Département MIDO, M1

Analyse des données

M1

Patrice Bertrand et Denis Pasquignon

(version du 3 septembre 2025)

Patrice Bertrand et Denis Pasquignon
CEREMADE, Université Paris-Dauphine, 75016 Paris, France.
E-mail : `denis.pasquignon@ceremade.dauphine.fr`

Ce polycopié a été créé avec LaTeX; pour la mise en forme, nous avons adapté des fichiers de style fournis par la Société Mathématique de France, notamment la classe `smfbook`.

ANALYSE DES DONNÉES

Patrice Bertrand et Denis Pasquignon

TABLE DES MATIÈRES

INTRODUCTION

L'analyse des données (AD), et plus généralement la fouille des données (FD), est constituée d'un ensemble de techniques qui ont pour but de déterminer les structures possédées par l'ensemble des données. Ces structures peuvent être de nature descriptive (partition, hiérarchie, plan factoriel,...) ou explicative (arbre de décision, analyse factorielle discriminante,...). L'analyse de données peut être considérée comme une science expérimentale : propriétés démontrées après avoir été observées, indice empirique pour l'interprétation des résultats, codages établis de façon heuristique.

Par ailleurs, les premiers résultats fournis par une analyse factorielle sont généralement évidents, alors que les résultats suivants ne sont pas triviaux et sont souvent intéressants.

Les données peuvent se présenter sous différentes formes : tableaux individus $\times$ variables (dans un but descriptif, l'interprétation établira des liens entre variables et groupes d'individus qui se ressemblent selon ces variables), tableaux de distances (représentation des individus dans un plan, sur une droite, etc ou partitionement de l'ensemble des individus), tableaux de contingence (ces tableaux croisent les ensembles de modalités de deux caractères qualitatifs), tableaux de présence-absence (0/1), tableaux de notes, tableaux de pourcentage...

Les techniques d'analyse de données se différencient non seulement par les outils mathématiques utilisés (algèbre linéaire dans le cas de l'analyse factorielle, théorie des graphes et combinatoire pour certaines méthodes de classification) mais aussi par les buts poursuivis qui peuvent être un but descriptif ou un but prévisionnel. Le but descriptif consiste à essayer d'obtenir une représentation simplifiée aussi proche que possible des données initiales, le but prévisionnel consiste à expliquer et prévoir une ou plusieurs variables en fonction d'autres variables. Dans ce cours, nous présenterons les techniques suivantes :

- Analyse en composantes principales (ACP) : rechercher des axes d'inertie d'un système de points affectés de poids, ce qui permet d'en déduire des sous-espaces de dimensions réduites sur lesquels la projection des points est la moins déformante.
- Analyse des correspondances (AC) : double ACP ayant un but à la fois descriptif et prévisionnel (étude de liens existants entre lignes et colonnes d'un tableau).

CHAPITRE 1

NUAGES DE POINTS

1. Tableau de données

On observe p variables quantitatives mesurées sur un échantillon de taille n. Les données sont rassemblées en un tableau ou matrice de n lignes et p colonnes. On note X ce tableau de données, son terme général x_i^j, situé à la $i$ème et $j$ème colonne, désigne la valeur prise par le $i$ème individu pour la variable j.

On note $I = [\![1, n]\!]$ et $J = [\![1, p]\!]$ qui sont les ensembles d'indices désignant respectivement les n individus et les p variables.

$$X = (x_i^j)_{i \in I, j \in J} \in \mathcal{M}_{n,p}(\mathbb{R}).$$

Ainsi les valeurs prises par la variable x^j pour les n individus se lisent sur la $j$ème colonne et les valeurs prise par l'individu i pour les p variables se lisent sur la $i$ème ligne. On note x^j la $j$ème variable et x_i le $i$ème individu :

$$\forall (i,j) \in [\![1, n]\!] \times [\![1, p]\!], \quad x^j = \begin{pmatrix} x_1^j \\ \vdots \\ x_n^j \end{pmatrix} \in \mathbb{R}^n \text{ et } x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^p \end{pmatrix} \in \mathbb{R}^p.$$

Ainsi

$$X = [x^1, \cdots, x^p] = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}.$$

2. Nuages des individus et nuages des variables

On munit $\mathbb{R}^p$ de la base canonique, O étant l'origine de ce repère, on peut alors associer à chaque individu i le point M_i tel que

$$\forall i \in [\![1, n]\!], \quad \overrightarrow{OM_i} = x_i.$$

Chaque axe représente une variable. L'ensemble des points $\mathcal{M}_X = \{M_i,\ 1 \leq i \leq n\}$ est appelé le nuage des individus et $\mathbb{R}^p$ est l'espace des individus.

De même, on munit $\mathbb{R}^n$ de la base canonique, on peut alors associer à chaque variable le point N^j tel que

$$\forall j \in [\![1, p]\!], \quad \overrightarrow{ON^j} = x^j.$$

Chaque axe représente un individu. L'ensemble des points $\mathcal{N}_X = \{N^j,\ 1 \leq j \leq p\}$ est appelé le nuage des variables et $\mathbb{R}^n$ est l'espace des variables.

Les ensembles $\mathbb{R}^n$ et $\mathbb{R}^p$ sont considérés comme des espaces affines. Dans l'annexe A, on rappelle les principales notions à connaitre pour ce cours.

3. Métrique dans un espace euclidien, distance dans un espace affine

Définition 1.1 – Métrique

On considère E un espace euclidien de dimension n, où n est un entier naturel non nul, muni d'un produit scalaire noté $< .,. >$. Soit $\mathcal{B} = (e_1, \cdots, e_n)$ une base de E. La matrice du produit scalaire dans la base $\mathcal{B}$ est la matrice M carrée d'ordre n et de terme courant

$$\forall (i,j) \in [\![1,n]\!]^2, \quad m_{i,j} =< e_i, e_j > .$$

On appellera M la métrique de E.

Proposition 1.2

Soit E un espace euclidien de dimension n, où n est un entier naturel non nul. Soit M la métrique de E dans la base $\mathcal{B} = (e_1, \cdots, e_n)$. La matrice M est symétrique, définie et positive. De plus soit u et v deux vecteurs de E, on note U et V les matrices représentant u et v dans $\mathcal{B}$, on a

$$< u, v >= U'MV = V'MU.$$

On note le produit scalaire avec M en indice pour indiquer la métrique utilisée : $< u, v >_M$. Réciproquement, toute matrice d'ordre n symétrique, définie positive permet de définir un produit scalaire dans E en utilisant la relation précédente.

Remarque 1.3. — La base $\mathcal{B}$ est orthonormée si et seulement si la matrice M est égale à l'identité.

Démonstration. — La matrice M est symétrique car le produit scalaire est symétrique. Par ailleurs, on a

$$u = \sum_{i=1}^{n} u_i e_i \quad v = \sum_{j=1}^{n} v_j e_j.$$

Donc d'après les propriétés du produit scalaire de bilinéarité et symétrie, on a

$$< u, v >= \sum_{i=1}^{n} \sum_{j=1}^{n} u_i v_j m_{i,j} = U'MV = V'MU.$$

On en déduit que pour toute matrice colonne U

$$U'MU = \|u\|^2 \geq 0,$$

donc M est positive. Enfin $U'MU = 0$ implique $u = 0$ soit $U = 0$ donc M est définie positive.

Réciproquement soit M matrice d'ordre n symétrique, définie positive, on se place dans $\mathbb{R}^n$ muni de la base canonique, pour tout vecteur u et v de $\mathbb{R}^n$, on peut associer les matrices U et V respectivement à u et v, on pose alors

$$\phi(u, v) = U'MV.$$

On vérifie que ϕ est un produit scalaire sur $\mathbb{R}^n$.

□

Pour étudier la proximité entre deux individus d'un même nuage de points, on introduit une distance notée d entre les individus.

Définition 1.4 – Distance

Soit $\mathcal{E}$ un espace affine associé à l'espace euclidien E. Soit M_1 et M_2 deux points de $\mathcal{E}$, la distance entre M_1 et M_2 est égale à la norme du vecteur joignant ces deux points :

$$d(M_1, M_2) = \|\overrightarrow{M_1M_2}\|.$$

Dans toute la suite de ce cours, nous noterons M la métrique de l'espace des individus $\mathbb{R}^p$. Si l'on suppose que la matrice M est diagonale $M = diag(m_1, \cdots, m_p)$, alors la distance entre les individus i et i' est donnée par

$$d(x_i, x_{i'}) = \sqrt{\sum_{j=1}^{p} m_j (x_i^j - x_{i'}^j)^2}.$$

Par ailleurs chaque individu i est muni d'une masse, appelée aussi poids, notée p_i et telle que

$$\forall i \in I, \;\; p_i > 0 \text{ et } \sum_{i=1}^{n} p_i = 1.$$

On note D_p la matrice diagonale définie par

$$D_p = diag(p_1, \cdots, p_n).$$

En général, les poids sont tous égaux à $1/n$, mais ce n'est pas toujours le cas comme par exemple en Analyse des Correspondances. Alors l'espace des variables $\mathbb{R}^n$ est muni de la métrique D_p. Par conséquent la distance entre deux variables est

$$d(x^j, x^{j'}) = \sqrt{\sum_{i=1}^{n} p_i (x_i^j - x_i^{j'})^2}.$$

4. Centre de gravité du nuage $\mathcal{M}_X$

Définition 1.5 – Centre de gravité

Le centre de gravité du nuage des individus M_i affecté du poids p_i est le point G tel que

$$G = \sum_{i=1}^{n} p_i M_i.$$

La $j$ème coordonnée de G est donnée par

$$g_j = \sum_{i=1}^{n} p_i x_i^j = \overline{x^j}.$$

Ainsi g_j est la moyenne de la variable x^j et les coordonnées du point G sont les p moyennes des p variables.

Proposition 1.6

On note 1_n le vecteur de $\mathbb{R}^n$ dont toutes les coordonnées sont égales à 1, on a

$$g = \overrightarrow{OG} = \begin{pmatrix} g_1 \\ \vdots \\ g_p \end{pmatrix} = X' D_p 1_n.$$

Démonstration. — Puisque $D_p 1_n = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}$, on en déduit que

$$\forall j \in [\![1,p]\!], \quad g_j = x^{j'} D_p 1_n,$$

ce qui donne le résultat. □

Remarque 1.7. — Lorsque l'on se place dans $\mathbb{R}^n$ muni de la métrique D_p, le vecteur 1_n est unitaire, soit $P_{\text{Vect}(1_n)}$ la projection orthogonale sur Vect 1_n, alors on a

$$\forall u \in \mathbb{R}^n, \quad P_{\text{Vect}(1_n)}(u) =< u, 1_n > 1_n = U' D_p 1_n.$$

On en déduit que g_j est l'abscisse de la projection orthogonale pour la métrique D_p de x^j sur $\text{Vect}(1_n)$.

Il est naturel de centrer le nuage des individus sur le centre de gravité G ce qui revient à construire un nouveau tableau Y tel que

$$\forall (i,j) \in [\![1,n]\!] \times [\![1,p]\!], \quad y_i^j = x_i^j - \overline{x^j},$$

soit

$$\forall i \in [\![1,n]\!], \quad y_i = M_i - G.$$

Ainsi dans ce nouveau tableau de données, toutes les variables y^j, $1 \leq j \leq p$, sont de moyennes nulles.

Proposition 1.8

On a

$$Y = X - 1_n g'$$

Par ailleurs

$$y^j = x^j - g_j 1_n = (Id - P_{\text{Vect}(1_n)})(x^j),$$

ce qui signifie que y^j est la projection de x^j sur l'hyperplan orthogonal à 1_n.

5. Support des nuages

Définition 1.9 – Support d'un nuage

On appelle support d'un nuage le plus petit sous-espace affine contenant les points du nuage. On note

$$S_X = supp(\mathcal{M}_X) \text{ et } S_Y = supp(\mathcal{M}_Y).$$

Puisque le nuage $\mathcal{M}_Y$ est centré, le support S_Y contient l'origine et est assimilé à un sous-espace vectoriel

$$S_Y = \text{Vect}(y_1, \cdots, y_n) = \operatorname{Im} Y'.$$

Proposition 1.10

Soit r le rang de la matrice Y, alors la dimension de S_Y est égale à r, le rang de Y.

