un élève obtient de bonnes notes dans chacune des 4 disciplines, plus il a un score élevé sur l'axe 1 ; réciproquement, plus ses notes sont mauvaises, plus son score est négatif.

En ce qui concerne l'axe 2, il oppose, d'une part, le français et l'anglais (corrélations positives), d'autre part, les mathématiques et la physique (corrélations négatives). Il s'agit donc d'un axe d'opposition entre disciplines littéraires et disciplines scientifiques, surtout marqué par l'opposition entre le français et les mathématiques.

Cette interprétation peut être précisée avec les graphiques et tableaux relatifs aux individus

que nous présentons maintenant.



Représentation des variables

4.7 Résultats sur les individus

Le tableau ci-dessous contient tous les résultats importants sur les individus

/	<i>poids</i>	<i>fact1</i>	<i>fact2</i>	<i>contg</i>	<i>cont1</i>	<i>cont2</i>
<i>Jean</i>	0.11	-8.61	1.42	20.99	29.19	01.83
<i>Alan</i>	0.11	-3.88	0.5	05.92	05.92	00.23
<i>Anni</i>	0.11	-3.21	-3.50	04.06	04.06	11.11
<i>Moni</i>	0.11	9.85	-0.5	38.19	38.19	00.33
<i>Didi</i>	0.11	6.41	2.17	16.15	16.15	03.87
<i>Andr</i>	0.11	-3.03	4.91	03.62	03.62	22.37
<i>Pier</i>	0.11	-1.03	-6.40	00.41	00.41	37.56
<i>Brig</i>	0.11	1.95	4.18	01.05	01.05	16.29
<i>Evel</i>	0.11	1.55	-2.66	00.95	00.95	06.41

On notera que chaque individu représente 1 élément sur 9, d'où un poids (une pondération) de $\frac{1}{9} = 0,11$, ce qui est fourni par la première colonne du tableau ci-dessus.

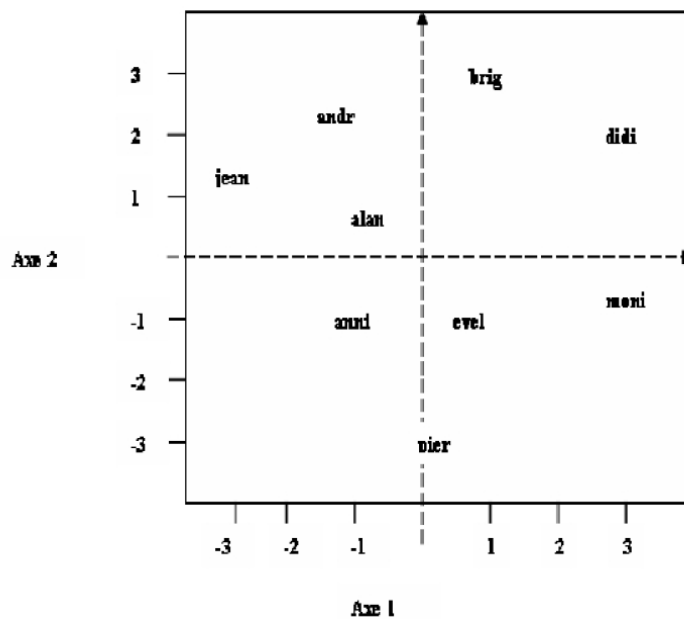
Les 2 colonnes suivantes fournissent les coordonnées des individus (les élèves) sur les deux premiers axes (les facteurs) et ont donc permis de réaliser le graphique des individus.

4.8 Interprétation

On peut ainsi voir que l'axe 1 représente le résultat d'ensemble des élèves (si on prend leur score ou coordonnée sur l'axe 1, on obtient le même classement que si on prend leur moyenne générale).

Par ailleurs, l'élève "le plus haut" sur le graphique, celui qui a la coordonnée la plus élevée sur l'axe 2, est Pierre dont les résultats sont les plus contrastés en faveur des disciplines littéraires (14 et 11,5 contre 7 et 5,5).

C'est exactement le contraire pour André qui obtient la moyenne dans les disciplines scientifiques (11 et 10) mais des résultats très faibles dans les disciplines littéraires (7 et 5,5).



Représentation des individus

On notera que Monique et Alain ont un score voisin de 0 sur l'axe 2 car ils ont des résultats très homogènes dans les 4 disciplines (Mais à des niveaux très distincts, ce qu'a déjà révélé l'axe 1).