Démonstration. — Une matrice et sa transposée ont même rang donc r est égale à la dimension de l'espace vectoriel générée par les vecteurs lignes soit la famille $(y_i)_{1 \leq i \leq n}$. □

6. Matrice Variance

Définition 1.11 – Matrice Variance

La matrice variance, notée V, des p variables pour les n individus est une matrice carré d'ordre p et de terme courant $v_{j,j'}$ donné par

$$\forall (j,j') \in [\![1,p]\!]^2, \quad v_{j,j'} = \mathrm{Cov}(x^j, x^{j'}) = \sum_{i=1}^{n} p_i (x_i^j - g_j)(x_i^{j'} - g_{j'}) =< y^j, y^{j'} >_{D_p} .$$

Proposition 1.12

En notation matricielle, on a

$$V = Y' D_p Y.$$

ou encore

$$V = (X - 1_n g')' D_p (X - 1_n g') = X' D_p X - g g'.$$

La matrice V se décompose de la manière suivante

$$V = \sum_{i=1}^{n} p_i \, y_i \, y_i'.$$

Démonstration. — C'est une preuve directe, soit $(j,j') \in [\![1,p]\!]^2$ le terme courant de $Y' D_p Y$ est

$$(Y' D_p Y)_{jj'} = \sum_{i=1}^{n} y_i^j p_i y_i^{j'} = v_{j,j'}.$$

D'où l'égalité. En remplaçant Y en fonction de X, on obtient

$$V = (X - 1_n g')' D_p (X - 1_n g').$$

Puis on développe cette expression donc

$$V = X' D_p X - g 1_n' D_p X - X' D_p 1_n g' + g 1_n' D_p 1_n g'.$$

Or 1_n est un vecteur unitaire pour la métrique D_p et par symétrie du produit scalaire, on a

$$1_n' D_p 1_n = 1, \quad \text{et } g 1_n' D_p X = X' D_p 1_n g' = g g'.$$

Il reste alors

$$V = X' D_p X - g g'.$$

La décomposition se prouve directement aussi : soit $(j,j') \in [\![1,p]\!]^2$ le terme courant de $\sum_{i=1}^{n} p_i \, y_i \, y_i'$ est

$$(\sum_{i=1}^{n} p_i \, y_i \, y_i')_{jj'} = \sum_{i=1}^{n} p_i (y_i \, y_i')_{j,j'} = \sum_{i=1}^{n} p_i \, y_i^j y_i^{j'} = v_{j,j'}.$$

D'où l'égalité. □

Remarque 1.13. — Si la matrice V est définie positive, elle fournit une métrique sur $\mathbb{R}^p$, métrique induite par D_p et Y. Si V n'est pas régulière, on aura seulement une pseudo métrique.

Proposition 1.14

Le rang de la matrice V est égal au rang de Y.

Démonstration. — On montre directement que le noyau de Y est égal au noyau de V. En effet soit u de $\operatorname{Ker} Y$, alors $Yu = 0$. donc $Vu = Y'D_pYu = 0$ ainsi $u \in \operatorname{Ker} V$. Réciproquement si $u \in \operatorname{Ker} M$, alors $Y'D_pYu = 0$ soit $u'Y'D_pYu = 0$ ce qui donne $\|Yu\|_{D_p} = 0$ donc $Yu = 0$ et $u \in \operatorname{Ker} Y$.

Ensuite on applique le théorème du rang et on conclut

$$rang(Y) = p - dim(\operatorname{Ker} Y) = p - dim(\operatorname{Ker} V) = rang(V).$$

□

Proposition 1.15

Soient $u = \begin{pmatrix} u_1 \\ \vdots \\ u_p \end{pmatrix}$ *et* $v = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix}$ *de* $\mathbb{R}^p$*, on définit deux nouvelles variables* z *et* t *par*

$$z = \sum_{j=1}^{p} u_j x^j \ \textit{ et } t = \sum_{j=1}^{p} v_j x^j.$$

Alors la covariance entre z *et* t *est donnée par*

$$\operatorname{Cov}(z, t) = u'Vv.$$

Si la matrice V *est définie positive,* V *définit une métrique pour laquelle la covariance entre* z *et* t *est le produit scalaire entre les vecteurs* z *et* t *et la variance de la variable* z *est le carré de la norme de* z *soit*

$$\operatorname{Cov}(z, t) = < u, v >_V \ \textit{ et } \ \mathrm{V}(z) = \|u\|_V^2.$$

Démonstration. — On a

$$\begin{aligned} \operatorname{Cov}(z, t) &= \operatorname{Cov}(\sum_{j=1}^{p} u_j x^j, \sum_{j'=1}^{p} v_{j'} x^{j'}), \\ &= (\sum_{j=1}^{p} \sum_{j'=1}^{p} u_j v_{j'} \operatorname{Cov}(x^j, x^{j'}), \text{ par bilinéarité de la covariance,} \\ &= u'Vv. \end{aligned}$$

□

7. Effet d'une transformation linéaire A du nuage des individus

On considère une application linéaire f de l'espace des individus $\mathbb{R}^p$ dans $\mathbb{R}^s$ où s est un entier naturel. Le nuage de points $\mathcal{M}_X$ est alors transformé en un autre nuage noté $\mathcal{M}_Z$. Si le paramètre s est inférieur à p, le nouveau nuage de points $\mathcal{M}_Z$ évolue alors dans un espace de dimension plus faible.

On note A la matrice qui représente l'application linéaire f. Ainsi A est une matrice de format $s \times p$. On note

$$\mathcal{M}_Z = \{z_1, \cdots, z_n\} \text{ avec } \forall i \in [\![1, n]\!], \ \ z_i = f(x_i) = Ax_i.$$

On obtient ainsi une nouvelle matrice Z dont les lignes sont les $z_1, \cdots, z_n$ soit

$$Z' = AX' \text{ donc } Z = XA'.$$

Proposition 1.16

Le centre de gravité de $\mathcal{M}_Z$ affectés des poids $p_1, \cdots, p_n$ noté g_Z est

$$g_Z = Ag.$$

La matrice de variance de Z noté $Var(Z)$ est

$$Var(Z) = Var(XA') = AVA'.$$

Démonstration. — On a

$$g_Z = \sum_{i=1}^{n} p_i z_i = \sum_{i=1}^{n} p_i A x_i = Ag.$$

On note Z_c la matrice centrée

$$Z_c = Z - 1_n g_Z' = XA' - 1_n g' A' = YA',$$

donc

$$Var(Z) = Z_c' D_p Z_c = AY' D_p YA' = AVA'.$$

□

8. Inerties

Dans tout ce chapitre, on se place dans $\mathbb{R}^p$ considéré comme un espace euclidien muni d'une métrique notée M.

Définition 1.17 – Inertie par rapport à un point

Soit A un point, l'inertie du nuage $\mathcal{M} = (x_i)_{1 \leq i \leq n}$ par rapport au point A est

$$I_A(\mathcal{M}) = \sum_{i=1}^{n} p_i \|x_i - A\|_M^2.$$

Si $A = G$ le centre de gravité , $I_G(\mathcal{M})$ est appelée inertie totale du nuage :

$$I_T(\mathcal{M}) = I_G(\mathcal{M}).$$

Si l'on suppose que $M = diag(m_1, \cdots, m_n)$ alors

$$I_T(\mathcal{M}) = \sum_{i=1}^{n} p_i \|y_i\|_M^2 = \sum_{i=1}^{n} p_i \sum_{j=1}^{p} m_j (y_i^j)^2 = \sum_{j=1}^{p} m_j \, \mathrm{V}(y^j),$$

où $\mathrm{V}(y^j)$ représente la variance de y^j. L'inertie totale est ainsi la somme pondérée des variances des variables initiales, elle mesure la dispersion du nuage autour du centre de gravité.

Proposition 1.18 – Théorème de Huyghens

On a

$$I_A(\mathcal{M}) = I_T(\mathcal{M}) + \|A - G\|_M^2.$$

Démonstration. —

$$\begin{aligned} \|x_i - A\|_M^2 &= \|x_i - G + G - A\|_M^2, \\ &= \|x_i - G\|_M^2 + \|G - A\|_M^2 + 2 < x_i - G, G - A >_M . \end{aligned}$$

On en déduit que

$$\begin{aligned} I_A(\mathcal{M}) &= \sum_{i=1}^{n} p_i \|x_i - A\|_M^2, \\ &= \sum_{i=1}^{n} p_i \|x_i - G\|_M^2 + \sum_{i=1}^{n} p_i \|G - A\|_M^2 + 2 < \sum_{i=1}^{n} p_i (x_i - G), G - A >_M, \\ &= I_T(\mathcal{M}) + \|G - A\|_M^2. \end{aligned}$$

□

Définition 1.19 – Inertie par rapport à un sous-espace affine

Soit $\mathcal{E}$ un sous-espace affine de $\mathbb{R}^p$ et E le sous-espace vectoriel associé. Soit A un point de $\mathcal{E}$ et B un point de $\mathbb{R}^p$, la distance de B à $\mathcal{E}$ est

$$d_M(B, \mathcal{E}) = \|(Id - P_E)(\overrightarrow{AB})\|_M,$$

où P_E est la projection orthogonale sur E.
On appelle inertie du nuage $\mathcal{M} = (M_i)_{1 \leq i \leq n}$ par rapport au sous-espace affine $\mathcal{E}$

$$I_{\mathcal{E}}(\mathcal{M}) = \sum_{i=1}^{n} p_i d_M^2(M_i, \mathcal{E}).$$

Remarque 1.20. — La définition de $d_M(B, \mathcal{E})$ ne dépend pas du point A de $\mathcal{E}$. En effet soit C un point de $\mathcal{E}$ distinct de A alors

$$(Id - P_E)(\overrightarrow{CB}) = (Id - P_E)(\overrightarrow{CA}) + (Id - P_E)(\overrightarrow{AB}),$$

or le vecteur $\overrightarrow{CA}$ est dans E donc sa projection sur E est lui-même, ainsi $(Id - P_E)(\overrightarrow{CA}) = 0$ donc

$$(Id - P_E)(\overrightarrow{CB}) = (Id - P_E)(\overrightarrow{AB}).$$

Ce qui montre que la définition ne dépend pas du choix du point A.

Remarque 1.21. — Si l'inertie est nulle $I_{\mathcal{E}}(\mathcal{M}) = 0$, cela signifie que le nuage $\mathcal{M}$ est inclus dans le sous-espace affine $\mathcal{E}$.

Proposition 1.22

Soit $\mathcal{E}_E$ un sous-espace affine de direction E et $\mathcal{E}_G$ le sous-espace affine de direction E passant par G, centre de gravité de $\mathcal{M}$, alors pour tout point A de $\mathcal{E}$, on a

$$I_{\mathcal{E}}(\mathcal{M}) = I_{\mathcal{E}_G}(\mathcal{M}) + \|(Id - p_E)(\overrightarrow{AG})\|_M^2.$$

Démonstration. —

$$\begin{aligned}
I_{\mathcal{E}}(\mathcal{M}) &= \sum_{i=1}^{n} p_i d_M^2(M_i, \mathcal{E}),\\
&= \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{AM_i})\|^2,\\
&= \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{AG} + \overrightarrow{GM_i})\|^2,\\
&= \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{AG})\|^2 + \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{GM_i}))\|^2\\
&\quad +2 < (Id - P_E)(\overrightarrow{AG}), \sum_{i=1}^{n} p_i (Id - P_E)(\overrightarrow{GM_i}) >_M,\\
&= \|(Id - P_E)(\overrightarrow{AG})\|^2 + I_{\mathcal{E}_G}(\mathcal{M}).
\end{aligned}$$

□

Ce résultat montre que parmi tous les sous-espaces affine parallèles à E, celui qui possède une inertie minimale est celui qui passe par le centre de gravité du nuage.

Par la suite, on recherche le ou les sous-espaces affines de dimension k donnée par rapport auquel(s) le nuage a une inertie minimale : c'est l'objectif de l'ACP.

On voit donc que ces sous-espaces optimaux passent nécessairement par G. C'est la raison pour laquelle on supposera, en général, par la suite que le tableau X est centré. Si ce n'est pas le cas, on raisonnera sur Y.

Proposition 1.23

On note $\mathcal{E}^{\perp}$ le sous espace affine passant par G et de direction $E^{\perp}$, on a

$$I_T = I_{\mathcal{E}}(\mathcal{M}) + I_{\mathcal{E}^{\perp}}(\mathcal{M}).$$

On pose

$$J_{\mathcal{E}}(\mathcal{M}) = I_{\mathcal{E}^{\perp}}(\mathcal{M}).$$

$J_{\mathcal{E}}(\mathcal{M})$ est l'inertie totale de la projection de $\mathcal{M}$ sur $\mathcal{E}$

Démonstration. — On a la relation $P_E + P_{E^{\perp}} = Id$, d'où en utilisant Pythagore

$$I_{\mathcal{E}}(\mathcal{M}) + I_{\mathcal{E}^{\perp}}(\mathcal{M}) = \sum_{i=1}^{n} p_i \|(Id - P_E)(\overrightarrow{GM_i})\|^2 + \sum_{i=1}^{n} p_i \|(Id - P_{E^{\perp}})(\overrightarrow{GM_i})\|^2 = I_T.$$

Pour le dernier point, il suffit d'appliquer la définition :

$$J_{\mathcal{E}}(\mathcal{M}) = \sum_{i=1}^{n} p_i \|(Id - P_E^{\perp})(\overrightarrow{GM_i})\|^2 = \sum_{i=1}^{n} p_i \|(P_E)(\overrightarrow{GM_i})\|^2.$$

□

Ainsi la recherche de $\mathcal{E}$ qui minimise $I_{\mathcal{E}}(\mathcal{M})$ est équivalent à rechercher $\mathcal{E}$ qui maximise $J_{\mathcal{E}}(\mathcal{M})$.

Remarque 1.24. — Si $J_{\mathcal{E}}(\mathcal{M}) = 0$, alors le nuage $\mathcal{M}$ est inclus dans $\mathcal{E}^{\perp}$.

Proposition 1.25 – Cas particulier d'une droite affine passant par G

Soit ϵ_1 un vecteur unitaire pour la métrique M de $\mathbb{R}^p$. Soit $\mathcal{E}_1$ la droite affine passant par G associée à $\text{Vect}(\epsilon_1)$. On a

$$J_{\mathcal{E}_1}(\mathcal{M}) = \epsilon_1' MVM\epsilon_1 =< \epsilon_1, VM\epsilon_1 >_M .$$

Démonstration. — Puisque pour tout vecteur u de $\mathbb{R}^p$, on a

$$P_{\text{Vect}(\epsilon_1)}(u) =< u, \epsilon_1 >_M \epsilon_1.$$

On en déduit que

$$\begin{aligned} J_{\mathcal{E}_1}(\mathcal{M}) &= \sum_{i=1}^n p_i \|(P_E)(\overrightarrow{GM_i})\|^2, \\ &= \sum_{i=1}^n p_i < y_i, \epsilon_1 >_M^2, \\ &= \sum_{i=1}^n p_i \epsilon_1' M y_i y_i' M \epsilon_1, \\ &= \epsilon_1' MVM\epsilon_1. \end{aligned}$$

□

Remarque 1.26. — Si $J_{\mathcal{E}_1}(\mathcal{M}) = 0$, alors le nuage $\mathcal{M}$ est inclus dans $\text{Vect}(\epsilon_1)^\perp$.

Proposition 1.27 – Décomposition de l'inertie

On considère $\mathcal{E}_G$ un sous-espace affine de $\mathbb{R}^p$ de direction E passant par G. Soit $(\epsilon_1, \cdots, \epsilon_k)$ une base orthonormale de E pour la métrique M, on complète cette base en une base orthonormale de $\mathbb{R}^p$ soit $(\epsilon_1, \cdots, \epsilon_k, \epsilon_{k+1}, \cdots, \epsilon_p)$. On a

$$I_{\mathcal{E}_G}(\mathcal{M}) = \sum_{l=k+1}^p J_{\mathcal{E}_l}(\mathcal{M}),$$

où $\mathcal{E}_l$ est la droite affine passant par G de direction $\text{Vect}(\epsilon_l)$.

Démonstration. — On a pour tout vecteur u

$$P_E(u) = \sum_{i=1}^k < u, \epsilon_i >_M \epsilon_i,$$

on en déduit que pour i fixé

$$\|(Id - P_E)(\overrightarrow{GM_i})\|^2 = \| \sum_{l=k+1}^p < \overrightarrow{GM_i}, \epsilon_l >_M \epsilon_l\|^2,$$

ce qui donne

$$\|(Id - P_E)(\overrightarrow{GM_i})\|^2 = \sum_{l=k+1}^p < \overrightarrow{GM_i}, \epsilon_l >_M^2,$$

et matriciellement

$$\|(Id - P_E)(\overrightarrow{GM_i})\|^2 = \sum_{l=k+1}^p \epsilon_l' M y_i y_i' M \epsilon_l.$$

Par conséquent on obtient par interversion de somme

$$\begin{aligned}
I_{\mathcal{E}_G}(\mathcal{M}) &= \sum_{i=1}^{n} p_i \sum_{l=k+1}^{p} \epsilon_l' M y_i y_i' M \epsilon_l, \\
&= \sum_{l=k+1}^{p} \epsilon_l' M \left(\sum_{i=1}^{n} p_i y_i y_i' \right) M \epsilon_l, \\
&= \sum_{l=k+1}^{p} \epsilon_l' M \, V \, M \epsilon_l, \\
&= \sum_{l=k+1}^{p} J_{\mathcal{E}_l}(\mathcal{M}).
\end{aligned}$$

□

Proposition 1.28 – Calcul de l'inertie totale

On a

$$I_T = tr(VM).$$

Démonstration. — On applique la proposition précédente en remarquant que l'inertie totale est l'inertie par rapport à l'espace $\mathbb{R}^p$ lui-même. Comme VM est une matrice associé à un endomorphisme symétrique, on choisit comme base orthonormale une base constituée de vecteurs propres de VM soit $(u_1, \cdots, u_p)$, on a

$$\forall j \in [\![1, p]\!], \;\; VMu_j = \lambda_j \, u_j.$$

$$\begin{aligned}
I_T &= I_{(\mathbb{R}^p)^\perp}(\mathcal{M}), \\
&= \sum_{j=1}^{p} u_j' MVM u_j, \\
&= \sum_{j=1}^{p} \lambda_j \|u_j\|_M^2, \\
&= \sum_{j=1}^{p} \lambda_j, \\
&= \mathrm{tr}(VM).
\end{aligned}$$

On peut aussi raisonner directement : puisque la trace de AB est égal à la trace de BA, on en déduit

$$\begin{aligned} I_T &= \sum_{i=1}^n p_i \|y_i\|_M^2, \\ &= \sum_{i=1}^n p_i y_i' M y_i, \\ &= \operatorname{tr}(\sum_{i=1}^n p_i y_i' M y_i), \\ &= \sum_{i=1}^n p_i \operatorname{tr}(y_i' M y_i), \\ &= \sum_{i=1}^n p_i \operatorname{tr}(M y_i y_i'), \\ &= \operatorname{tr}(M \sum_{i=1}^n p_j y_i y_i'), \\ &= \operatorname{tr}(MV) = \operatorname{tr}(VM). \end{aligned}$$

□

CHAPITRE 2

ANALYSE EN COMPOSANTES PRINCIPALES

Soit $\mathcal{N} = \{x_i, i \in I\} \subset \mathbb{R}^p$ un nuage de points de l'espace $\mathbb{R}^p$ muni de la métrique M. Chaque point x_i est muni de la masse $p_i > 0$ avec $\sum_{i \in I} p_i = 1$.

1. Recherche du meilleur sous-espace de dimension k représentant $\mathcal{N}$

L'objectif de l'ACP est de rechercher pour un entier k fixé le ou les sous-espaces affine de dimension k par rapport auquel(s) le nuage a une inertie minimale. D'après ce qui précède, on sait que le meilleur sous-espace $\mathcal{E}_k$ passe par G le centre de gravité de $\mathcal{N}$. On peut donc prendre l'origine en $O = G$ et il est équivalent de rechercher un sous-espace vectoriel E_k de dimension k tel que l'inertie $In(E_k)$ soit minimale. Comme

$$I_T = I_{E_k} + J_{E_k},$$

il est équivalent de rechercher E_k tel que J_{E_k} soit maximale.

Le théorème suivant décrit l'espace qui maximise J_{E_k} parmi tous les sous-espaces vectoriels de dimension k.

Théorème 2.1 – ACP

La matrice VM est une matrice M-symétrique, positive. On en déduit que VM est diagonalisable, que ses valeurs propres sont des réels et il existe une base M-orthonormale $(u_1, \cdots u_p)$ constituée de vecteurs propres de VM associés aux valeurs propres respectives

$$\lambda_1 \geq \cdots \geq \lambda_p \geq 0.$$

On pose

$$\forall k \in [\![1, p]\!], \quad E_k = \mathrm{Vect}(u_1, \cdots, u_k).$$

Alors on a

$$\forall k \in [\![1, p]\!], \quad J_{E_k} = \sum_{i=1}^{k} \lambda_i = \max_{E e.v. dim E = k} (J_E).$$

Réciproquement si F est un sous-espace vectoriel de dimension k tel que $J_F = \sum_{i=1}^{k} \lambda_i$, alors il existe une base orthonormale $(v_1, \cdots v_p)$ constituée de vecteurs propres de VM associé aux valeurs propres respectives $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$, telle que

$$F = \mathrm{Vect}(v_1, \cdots, v_k).$$

Démonstration. — Pour le premier point, on a pour tout vecteur u et v de $\mathbb{R}^p$:

$$\begin{aligned} < u, VMv >_M &= u'MVMv, \\ &= (VMu)'Mv, \\ &= < VMu, v >_M . \end{aligned}$$

donc la matrice VM est une matrice M-symétrique. De plus

$$< VMu, u >_M = u'MVMu.$$

or $u'MVMu$ est la variance d'une combinaison linéaire des variables initiales donc est positif. Ainsi la matrice VM est une matrice M-symétrique, positive.

Pour le second point, soit k un entier entre 1 et p, on a

$$J_{E_k} = \sum_{j=1}^{k} < u_j, VMu_j >_M = \sum_{j=1}^{k} \lambda_j.$$

Puis on considère un sous-espace vectoriel E de $\mathbb{R}^p$ de dimension k. Soit $(h_1, \cdots, h_k)$ une base orthonormale de E, on a

$$J_E = \sum_{i=1}^{k} < h_i, VMh_i >_M .$$

On décompose le vecteur h_i dans la base $(u_1, \cdots u_p)$, on a

$$\forall i \in [\![1, k]\!], \;\; h_i = \sum_{j=1}^{p} < h_i, u_j >_M \; u_j.$$

On en déduit que

$$\forall i \in [\![1, k]\!], \;\; VMh_i = \sum_{j=1}^{p} < h_i, u_j >_M \; \lambda_j \, u_j.$$

D'où

$$J_E = \sum_{i=1}^{k} \sum_{j=1}^{p} \lambda_j \; < h_i, u_j >_M^2 = \sum_{j=1}^{p} \lambda_j \, q_j$$

avec

$$q_j = \sum_{i=1}^{k} < h_i, u_j >_M^2 .$$

Or en notant P_E la projection orthogonale sur E, on a

$$0 \leq q_j = \|P_E(u_j)\|^2 \leq \|u_j\|^2 = 1,$$

ainsi que

$$\sum_{j=1}^{p} q_j = \sum_{i=1}^{k} \left(\sum_{j=1}^{p} < h_i, u_j >_M^2 \right) = \sum_{i=1}^{k} \|h_i\|_M^2 = k.$$

On en déduit donc

$$\begin{aligned} J_E - \sum_{j=1}^{k} \lambda_j &= \sum_{j=1}^{p} \lambda_j \, q_j - \sum_{j=1}^{k} \lambda_j, \\ &= \sum_{j=1}^{k} \lambda_j \, (q_j - 1) + \sum_{j=k+1}^{p} \lambda_j \, q_j, \\ &\leq \sum_{j=1}^{k} \lambda_k \, (q_j - 1) + \sum_{j=k+1}^{p} \lambda_k \, q_j, \\ &= \lambda_k \, k - k \, \lambda_k = 0. \end{aligned}$$