Les 3 colonnes suivantes du tableau fournissent des contributions des individus à diverses dispersions : *CONT1* et *CONT2* donnent les contributions des individus à la variance selon les axes 1 et 2 (rappelons que c'est la variance qui caractérise la dispersion); *CONTG* les contributions à la dispersion en dimension 4 (il s'agit de ce que l'on appelle l'inertie du nuage

des élèves ; la notion d'inertie généralise celle de variance en dimension quelconque, la variance étant toujours relative à une seule variable).

Ces contributions sont fournies en pourcentages (chaque colonne somme à 100) et permettent de repérer les individus les plus importants au niveau de chaque axe (où du nuage en dimension 4).

Elles servent en général à affiner l'interprétation des résultats de l'analyse.

Ainsi, par exemple, la variance de l'axe 1 vaut 28.23 (première valeur propre).

On peut la retrouver en utilisant la formule de définition de la variance

$$\text{var}(c^1) = \frac{1}{9} \sum_{i=1}^q (c_i^1)^2.$$

Il faut noter que, dans une ACP, les variables étant centrées, il en va de même pour les facteurs ; ainsi, la moyenne de c^1 est nulle et n'apparaît pas dans la formule de la variance.

La coordonnée de Jean (le premier individu du fichier) sur l'axe 1 vaut $c_1^1 = 8,61$; sa contribution est donc

$$\frac{\frac{1}{9}(-8.61)}{28.23} \times 100 = 29.19\%.$$

A lui seul, cet individu représente près de 30% de la variance il est prépondérant (au même titre que Monique) dans la définition de l'axe 1 ; cela provient du fait qu'il a le résultat le plus faible, Monique ayant, à l'opposé, le résultat le meilleur.

Enfin, les 2 dernières colonnes du tableau sont des cosinus carrés qui fournissent la (qualité de la représentation) de chaque individu sur chaque axe. Ces quantités s'additionnent axe par axe, de sorte que, en dimension 2, Evelyne est représentée à 98% (0,25 + 0,73), tandis que les 8 autres individus le sont à 100%.

Lorsqu'on considère les données initiales, chaque individu (chaque élève) est représenté par un vecteur dans un espace de dimension 4 (les éléments ou coordonnées de ce vecteur sont les notes obtenues dans les 4 disciplines).

Lorsqu'on résume les données en dimension 2, et donc qu'on les représente dans un plan, chaque individu est alors représenté par la projection du vecteur initial sur le plan en question.

Le cosinus carré relativement aux deux premières dimensions (par exemple, pour Evelyne, 0,98 ou 98%) est celui de l'angle formé par le vecteur initial et sa projection dans le plan.

Plus le vecteur initial est proche du plan, plus l'angle en question est

petit et plus le cosinus, et son carré, sont proches de 1 (ou de 100%), la représentation est alors très bonne.

Au contraire, plus le vecteur initial est loin du plan, plus l'angle en question est grand (proche de 90 degrés) et plus le cosinus, et son carré, sont proches de 0 (ou de 0%) : la représentation est alors très mauvaise.

On utilise les carrés des cosinus, parce qu'ils s'additionnent suivant les différentes dimensions.

CONCLUSION

L'exécution de l'ACP sur des variables quantitatives donne plusieurs solutions :

- La représentation des n individus x_1, x_2, \dots, x_n dans un sous-espace de F_k de dimension k à travers la recherche de k nouvelles variables. Ces nouvelles variables combinaisons linéaires des p variables initiales qui feront perdre le moins d'information possible, appelées composantes principales.
- L'ACP permet d'explorer les liaisons entre les variables et les ressemblances entre individus. Cette exploration donne une classification des points.

Finalement, nous intéressons à généraliser cette notion de l'ACP au cas fonctionnel ou l'ACP sous contraintes.

Bibliographie

- [1] **Anderson T.W**; *"Introduction to multivariate statistical analysis"*. New York, Wiley 1958.
- [2] **Ben Zecri.J.P**; *"L'analyse des données "(tome 2)*, Paris -Dunod 1973.
- [3] **F.CAILLIEZ**; *"Introduction à l'analyses des données"*. Smash, Paris, 1976.
- [4] **Gilbert Saporta**; *"Théorie et méthodes de la statistique"* ; Editions Technip, Paris, 1978.
- [5] **I.T.Jolliffe**; *"Principal Component Analysis"*; Springer Series in Statistics, New York, 2002.
- [6] **J.M.BOUROCHE**; *"Analyse des données en marketing"*. Masson, Paris, 1977.
- [7] **Jean-Marie Monier**; *"Cours de mathématiques , Algèbre et géométrie"*, Dunod, Paris 2007.
- [8] **Philippe G. Ciarlet**; *"Introduction à l'analyse numérique matricielle et à l'optimisation"*, Dunod, Paris, 1998.