Par conséquent

$$J_E \leq \sum_{j=1}^{k} \lambda_j = J_{E_k}.$$

Réciproquement, on considère un sous-espace vectoriel E réalisant l'égalité, on note $(h_1, \cdots, h_k)$ une base orthonormale de E. De l'égalité

$$J_E = \sum_{j=1}^{k} \lambda_j$$

on déduit que

$$\sum_{j=1}^{k} \lambda_j\,(q_j - 1) + \sum_{j=k+1}^{p} \lambda_j\, q_j = \sum_{j=1}^{k} \lambda_k\,(q_j - 1) + \sum_{j=k+1}^{p} \lambda_k\, q_j,$$

soit

$$\sum_{j=1}^{k} (\lambda_j - \lambda_k)\,(q_j - 1) = \sum_{j=k+1}^{p} (\lambda_k - \lambda_j)\, q_j.$$

Le premier membre de l'égalité est négatif tandis que le second membre est positif puisque les valeurs propres sont dans l'ordre décroissant. Par conséquent

$$\sum_{j=1}^{k} (\lambda_j - \lambda_k)\,(q_j - 1) = 0 \text{ et } \sum_{j=k+1}^{p} (\lambda_k - \lambda_j)\, q_j = 0.$$

Ce que l'on peut écrire

$$\forall j \in [\![1, k]\!],\ \ (\lambda_j - \lambda_k)\,(q_j - 1) = 0 \text{ et } \forall j \in [\![k+1, p]\!],\ \ (\lambda_k - \lambda_j)\, q_j = 0.$$

On considère les deux indices j_0 et j_1 tels que

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{j_0} > \lambda_{j_0+1} = \cdots = \lambda_k = \cdots = \lambda_{j_1-1} > \lambda_{j_1} \geq \cdots \geq \lambda_p \geq 0.$$

On a donc

$$\forall j \in [\![1, j_0]\!],\ \ q_j = 1 \text{ et } \forall j \in [\![j_1, p]\!],\ \ q_j = 0.$$

Or q_j représente la norme au carré de la projection orthogonale de u_j sur E. Lorsque q_j vaut 1, la norme de u_j, cela signifie que le vecteur u_j est dans E. De même lorsque $q_j = 0$, alors u_j est dans l'orthogonale de E. On en conclut que

$$vect(u_1, \cdots, u_{j_0}) \subset E \text{ et } vect(u_{j_1}, \cdots, u_p) \subset E^{\perp},$$

soit

$$vect(u_1, \cdots, u_{j_0}) \subset E \subset vect(u_1, \cdots, u_{j_1-1}).$$

On peut donc construire une base orthonormale de E à partir de $(u_1, \cdots, u_{j_0})$ en rajoutant des vecteurs du sous-espace propre associé à λ_k. Ainsi E admet une base orthonormale constitué de vecteurs propres de VM selon le théorème. □

On peut introduire les définition suivantes :

Définition 2.2

Soit $(u_1, \cdots u_p)$ une base orthonormale de vecteurs propres de VM associé aux valeurs propres respectives

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0,$$

pour tout entier $1 \leq \alpha \leq p$,

- l'axe $\mathrm{Vect}(u_\alpha)$ est appelé le αième axe factoriel du nuage de points $\mathcal{N}$.
- $\varphi_\alpha = M u_\alpha$ est appelé le αième facteur,
- $\forall i \in [\![1, n]\!]$, $\psi_{\alpha,i} =< y_i, u_\alpha >_M = y_i' M u_\alpha = y_i' \varphi_\alpha$ est l'abscisse de la projection de y_i sur $\mathrm{Vect}(u_\alpha)$:

$$\psi_\alpha = \begin{pmatrix} \psi_{\alpha,1} \\ \vdots \\ \psi_{\alpha,n} \end{pmatrix} = Y\varphi_\alpha = YMu_\alpha \text{ est appelée } \alpha \text{ ième composante principale.}$$

- le taux d'inertie expliquée par le αième axe factoriel, noté τ_α, est la quantité

$$\tau_\alpha = \frac{\lambda_\alpha}{I_T} = \frac{\lambda_\alpha}{\sum_{i=1}^{p} \lambda_i}.$$

- le taux d'inertie expliquée par E_α, noté $\tau_{1\cdots\alpha}$, est la quantité

$$\tau_{1\cdots\alpha} = \frac{\lambda_1 + \cdots + \lambda_\alpha}{I_T} = \sum_{i=1}^{\alpha} \tau_i.$$

2. Axes de totale M-symétrie

Définition 2.3 – Axe de totale M-symétrie

Un axe Δ est un axe de totale M symétrie pour le nuage $\mathcal{N} = \{x_i,\ 1 \leq i \leq n\}$ s'il y a symétrie des points et des poids c'est-à-dire, en notant s_Δ la symétrie orthogonale par raport à l'axe Δ, pour tout entier i compris entre 1 et n, on a

- $s_\Delta(x_i) \in \mathcal{N}$,
- $s_\Delta(x_i)$ et x_i ont le même poids p_i.

Proposition 2.4

Tout axe de totale M-symétrie est un axe factoriel dans l'ACP du nuage $\mathcal{N}$.

Démonstration. — On note g le centre de gravité du nuage $\mathcal{N}$. On a

$$s_\Delta(g) = \sum_{i=1}^{n} p_i\, s_\Delta(x_i) = \sum_{i=1}^{n} p_i\, x_{i'} = \sum_{i=1}^{n} p_{i'}\, x_{i'} = g.$$

Donc g est sur l'axe Δ. On en déduit que l'axe Δ est toujours un axe de totale M-symétrie pour le nuage centré.

En notant Π_Δ la projection orthogonale sur Δ, on a la relation

$$s_\Delta = 2\Pi_\Delta \ - \ Id.$$

De plus soit u un vecteur unitaire de l'axe Δ, on a pour tout $1 \leq i \leq n$

$$< y_i, u >_M =< s_\Delta(y_i), u >_M \ \text{ et } y_i + s_\Delta(y_i) = 2 < y_i, u >_M \ u \in \Delta.$$

Avec ces propriétés, on prouve que u est un vecteur propre de VM, en effet

$$\begin{aligned}
VMu &= \left(\sum_{i=1}^{n} p_i\, y_i\, y_i'\right) Mu, \\
&= \sum_{i=1}^{n} p_i\, y_i\ < y_i, u >_M, \\
&= \sum_{i=1}^{n} p_i\ < y_i, u >_M\ y_i, \\
&= \sum_{i,\ y_i \in \Delta} p_i\ < y_i, u >_M\ y_i\ +\ \sum_{i,\ y_i \notin \Delta} p_i\ < y_i, u >_M\ y_i, \\
&= \sum_{i,\ y_i \in \Delta} p_i\ < y_i, u >_M\ y_i\ +\ \sum_{i,\ y_i \notin \Delta} p_i\ < y_i, u >_M\ (y_i + s_\Delta(y_i))/2, \\
&= \sum_{i=1}^{n} p_i\ < y_i, u >_M\ (y_i + s_\Delta(y_i))/2, \\
&= \left(\sum_{i=1}^{n} p_i\,(< y_i, u >_M)^2\right) u.
\end{aligned}$$

Par conséquent Δ est un axe factoriel associé à la valeur propre $\sum_{i=1}^{n} p_i\,(< y_i, u >_M)^2$. □

3. Représentations des individus

Un individu i du nuage $\mathcal{N}$ possède de nouvelles coordonnées dans la base $(u_1, \cdots u_p)$ qui est une base orthonormale de vecteurs propres de VM associé aux valeurs propres respectives

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0.$$

Les nouvelles coordonnées de l'individu i sont données par les composantes principales :

$$\begin{pmatrix} \psi_{1,i} \\ \vdots \\ \psi_{\alpha,i} \\ \vdots \\ \psi_{p,i} \end{pmatrix}.$$

Proposition 2.5 – Support du nuage des individus

Soit r le rang de Y. Alors seules les r premières valeurs propres sont non nulles

$$\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_p.$$

Ainsi le nuage $\mathcal{N}$ centré a pour support $E_r = \mathrm{Vect}(u_1, \cdots, u_r)$.

Démonstration. — Comme V et VM ont même rang puisque M est inversible, le rang de VM est celui de V qui est aussi celui de Y donc r, cela implique que VM possède exactement r valeurs propres non nulles donc

$$\lambda_1 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_p = 0.$$

Or une valeur propre λ_α est aussi $J_{\mathrm{Vect}(u_\alpha)}(\mathcal{N})$ et comme

$$\forall \alpha \in [\![r+1, p]\!],\quad J_{\mathrm{Vect}(u_\alpha)}(\mathcal{N}) = 0,$$

on en déduit que le nuage $\mathcal{N}$ centré est inclus dans $\text{Vect}(u_{r+1}, \cdots, u_p)^{\perp}$ soit $E_r = \text{Vect}(u_1, \cdots, u_r)$. □

Ainsi lorsque Y est de rang r, un individu i a $p-r$ coordonnées nulles donc est caractérisé par r valeurs $\psi_{i,1}, \cdots, \psi_{i,r}$ au lieu des p coordonnées initiales dans la base canonique. Les nouvelles coordonnées de l'individu i sont données par les composantes principales :

$$\begin{pmatrix} \psi_{1,i} \\ \vdots \\ \psi_{r,i} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Si le taux $\tau_{1,2}$ est proche de 1, on visualise le nuage $\mathcal{N}$ dans le plan $\text{Vect}(u_1, u_2)$, noté plan 1×2. Sinon on complète cette représentation par les projections sur les plans 1×3, 2×3, voire si $\tau_{1,2,3}$ est trop faible, sur les plans 1×4, 2×4, etc.

Définition 2.6 – Qualité de représentation

La qualité de la représentation de l'individu i sur E_k est

$$QLT(y_i, E_k) = \cos^2(\theta_{i,E_k}),$$

où θ_{i,E_k} est l'angle entre y_i et E_k.

Proposition 2.7

On a

$$QLT(y_i, E_k) = \sum_{\alpha=1}^{k} QLT(y_i, \text{Vect}(u_\alpha)) = \sum_{\alpha=1}^{k} (\frac{\psi_{\alpha,i}}{\|y_i\|_M})^2.$$

Démonstration. — On note P la projection orthogonale sur E_k, on a

$$< y_i, P(y_i) >_M = \|y_i\| \, \|P(y_i)\| \, \cos(\theta_{i,E_k}).$$

Or on a

$$P(y_i) = \sum_{\alpha=1}^{k} < y_i, u_\alpha >_M u_\alpha \text{ et } \|P(y_i)\|_M^2 = \sum_{\alpha=1}^{k} < y_i, u_\alpha >_M^2 .$$

donc

$$< y_i, P(y_i) > = \sum_{\alpha=1}^{k} < y_i, u_\alpha >^2 = \sum_{\alpha=1}^{k} \psi_{\alpha,i}^2 = \|P(y_i)\|_M^2.$$

On en conclut que

$$\cos^2(\theta_{i,E_k}) = \left(< \frac{y_i}{\|y_i\|}, \frac{P(y_i)}{\|P(y_i)\|} >_M \right)^2 = \sum_{\alpha=1}^{k} (\frac{\psi_{\alpha,i}}{\|y_i\|_M})^2.$$

□

Plus ce facteur de qualité se rapproche de 1, mieux est représenté l'individu i. S'il vaut 1, alors y_i est dans E_k.

On note parfois sur les listings, $COR_\alpha(i)$ pour désigner $1000 \times \cos^2(\theta_{i,\text{Vect}(u_\alpha)})$ et aussi $QLT_{E_k}(i) = 1000 \times \cos^2(\theta_{i,E_k})$.

4. Représentation des variables

Les variables y^j sont représentés par les vecteurs de l'espace $\mathbb{R}^n$ muni de la métrique D_p. Pour cette métrique, la norme d'un vecteur est l'écart-type de la variable et le produit scalaire entre deux vecteurs est la covariance entre les deux variables. La composante principale ψ_α est un vecteur de $\mathbb{R}^n$.

Proposition 2.8

La composante principale ψ_α associée à la valeur propre λ_α est centrée, de variance égale à la valeur propre λ_α et les composantes principales ne sont pas corrélées entre elles :

$$\forall (\alpha,\beta) \in [\![1,p]\!]^2, \quad \mathrm{Cov}(\psi_\alpha,\psi_\beta) = <\psi_\alpha,\psi_\beta>_{D_p} = \begin{cases} \lambda_\alpha & \textit{si } \alpha=\beta, \\ 0 & \textit{si } \alpha \neq \beta \end{cases}$$

Démonstration. — Toute composante principale est une combinaison linéaire des variables y^j qui sont toutes centrées donc ψ_α est centrée. De plus on a

$$\begin{aligned} <\psi_\alpha,\psi_\beta>_{D_p} &= \psi'_\alpha D_p \psi_\beta, \\ &= u'_\alpha M Y' D_p Y M u_\beta, \\ &= u'_\alpha M V M u_\beta, \\ &= \lambda_\beta <u_\alpha, u_\beta>_M . \end{aligned}$$

D'où le résultat. □

Proposition 2.9

Soit r le rang de Y. Alors seules les r premières valeurs propres sont non nulles

$$\lambda_1 \geq \cdots \geq \lambda_r > 0 = \lambda_{r+1} = \cdots = \lambda_p.$$

On pose

$$\forall \alpha \in [\![1,r]\!], \quad v_\alpha = \frac{\psi_\alpha}{\sqrt{\lambda_\alpha}}.$$

La famille $(v_1,\cdots,v_r)$ est une base orthonormale de $\mathrm{Vect}(y^1,\cdots,y^p) = \mathrm{Im}\, Y$, la αème coordonnée de y^j est donnée par

$$\forall 1 \leq \alpha \leq r, \quad \eta_{\alpha,j} = <y^j, \frac{\psi_\alpha}{\sqrt{\lambda_\alpha}}>_{D_p} .$$

On a

$$\eta_\alpha = \begin{pmatrix} \eta_{\alpha,1} \\ \vdots \\ \eta_{\alpha,p} \end{pmatrix} = \sqrt{\lambda_\alpha} u_\alpha \ \textit{et} \ \|\eta_\alpha\|_M^2 = \lambda_\alpha.$$

Démonstration. — Le premier point a déjà été démontré. Puis pour tout $1 \leq k \leq r$, le vecteur v_k est une combinaison linéaire des variables y^j, $1 \leq j \leq p$, on en déduit que $\mathrm{Vect}(v_1,\cdots,v_r)$ est inclus dans $\mathrm{Im}\, Y$. Par ailleurs ces deux espaces ont même dimension r donc ils sont égaux. On en conclut que $(v_1,\cdots,v_r)$ est une base orthonormale de $\mathrm{Vect}(y^1,\cdots,y^p) = \mathrm{Im}\, Y$. Puis on a

$$\eta_{\alpha,j} = \frac{1}{\sqrt{\lambda_\alpha}} {y^j}' D_p Y M u_\alpha$$

donc

$$\eta_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} Y' D_p Y M u_\alpha = \sqrt{\lambda_\alpha} u_\alpha.$$

$\square$

On représente donc la variable y^j dans ce nouveau repère et les nouvelles coordonnées de y^j sont

$$\begin{pmatrix} \eta_{1,j} \\ \vdots \\ \eta_{\alpha,j} \\ \vdots \\ \eta_{r,j} \end{pmatrix}.$$

Définition 2.10 – Qualité de représentation

La qualité de la représentation de la variable y^j sur $F_k = \text{Vect}(v_1, \cdots, v_k)$, avec $1 \leq k \leq r$ est

$$QLT(y^j, F_k) = \cos^2(\theta_{j,F_k}),$$

où θ_{j,F_k} est l'angle entre y^j et F_k.

Proposition 2.11

On a

$$QLT(y^j, F_k) = \sum_{\alpha=1}^{k} QLT(y^j, \text{Vect}(v_\alpha)) = \sum_{\alpha=1}^{k} < \frac{y^j}{\|y^j\|}, v_\alpha >_{D_p}^2 = \sum_{\alpha=1}^{k} r_{i,\alpha}^2.$$

où $r_{j,\alpha}$ désigne la corrélation entre y^j et v_α puisque ces deux variables sont centrées.

Démonstration. — On note P la projection orthogonale sur F_k, on a

$$< y^j, P(y^j) >_{D_p} = \|y^j\| \, \|P(y^j)\| \, \cos(\theta_{j,F_k}).$$

Or on a

$$P(y^j) = \sum_{\alpha=1}^{k} < y^j, v_\alpha >_{D_p} v_\alpha \text{ et } \|P(y^j)\|_{D_p}^2 = \sum_{\alpha=1}^{k} < y^j, v_\alpha >_{D_p}^2 .$$

donc

$$< y^j, P(y^j) >= \sum_{\alpha=1}^{k} < y^j, v_\alpha >^2 = \|P(y^j)\|_{D_p}^2.$$

On en conclut que

$$\cos^2(\theta_{i,F_k}) = \left(< \frac{y^j}{\|y^j\|}, \frac{P(y^j)}{\|P(y^j)\|} >_{D_p} \right)^2 = \sum_{\alpha=1}^{k} < \frac{y^j}{\|y^j\|}, v_\alpha >_{D_p}^2 = \sum_{\alpha=1}^{k} r_{i,\alpha}^2.$$

$\square$

5. Décompositions de l'inertie

5.1. Décomposition de l'inertie selon les individus. — Puisque l'inertie totale I_T est égale à la somme des valeurs propres et comme chaque valeur propre λ_α est le carré de la norme de la composante principale associée ψ_α pour la métrique D_p, on a

$$I_T = \sum_{\alpha=1}^{r} \lambda_\alpha = \sum_{\alpha=1}^{r} \|\psi_\alpha\|_{D_p}^2 = \sum_{\alpha=1}^{r} \sum_{i=1}^{n} p_i(\psi_{\alpha,i})^2 = \sum_{i=1}^{n} p_i ||y_i||_M^2, \quad \text{et } \lambda_\alpha = \sum_{i=1}^{n} p_i(\psi_{\alpha,i})^2.$$

On en déduit la définition suivante :

Définition 2.12 – Contribution relative

La contribution relative de l'individu y_i à l'inertie totale est

$$INR(i) = \frac{p_i||y_i||_M^2}{I_T}.$$

La contribution relative de l'individu y_i à l'inertie de l'axe α est

$$CTR_\alpha(i) = \frac{p_i(< y_i, u_\alpha >_M)^2}{\lambda_\alpha} = \frac{p_i(\psi_{\alpha,i})^2}{\lambda_\alpha},$$

Remarque 2.13. — De même puisque $\|y_i\|_M^2 = \sum_{\alpha=1}^{p} \psi_{\alpha,i}^2$, on peut définir la contribution relative de l'axe α à l'inertie de l'individu y_i de la manière suivante

$$CTR_i(\alpha) = \frac{(\psi_{\alpha,i})^2}{\sum_{\alpha=1}^{p}(\psi_{\alpha,i})^2} = \cos^2(\theta_{\alpha,i}) = COR_\alpha(i),$$

où $\theta_{i,\alpha}$ est l'angle entre y_i et u_α.

Sur les listings, $CTR_\alpha(i)$ et $COR_\alpha(i)$ sont souvent multipliés par 1000.

5.2. Décomposition de l'inertie selon les variables. — On suppose que la matrice M est diagonale :

$$M = diag(m_1, \cdots, m_p) \text{ où les réels } m_j,\ 1 \leq j \leq p, \text{ sont strictement positifs.}$$

Puisque l'inertie totale I_T est dans ce cas égale à la somme des variances de chaque variable pondéré par m_j

$$I_T = \sum_{j=1}^{p} m_j \|y^j\|_{D_p}^2,$$

et puisque chaque valeur propre λ_α est le carré de la norme de η_α pour la métrique M, on a

$$\lambda_\alpha = \|\eta_\alpha\|_M^2 = \sum_{j=1}^{p} m_j(\eta_{\alpha,j})^2,$$

on en déduit la définition suivante :

Définition 2.14 – Contribution relative

La contribution relative de la variable y^j à l'inertie totale est

$$INR(j) = \frac{m_j\|y^j\|_{D_p}^2}{I_T}.$$

La contribution relative de la variable y^j à l'inertie de l'axe α est

$$CTR_\alpha(j) = \frac{m_j(< y^j, v_\alpha >_{D_p})^2}{\lambda_\alpha} = m_j u_{\alpha,j}^2,$$

Remarque 2.15. — De même on peut définir la contribution relative de l'axe α à l'inertie de la variable y^j par

$$COR_\alpha(j) = r_{i,\alpha}^2 = \cos^2(\theta_{j,\alpha}),$$

où $\theta_{j,\alpha}$ est l'angle entre y^j et v_α.

Sur les listings, $CTR_\alpha(j)$ et $COR_\alpha(j)$ sont souvent multipliés par 1000.

Remarque 2.16. — Il est possible de retrouver cette formule en remarquant que pour tout $1 \leq \alpha \leq p$, le vecteur u_α est unitaire pour la métrique M donc $\sum_{j=1}^{p} m_j u_{\alpha,j}^2 = 1$, soit en multipliant par λ_α

$$\lambda_\alpha = \sum_{j=1}^{p} \lambda_\alpha m_j u_{\alpha,j}^2.$$

On retrouve les formules de contributions relatives pour une variable y^j.

5.3. Eléments supplémentaires. — On peut prendre comme éléments supplémentaires une observation douteuse, un élément aberrant, un cas nouveau, le centre de gravité d'un groupe ("homme", "femme"), des éléments de nature différente (opinion/CSP).

Un individu supplémentaire est un individu y_s de $\mathbb{R}^p$ n'ayant pas participé à l'analyse. L'abscisse $\psi_{\alpha,s}$ de sa projection sur $\mathrm{Vect}(u_\alpha)$ vérifie

$$\psi_{\alpha,s} = y_{s'} M u^\alpha.$$

Il est clair que $\psi_{\alpha,s}$ s'obtient en effectuant l'analyse factorielle du tableau $X_1 = \begin{pmatrix} X \\ x'_s \end{pmatrix}$ et en donnant un poids nul à s. En effet dans ce cas, les seuls points ayant une inertie non nulle sont les x_i pour $1 \leq i \leq n$.

De même, une variable supplémentaire est une variable x^s de $\mathbb{R}^n$ n'ayant pas participé à l'analyse. Elle peut être représentée par ses projections sur les nouveaux axes v_α, on note y^s la variable centrée

$$\eta_{\alpha,s} =< y^s, \frac{\psi_\alpha}{\sqrt{\lambda_\alpha}} >_M .$$

Exercice de manipulation 2.1. — : Montrer que l'on peut exprimer $\psi_{\alpha,s}$ en fonction de $\psi_{\alpha,i}$ selon la formule :

$$\psi_{\alpha,s} = \frac{1}{\lambda_\alpha} \sum_{i=1}^{n} w_{s,i} p_i \psi_{\alpha,i} \text{ avec } w_{s,i} =< y_s, y_i >_M .$$

solution :

$$\begin{aligned}
\sum_{i=1}^{n} w_{s,i} p_i \psi_{\alpha,i} &= \sum_{i=1}^{n} p_i < y_s, y_i >_M \psi_{\alpha,i}, \\
&= < y_s, \sum_{i=1}^{n} p_i \psi_{\alpha,i} y_i >_M, \\
&= < y_s, \sum_{i=1}^{n} p_i < y_i, u_\alpha > y_i >_M, \\
&= < y_s, \sum_{i=1}^{n} p_i y_i < y_i, u_\alpha >>_M, \\
&= < y_s, \sum_{i=1}^{n} p_i y_i y'_i, M u_\alpha >_M, \\
&= < y_s, VMu_\alpha >_M, \\
&= \lambda_\alpha \psi_{\alpha,s}.
\end{aligned}$$

6. Analyse en composantes principales

Etant donnée un nuage de n points, muni chacun d'un poids, dans $\mathbb{R}^p$ muni de la métrique M, effectuer une Analyse en Composantes Principales, ACP, du tableau X associé avec les métriques M et D_p pour les espaces $\mathbb{R}^p$ et $\mathbb{R}^n$, consiste à rechercher les composantes principales associées aux axes factoriels. Pour cela, on diagonalise la matrice VM, ce qui fournit les valeurs propres et les axes factoriels.

Définition 2.17 – ACP sur matrice variance

Effectuer une ACP sur matrice variance du tableau X consiste à prendre comme métrique

$$M = I_p \text{ et } D_p = diag(p_1, \cdots, p_n).$$

Dans ce cas, la matrice VM est la matrice de variance-covariance.

On effectue souvent la représentation des variables dans le cercle de corrélations c'est-à-dire au lieu de représenter les variables selon leurs covariances avec les axes v_α, on les représente par leurs corrélations avec les axes v_α. Cette opération revient à représenter non pas le vecteur y^j mais $\frac{y^j}{\|y^j\|}$. Dans ce cas, toutes les variables sont des vecteurs unitaires et les extrémités des variables sont sur la sphère unité de $\mathbb{R}^n$.

Lorsque l'on projette une variable sur un plan factoriel, la sphère unité est projeté sur le cercle unité, appelé cercle de corrélations, et les projections des variables sont des vecteurs de norme inférieure à 1 donc dans le cercle unité. Mais si une variable est sur le cercle de corrélations, alors le vecteur $\frac{y^j}{\|y^j\|}$ est égal à sa projection ce sui signifie que la variable est parfaitement représentée, donc expliquée, par les deux facteurs associés.

6.1. ACP sur matrice de corrélation ou ACP normée. —

Définition 2.18 – ACP normée

Effectuer une ACP normée du tableau X consiste à prendre comme métrique

$$M = \Delta^2 = diag(\frac{1}{v_{11}}, \cdots, \frac{1}{v_{pp}}) \text{ et } D_p = diag(p_1, \cdots, p_n),$$

avec $\Delta = diag(\frac{1}{\sqrt{v_{11}}}, \cdots, \frac{1}{\sqrt{v_{pp}}})$ où v_{jj} est la variance de y^j.

Proposition 2.19

Etant donné un tableau X, on centre et on divise chaque variable par son écart-type, on obtient un nouveau tableau Z dont les variables sont toutes centrées et réduites. On a

$$Z = Y\Delta \text{ où } \Delta = diag(\frac{1}{\sqrt{v_{11}}}, \cdots, \frac{1}{\sqrt{v_{pp}}}).$$

On réalise une ACP normée sur X en effectuant une ACP sur Z avec $M = I_p$.

Démonstration. — Il s'agit de prouver que l'on retrouve les même composantes principales dans les deux ACP. Dans le cas de l'ACP normée, on diagonalise $VM = Y'D_pYM$. On note u_α un axe factoriel associé à la valeur propre λ_α :

$$VMu_\alpha = \lambda_\alpha u_\alpha.$$

Par ailleurs, l'ACP sur la matrice Z revient à diagonaliser $Z'D_pZI_p = Z'D_pZ$ qui est la matrice de corrélations. On exprime cette matrice en fonction de V :

$$Z'D_pZ = \Delta Y'D_pY\Delta = \Delta\, V\, \Delta.$$

Or on remarque que

$$Z'D_pZ(\Delta u_\alpha) = \Delta V\Delta^2 u_\alpha = \Delta VMu_\alpha = \lambda_\alpha \Delta u_\alpha.$$

De plus le vecteur Δu_α est non nul puisque de norme 1 :

$$\|\Delta u_\alpha\|^2_{I_p} = (\Delta u_\alpha)'\Delta u_\alpha = u'_\alpha M u_\alpha = 1,$$

ainsi Δu_α est un axe factoriel dans l'analyse de Z associé à la valeur propre λ_α. De plus pour $\alpha \neq \beta$, on a

$$< \Delta u_\alpha, \Delta u_\beta >_{I_p} = (\Delta u_\alpha)' \Delta u_\beta = u'_\alpha M u_\beta = 0.$$

On note ψ_α la composante principale associée à λ_α dans l'ACP normée, on a

$$\psi_\alpha = Y M u_\alpha = Y \Delta \, \Delta u_\alpha = Z \Delta u_\alpha,$$

or $Z \Delta u_\alpha$ est la composante principale associée à λ_α dans l'ACP sur Z avec MI_p.

On en conclut que les deux ACP sont équivalentes. □

Proposition 2.20 – Inertie totale d'une ACP

Dans le cas d'une ACP normée, l'inertie totale du nuage est égal à p, le nombre de variables.

Démonstration. — Dans ce type d'ACP, on diagonalise la matrice des corrélations. Or cette matrice ne comporte que des 1 sur la diagonale (car la corrélation d'une variable avec elle-même est 1). Dès lors l'inertie totale est la trace de cette matrice et vaut p. □

7. Analyse factorielle d'un système de points munis de poids et de distances

On considère un système de points M_i, $i \in I$, munis de poids p_i dans un espace affine. On identifie les points M_i à leurs vecteurs de coordonnées x_i dans un espace euclidien muni de la métrique M.

Théorème 2.21 – Tableau de distances

Dans un espace affine $\mathbb{R}^p$ de métrique M, on considère le nuage $\mathcal{N}_X$ constitué de n points M_i, $1 \leq i \leq n$. Le tableau de distance noté D entre les points du nuage est une matrice carrée d'ordre n de terme courant $d(i, i')$ avec

$$\forall (i, i') \in [\![1, n]\!]^2, \;\; d(i, i') = d^2(M_i, M_{i'}) = \|x_i - x_{i'}\|_M^2 = \|y_i - y_{i'}\|_M^2.$$

La représentation du nuage $\mathcal{N}_X$ des points M_i affectés des poids p_i dans le système des axes factoriels ne dépend que des poids p_i et de la matrice D.

Démonstration. — Soit ψ_α la composante principale associée au αième axe factoriel u_α, on a

$$Y M Y' D_p \psi_\alpha = Y M Y' D_p Y M u_\alpha = Y M V M u_\alpha = \lambda_\alpha Y M u_\alpha = \lambda_\alpha \psi_\alpha,$$

où Y est le tableau centré associé à X. Ceci montre que la diagonalisation de la matrice $YMY'D_p$ fournit les valeurs propres de l'ACP et pour chaque valeur propre non nulle, tout vecteur propre de norme égale à $\sqrt{\lambda_\alpha}$ est une composante principale.

Pour démontrer le résultat, il s'agit donc de prouver que l'on peut construire la matrice YMY' à partir des poids p_i et des distances $d^2(M_i, M_{i'})$. On pourra alors en déduire $YMY'D_p$.

La matrice YMY' est une matrice carrée d'ordre n et le terme courant de cette matrice est

$$\forall (i, i') \in [\![1, n]\!]^2, \;\; (YMY')_{ii'} = < y_i, y_{i'} >_M .$$

Une telle matrice s'appelle matrice de Gram associée à la famille de vecteurs $(y_i)_{1 \leq i \leq n}$. On pose

$$\forall i \in [\![1, n]\!], \;\; d(\cdot, i) = \sum_{i'=1}^{n} p_{i'} d(i, i'), \;\; \text{et } d(\cdot, \cdot) = \sum_{i=1}^{n} p_i d(\cdot, i).$$

On exprime le produit scalaire en fonction de la norme soit

$$\forall (i, i') \in [\![1, n]\!]^2, \;\; < y_i, y_{i'} >_M = \frac{-1}{2} (\|y_i - y_{i'}\|_M^2 - \|y_i\|_M^2 - \|y_{i'}\|_M^2).$$

On somme de $i' = 1$ à n en pondérant par $p_{i'}$ d'où

$$\forall i \in [\![1, n]\!], \quad \sum_{i'=1}^{n} p_{i'} < y_i, y_{i'} >_M = \frac{-1}{2}(\sum_{i'=1}^{n} p_{i'} \|y_i - y_{i'}\|_M^2 - \sum_{i'=1}^{n} p_{i'} \|y_i\|_M^2 - \sum_{i'=1}^{n} p_{i'} \|y_{i'}\|_M^2).$$

On note I_T l'inertie totale : $I_T = \sum_{i=1}^{n} p_i \|y_i\|_M^2$. Par ailleurs, $\sum_{i'=1}^{n} p_{i'} y_{i'} = 0$ donc

$$0 = \frac{-1}{2}(d(\cdot, i) - \|y_i\|_M^2 - I_T) \Longrightarrow \|y_i\|_M^2 = d(\cdot, i) - I_T,$$

Puis on somme de $i = 1$ à n en pondérant par p_i d'où

$$I_T = d(\cdot, \cdot) - I_T.$$

De cette dernière relation, on déduit que

$$I_T = \frac{1}{2} d(\cdot, \cdot).$$

Par conséquent

$$\forall (i, i') \in [\![1, n]\!]^2, \quad < y_i, y_{i'} >_M = \frac{-1}{2}(d(i, i') - d(\cdot, i) + I_T - d(\cdot, i') + I_T),$$

donc

$$\forall (i, i') \in [\![1, n]\!]^2, \quad < y_i, y_{i'} >_M = \frac{-1}{2}(d(i, i') - d(\cdot, i) - d(\cdot, i') + d(\cdot, \cdot)).$$

□

8. Reconstruction du nuage

Cette proposition donne une décomposition du tableau centré Y et montre comment à partir de la connaissance des composantes principales et des axes factoriels u_α on peut reconstruire le tableau centré Y.

Proposition 2.22

On effectue une ACP sur le tableau X avec comme métrique M pour $\mathbb{R}^p$ et D_p pour $\mathbb{R}^n$. On note r le rang de Y, on a la relation suivante avec les notations habituelles

$$Y = \sum_{\alpha=1}^{r} \sqrt{\lambda_\alpha}\, v_\alpha\, u'_\alpha = \sum_{\alpha=1}^{r} \psi_\alpha\, u'_\alpha.$$

Démonstration. — On pose

$$T = \sum_{\alpha=1}^{r} \psi_\alpha\, u'_\alpha.$$

Soit $1 \leq \beta \leq p$, on a

$$TMu_\beta = \sum_{\alpha=1}^{r} \psi_\alpha\, u'_\alpha M u_\beta = \psi_\beta = YMu_\beta.$$

Puisque $(u_\beta)_{1 \leq \beta \leq p}$ est une base de $\mathbb{R}^p$, on en déduit que $TM = YM$ et comme M est inversible on en conclut que $T = Y$.

□

Remarque 2.23. — Dans cette reconstruction, on pourrait négliger les termes de la somme associées aux plus faibles valeurs propres, et dans ce cas on obtiendrait un tableau $\hat{Y}$ qui serait une approximation de Y.

Il y a une réciproque à la proposition précédente c'est-à-dire :

Proposition 2.24

Soit Y un tableau centrée, s'il existe une famille $(u_\alpha)_{1\leq\alpha\leq r}$ orthonormale de $\mathbb{R}^p$ muni de la métrique M et une famille $(v_\alpha)_{1\leq\alpha\leq r}$ orthonormale de $\mathbb{R}^n$ muni de la métrique D_p et r réels strictements positifs $(\lambda_\alpha)_{1\leq\alpha\leq r}$ vérifiant

$$Y = \sum_{\alpha=1}^{r} \sqrt{\lambda_\alpha}\, v_\alpha\, u'_\alpha.$$

Alors le rang de Y est r, les vecteurs$(u_\alpha)_{1\leq\alpha\leq r}$ sont les axes factoriels non triviaux associés aux valeurs propres $(\lambda_\alpha)_{1\leq\alpha\leq r}$ et les composantes principales sont $\psi_\alpha = \sqrt{\lambda_\alpha} v_\alpha$, $1 \leq \alpha \leq r$.

Démonstration. — Avec la décomposition on a

$$\begin{aligned}
VMu_\alpha &= Y'D_pYMu_\alpha,\\
&= \left(\sum_{\beta=1}^{r}\sqrt{\lambda_\beta}\, v_\beta\, u'_\beta\right)' D_p \left(\sum_{\gamma=1}^{r}\sqrt{\lambda_\gamma}\, v_\gamma\, u'_\gamma\right) Mu_\alpha,\\
&= \left(\sum_{\beta=1}^{r}\sqrt{\lambda_\beta}\, u_\beta\, v'_\beta\right) D_p \sum_{\gamma=1}^{r}\sqrt{\lambda_\gamma}\, v_\gamma\, u'_\gamma\, Mu_\alpha,\\
&= \sum_{\beta=1}^{r}\sum_{\gamma=1}^{r}\sqrt{\lambda_\beta\,\lambda_\gamma}\, u_\beta \;<v_\beta, v_\gamma>_{D_p}\, u'_\gamma\, Mu_\alpha,\\
&= \sum_{\beta=1}^{r}\lambda_\beta\, u_\beta \;<u_\beta, u_\alpha>_M,\\
&= \lambda_\alpha u_\alpha.
\end{aligned}$$

On en conclut que les vecteurs u_α, $1 \leq \alpha \leq r$, sont des axes factoriels de l'ACP de Y associées aux valeurs propres λ_α. De plus

$$\psi_\alpha = YMu_\alpha = \sqrt{\lambda_\alpha} v_\alpha,$$

on retrouve donc les composantes principales de l'ACP de Y. Enfin le rang de Y est r, puisque pour tout vecteur x de $\mathbb{R}^p$, Yx est une combinaison linéaire des vecteurs v_α, $1 \leq \alpha \leq r$. Comme chaque vecteur v_α est dans l'image de Y en tant que vecteur propre associé à une valeur propre non nulle, on en conclut que l'image de Y est l'espace généré par la famille $(v_\alpha)_{1\leq\alpha\leq r}$ donc la dimension de l'image de Y est r.

□

Ce que montre cette proposition est que si l'on arrive à décomposer le tableau Y sous la forme donnée dans la proposition précédente les résultats de l'ACP se lisent directement. Or cette décompostion est connue sous le nom de décomposition en valeurs singulières notée SVD en utilisant les métriques canoniques c'est-à-dire $M = I_p$ et $D_p = I_n$, nous rappelons ce résultat

Théorème 2.25 – Décomposition en Valeurs Singulières, SVD

Soit Y une matrice de format $n \times p$ à coefficients réels. On note r le rang de Y, $r \leq s = min(n, p)$. Les espaces $\mathbb{R}^n$ et $\mathbb{R}^p$ sont munis d'une structure euclidienne canonique. Alors il existe

1. *$(u_1, \cdots, u_n)$ une base orthonormale de $\mathbb{R}^n$,*
2. *$(v_1, \cdots, v_p)$ une base orthonormale de $\mathbb{R}^p$,*
3. *r réels positifs : $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$,*

tel que

$$Y = \sum_{i=1}^{r} \sigma_i u_i v_i'.$$

Les r réels σ_i sont uniques. On les appelle valeurs singulières de Y.
Matriciellement, on pose

$$U = [u_1, \cdots, u_r] \text{ et } V = [v_1, \cdots, v_r], \quad \Sigma = Diag(\sigma_1, \cdots, \sigma_r),$$

L'équation précédente s'écrit

$$Y = U \Sigma V',$$

ou encore on peut poser

$$U = [u_1, \cdots, u_n] \text{ et } V = [v_1, \cdots, v_n],$$

et Σ est une matrice $n \times p$ dont les coefficients diagonaux sont des réels positifs ou nuls et tous les autres sont nuls. Les termes diagonaux de Σ sont rangés par ordre décroissant . Les matrices U et V sont deux matrices orthogonales d'ordre respectif n et p ($U'U = UU' = I_n$ et $VV' = V'V = I_p$) Dans les deux cas, la matrice Σ est unique.

Exemple 2.26. — On considère le tableau de données suivant associé aux résultats de trois variables x, y et z mesurées sur un échantillon I de six individus.

$I \setminus J$	x	y	z
1	1	6	5
2	2	5	3
3	3	1	-2
4	4	3	-1
5	2	2	0
6	6	1	-5

On suppose que chaque individu i de I $(1 \leq i \leq 6)$ est muni de la masse 1/6 donc $D_p = \frac{1}{6} I_6$ et $M = I_3$. On note X le tableau associé. Le tableau centré est

$$Y = \begin{pmatrix} -2 & 3 & 5 \\ -1 & 2 & 3 \\ 0 & -2 & -2 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \\ 3 & -2 & -5 \end{pmatrix}$$

Puis on applique la décomposition en valeurs singulières de la matrice Y à l'aide de la commande *svd* de R, on obtient

$$Y = U\Sigma V'$$

avec

$$U = \begin{pmatrix} -0.627137895 & 0.1483193 & 0.676592730 \\ -0.377938565 & 0.2072554 & -0.695626901 \\ 0.257478469 & -0.5323830 & 0.124438341 \\ 0.120460096 & 0.3251276 & 0.065324265 \\ 0.008279139 & -0.5913191 & -0.003105094 \\ 0.618858757 & 0.4429998 & 0.196376400 \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} 9.813206 & 0 & 0 \\ 0 & 2.387675 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$V = \begin{pmatrix} 0.3669516 & 0.72939234 & 0.5773503 \\ -0.4481965 & 0.68248558 & -0.5773503 \\ -0.8151481 & -0.04690676 & 0.5773503 \end{pmatrix}'$$

On en déduit que les résultats d'une ACP sur Y avec comme métrique $D_p = 1/6 I_6$ et $M = I_3$ sont

- Pour les valeurs propres, ce sont les carrés des termes de la matrice diagonale divisés par 6

$$\lambda_1 = 96/6 = 16 > \lambda_2 = 5.7/6 = 0.95 > \lambda_3 = 0.$$

- Pour les axes factoriels ce sont les colonnes de V.
- Pour les vecteurs v_α, il faut rendre la matrice $U\ D_p$ orthogonale alors qu'elle est orthogonale pour la métrique I_3, il suffit donc de multiplier par $\sqrt{6}$ car $\sqrt{6}VD_p\sqrt{6}V' = VV' = I_3$.

Ces résultats ont été obtenus en utilisant la commande svd de R

Programme en R.

```
> X=matrix(c(1,2,3,4,5,6,6,5,4,3,2,1,0,1,2,2,1,0),6,3)
> moy=apply(X,MARGIN=2,mean)
> Y=X-t(moy\%*\%matrix(c(1,1,1,1,1,1),1,6))
> s=svd(Y)
```

CHAPITRE 3

ANALYSE FACTORIELLE DES CORRESPONDANCES

1. Introduction

L'analyse Factorielle des Correspondances (AFC) a été introduite pour analyser les tableaux de contingence. Un tableau de contingence croise les ensembles I et J de deux variables qualitatives X et Y. Un tel tableau peut se noter k_{IJ} et a alors pour terme général le nombre $k(i,j)$ d'individus qui ont pris simultanément la modalité i pour la variable X et la modalité j pour la variable Y.

L'AFC consiste à effectuer deux ACP, l'une sur le tableau des profils lignes, l'autre sur celui des profils colonnes de k_{IJ}.

L'AFC peut être appliquée à des tableaux de nombres positifs de types divers : tableaux de contingence, tableaux homogènes de nombres positifs, tableaux d'échanges, tableau de rangs, tableaux de présence /absence, tableau disjonctifs complets,....

2. Définition des nuages étudiés par l'AFC

2.1. Notations. — On étudie deux variables qualitatives X et Y, X a p modalités et Y q modalités. Le tableau de contingence k_{IJ} est une matrice de format $p \times q$. On pose $I = \{1, \cdots p\} = [\![1, p]\!]$ et $J = \{1, \cdots q\} = [\![1, q]\!]$.

On note

$$k_I = (k(i, \cdot))_{i \in I} \in \mathbb{R}^p \text{ avec } k(i, \cdot) = \sum_{j=1}^{q} k(i,j),$$

$$k_J = (k(\cdot, j))_{j \in J} \in \mathbb{R}^q \text{ avec } k(\cdot, j) = \sum_{i=1}^{p} k(i,j),$$

$$k = \sum_{j=1}^{q} \sum_{i=1}^{p} k(i,j).$$

Ces définitions dans le tableau ci-dessous :

	1	$\cdots$	j	$\cdots$	q	total
1	$k(1,1)$	$\cdots$	$k(1,j)$	$\cdots$	$k(1,q)$	$k(1,\cdot)$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
i	$k(i,1)$	$\cdots$	$k(i,j)$	$\cdots$	$k(i,q)$	$k(i,\cdot)$
$\vdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
p	$k(p,1)$	$\cdots$	$k(p,j)$	$\cdots$	$k(p,q)$	$k(p,\cdot)$
total	$k(\cdot,1)$	$\cdots$	$k(\cdot,j)$	$\cdots$	$k(\cdot,p)$	k

On transforme les effectifs en **fréquences** : on obtient un nouveau tableau F_{IJ} ou F de terme courant

$$\forall (i,j) \in I \times J, \quad f_{i,j} = \frac{k(i,j)}{k}.$$

On a les lois marginales :

$$f_I = (f_{i\cdot})_{i \in I} \in \mathbb{R}^p \text{ avec } f_{i\cdot} = \sum_{j=1}^{q} f_{i,j} = \frac{k(i\cdot)}{k},$$

$$f_J = (f.j)_{j \in J} \in \mathbb{R}^q \text{ avec } f_{\cdot j} = \sum_{i=1}^{p} f_{i,j} = \frac{k(\cdot j)}{k}.$$

f_I est la loi marginale sur I et f_j sur J. Ainsi f_I et f_J sont des distributions de probabilités donc

$$\sum_{i \in I} \sum_{j \in J} f_{i,j} = \sum_{i \in I} f_{i\cdot} = \sum_{j \in J} f_{\cdot j} = 1.$$

On peut aussi introduire la loi conditionnelle sur I sachant j appelé profil de la colonne j :

$$f_I^J = (f_i^j)_{i \in I, j \in J} \text{ avec } f_i^j = \frac{f_{i,j}}{f_{\cdot j}} = \frac{k(i,j)}{k(\cdot, j)},$$

Ainsi f_I^J est une matrice de format $p \times q$ et f_I^j est le $j$ième vecteur colonne de $\mathbb{R}^p$.

De même on a la loi conditionnelle sur J sachant i appelé profil de la ligne i :

$$f_J^I = (f_j^i)_{i \in I, j \in J} \text{ avec } f_j^i = \frac{f_{i,j}}{f_{i\cdot}} = \frac{k(i,j)}{k(i, \cdot)},$$

Ainsi f_J^I est une matrice de format $q \times p$ et f_J^i est le $i$ème vecteur de $\mathbb{R}^q$.

Puisque f_I^j et f_J^i sont des distributions de probabilités, on a

$$\sum_{i \in I} f_i^j = \sum_{j \in J} f_j^i = 1.$$

S'il n'y a pas d'ambiguité, on note f_i pour $f_{i\cdot}$, f_j pour $f_{\cdot j}$, $k(i)$ pour $k(i, \cdot)$ et $k(j)$ pour $k(\cdot, j)$.

On suppose qu'aucune ligne ou colonne de $K_{I,J}$ n'est nulle. Donc f_i et f_j sont non nulles et f_i^j et f_j^i sont bien définies.

2.2. Nuages et métriques. —

Définition 3.1 – AFC

L'AFC consiste à effectuer deux ACP sur deux nuages différents mais présentant une certaine symétrie. On note

$$D_{f_I} = Diag(f_i)_{i\in I} \in \mathcal{M}_p(\mathbb{R}) \text{ et } D_{f_J} = Diag(f_j)_{j\in J} \in \mathcal{M}_q(\mathbb{R}).$$

On a

$$D_{f_I}^{-1} = Diag(1/f_i)_{i\in I} = D_{1/f_I} \text{ et } D_{f_J}^{-1} = Diag(1/f_j)_{j\in J} = D_{1/f_J}.$$

On dit que D_{1/f_I} (respectivement D_{1/f_J}) est la métrique du chi-deux de centre f_I (respectivement f_J).

On considère les nuages suivants :

- $\mathcal{N}(J) = \{f_I^j, j \in J\}$**, appelé nuage des profils colonnes**, où chaque point f_I^j de $\mathbb{R}^p$ est muni du poids f_j et $\mathbb{R}^p$ est muni de la métrique D_{1/f_I}.
- $\mathcal{N}(I) = \{f_J^i, i \in I\}$**, appelé nuage des profils lignes**, où chaque point f_J^i de $\mathbb{R}^q$ est muni du poids f_i et $\mathbb{R}^q$ est muni de la métrique D_{1/f_J}.

On note

$$F_1 = f_I^J = (f_I^1, \cdots, f_I^q) \text{ et } F_2 = f_J^I = (f_J^1, \cdots, f_J^p).$$

F_1 est le tableau des profils colonnes et F_2 des profils lignes. On peut remarquer que F_1' et F_2' sont les matrices correspondantes à X.

Proposition 3.2

On a

$$F_1 = FD_{1/f_J} \textit{ et } F_2 = F'D_{1/f_I}.$$

On en déduit que le rang de F est égal au rang de F_1 et à celui de F_2.

Démonstration. — Les matrices D_{1/f_J} et D_{1/f_I} sont inversibles d'où le résultat. □

3. Nuage $\mathcal{N}(J)$

3.1. Support, centre de gravité. — Puisque l'on a pour tout $j \in J$

$$\sum_{i\in I} f_i^j = 1,$$

on en déduit que tous les points du nuage $\mathcal{N}(J)$ sont dans l'hyperplan affine de $\mathbb{R}^p$ d'équation

$$\sum_{i\in I} x_i = 1.$$

Proposition 3.3 – Centre de gravité

Le centre de gravité du nuage $\mathcal{N}(J)$ est f_I.
Le support du nuage $\mathcal{N}(J)$ est inclus dans l'hyperplan affine passant par f_I et D_{1/f_I}-orthogonal à f_I.

Démonstration. — En effet soit G_I ce centre de gravité, on a

$$G_I = f_I^J D_{f_J} 1_q = F1_q = f_I.$$

Par ailleurs

$$< f_I^j - f_I, f_I >_{D_{1/f_I}} = (f_I^j - f_I)' D_{1/f_I} f_I = (f_I^j - f_I)' 1_p = 0.$$

□

3.2. Effet du non centrage. — On effectue une ACP sur la matrice F_1' avec les métriques $M = D_{1/f_I}$ et $D_p = D_{f_J}$. Le centre de gravité des individus pondérés par D_p est f_I, et la matrice de variance V est

$$V = (F_1 - f_I 1_q') D_{f_J} (F_1 - f_I 1_q')' = F_1 D_{f_J} F_1' - f_I f_I'.$$

Le support de $\mathcal{N}(J)$ est inclus dans l'hyperplan affine d'équation $\sum_{i \in I} x_i = 1$, ce qui se traduit par

$$F_1' 1_p = 1_q.$$

Proposition 3.4

La matrice $VM = VD_{1/f_I}$ et la matrice $F_1 D_{f_J} F_1' D_{1/f_I}$ ont les mêmes vecteurs propres et

$$Spectre(VD_{1/f_I}) = \{\lambda_p = 0 \leq \lambda_{p-1} \cdots \leq \lambda_1\} \text{ et } Spectre(F_1 D_{f_J} F_1' D_{1/f_I}) = \{1, \lambda_{p-1}, \cdots, \lambda_1\}.$$

Démonstration. — En effet on a

$$F_1 D_{f_J} F_1' D_{1/f_I} f_I = F_1 D_{f_J} F_1' 1_p = F_1 D_{f_J} 1_q = f_I, \text{ et } f_I f_I' D_{1/f_I} f_I = f_I f_I' 1_p = f_I.$$

On en déduit que f_I est un vecteur propre associé à la matrice VD_{1/f_I} et à la matrice $F_1 D_{f_J} F_1' D_{1/f_I}$ avec les valeurs propres 0 et 1 respectivement. Comme ces matrices représentent des endomorphismes D_{1/f_I}-symétriques, l'orthogonal de $\mathrm{Vect}(f_I)$ est stable pour ces deux matrices. Or soit u un vecteur de $\mathrm{Vect}(f_I)^\perp$, on a

$$f_I f_I' D_{1/f_I} u = < f_I, u >_{D_{1/f_I}} f_i = 0.$$

Ainsi

$$\forall u \in \mathrm{Vect}(f_I)^\perp, \; VD_{1/f_I} u = F_1 D_{f_J} F_1' D_{1/f_I} u,$$

par conséquent la restriction à $\mathrm{Vect}(f_I)^\perp$ des endomorphismes représentés par VD_{1/f_I} et par $F_1 D_{f_J} F_1' D_{1/f_I}$ sont identiques donc les deux matrices ont mêmes valeurs propres et même vecteurs propres.

□

On en déduit que pour obtenir les axes factoriels de l'ACP, le centrage n'est pas nécéssaire. Pour le calcul des composantes principales, il n'est pas nécessaire de centrer non plus :

Remarque 3.5. — Soit u_I un axe factoriel orthogonal à f_I, la composante principale ψ^J associée à l'axe u_I est

$$\forall j \in J, \;\; \psi^j = < (f_I^j - f_I), u_I > = < f_I^j, u_I > .$$

A l'axe factoriel trivial f_I, on associe la composante triviale $\psi_o = F_1' D_{1/f_I} f_I = 1_q$.

3.3. Axes Factoriels, facteurs et composantes principales. — On le résultat suivant

Proposition 3.6 – Axes Factoriels, facteurs et composantes principales

L'ACP du nuage $\mathcal{N}(J)$ consiste à diagonaliser F_1F_2.
Les axes factoriels sont solutions de

$$\begin{cases} F_1F_2u_I^\alpha = \lambda_\alpha u_I^\alpha, \\ < u_I^\alpha, u_I^\beta >_{D_{1/f_I}} = \delta_{\alpha,\beta}, \\ < u_I^\alpha, f_I >_{D_{1/f_I}} = 0. \end{cases}$$

Les facteurs φ_α^I sont solutions de

$$\begin{cases} F_2'F_1'\varphi_\alpha^I = \lambda_\alpha \varphi_\alpha^I, \\ < \varphi_\alpha^I, \varphi_\beta^I >_{D_{f_I}} = \delta_{\alpha,\beta}, \\ < \varphi_\alpha^I, 1_I >_{D_{f_I}} = 0. \end{cases}$$

Les composantes principales ψ_α^J sont solutions de

$$\begin{cases} F_1'F_2'\psi_\alpha^J = \lambda_\alpha \psi_\alpha^J, \\ < \psi_\alpha^J, \psi_\beta^J >_{D_{f_J}} = \lambda_\alpha\delta_{\alpha,\beta}, \\ < \psi_\alpha^J, 1_J >_{D_{f_J}} = 0. \end{cases}$$

Toutes les valeurs propres λ_α sont positives et inférieures à 1.

Démonstration. — On réalise l'ACP de $\mathcal{N}(J)$ donc $X = F_1'$ pour que les individus soient en ligne, la matrice des poids D_p est D_{f_J} et la matrice M est D_{1/f_I}.

Pour trouver les axes factoriels, on diagonalise la matrice sans le centrage de F_1' soit $X'D_pXM$ ce qui donne

$$F_1D_{f_J}F_1'D_{1/f_I},$$

or $F_1 = FD_{1/f_J}$, on a

$$F_1D_{f_J}F_1'D_{1/f_I} = F_1D_{f_J}D_{1/f_J}F'D_{1/f_I}, = F_1F_2.$$

Les facteurs sont vecteurs propres de MV. En effet le facteur ϕ_α est par définition Mu_α donc

$$MV\phi_\alpha = MVMu_\alpha = \lambda_\alpha\, Mu_\alpha = \lambda_\alpha\phi_\alpha.$$

Pour la même raison que dans la propriété sur le non centrage, il n'est pas nécessaire de centrer

$$D_{1/f_I}F_1D_{f_J}F_1' = F_2'F_1' \text{ puisque } F_2 = F'D_{1/f_I}.$$

Enfin les composantes principales sont vecteurs propres de $YMY'D_p$ et comme le centrage n'est pas nécessaire, on a

$$F_1'D_{1/f_I}F_1D_{f_J} = F_1'F_2'.$$

Enfin les valeurs propres sont positives. De plus le terme courant (j,k) de $F_1'F_2'$ est

$$\sum_{i=1}^{p} f_i^j\, f_k^i$$

donc l'égalité $F_1'F_2'\psi = \lambda\psi$ devient

$$\sum_{k=1}^{q}\sum_{i=1}^{p} f_i^j\, f_k^i\psi(k) = \lambda\psi(j),$$

en notant $\psi(j_0)$ la plus grande coordonnée de ψ, on a

$$\lambda\psi(j) \leq \sum_{k=1}^{q}\sum_{i=1}^{p} f_i^j\, f_k^i\, \psi(j_0) = \psi(j_0),$$

on en déduit que $0 \leq \lambda \leq 1$.

□

4. Le nuage $\mathcal{N}(I)$

L'étude de $\mathcal{N}(I)$ se déduit de celle de $\mathcal{N}(J)$ en intervertissant les rôles de I et de J. Ainsi le centre de gravité de $\mathcal{N}(I)$ est f_J, le support de $\mathcal{N}(I)$ est inclus dans l'hyperplan affine d'équation $\sum_{j\in J} x_j = 1$. On échange F_1 et F_2, donc pour trouver les axes factoriels on diagonalise F_2F_1, les facteurs, on diagonalise $F_1'F_2'$ et les composantes principales, on diagonalise $F_2'F_1'$. On a donc

Proposition 3.7

L'ACP du nuage $\mathcal{N}(I)$ consiste à diagonaliser F_2F_1.
Les axes factoriels sont solutions de

$$\begin{cases} F_2F_1u_J^\alpha = \lambda_\alpha u_J^\alpha, \\ < u_J^\alpha, u_J^\beta >_{D_{1/f_J}} = \delta_{\alpha,\beta}, \\ < u_J^\alpha, f_J >_{D_{1/f_J}} = 0. \end{cases}$$

Les facteurs φ_α^J sont solutions de

$$\begin{cases} F_1'F_2'\varphi_\alpha^J = \lambda_\alpha \varphi_\alpha^J, \\ < \varphi_\alpha^J, \varphi_\beta^J >_{D_{f_J}} = \delta_{\alpha,\beta}, \\ < \varphi_\alpha^J, 1_J >_{D_{f_J}} = 0. \end{cases}$$

Les composantes principales ψ_α^I sont solutions de

$$\begin{cases} F_2'F_1'\psi_\alpha^I = \lambda_\alpha \psi_\alpha^I, \\ < \psi_\alpha^I, \psi_\beta^I >_{D_{f_I}} = \lambda_\alpha\delta_{\alpha,\beta}, \\ < \psi_\alpha^I, 1_I >_{D_{f_I}} = 0. \end{cases}$$

Toutes les valeurs propres λ_α sont positives et inférieures à 1.

La proposition suivante établit des relations entre les deux ACP :

Proposition 3.8 – Formules de transition

On a

$$\psi_\alpha^J = F_1'\varphi_\alpha^I = \sqrt{\lambda_\alpha}\,\varphi_\alpha^J, \text{ et } \psi_\alpha^I = F_2'\varphi_\alpha^J = \sqrt{\lambda_\alpha}\varphi_\alpha^I.$$

Démonstration. — Soit λ_α une valeur propre non nulle de $F_2'F_1'$

$$F_2'F_1'\varphi_\alpha^I = \lambda_\alpha\varphi_\alpha^I,$$

en multipliant par F_1', on obtient que $F_1'\varphi_\alpha^I$ est non nul et est donc un vecteur propre de $F_1'F_2'$. On normalise ce vecteur propre, pour cela on calcule sa norme

$$\begin{aligned} \|F_1'\varphi_\alpha^I\|^2 &= \varphi_\alpha^I F_1 D_{f_J} F_1'\varphi_\alpha^I, \\ &= \varphi_\alpha^{I\prime} F F_1'\varphi_\alpha^I, \\ &= \varphi_\alpha^{I\prime} D_{f_I} F_2'F_1'\varphi_\alpha^I, \\ &= \lambda_\alpha \|\varphi_\alpha^I\|^2, \\ &= \lambda_\alpha. \end{aligned}$$

Par conséquent $\dfrac{1}{\sqrt{\lambda_\alpha}}F_1'\varphi_\alpha^I$ est un vecteur propre unitaire de $F_1'F_2'$ associé à la valeur propre λ_α.

De plus soit k et l deux indices distincts, on a

$$\begin{aligned} < \frac{F_1'\varphi_k^I}{\sqrt{\lambda_k}}, \frac{F_1'\varphi_l^I}{\sqrt{\lambda_l}} > &= \frac{1}{\sqrt{\lambda_k\lambda_l}} \varphi_k^{I,} F_1 D_{f_J} F_1' \varphi_l^I, \\ &= \frac{1}{\sqrt{\lambda_k\lambda_l}} \varphi_k^{I,} D_{f_I} D_{1/f_I} F D_{1/f_J} F' \varphi_l^I, \\ &= \frac{1}{\sqrt{\lambda_k\lambda_l}} \varphi_k^{I,} D_{f_I} F_2' F_1' \varphi_l^I, \\ &= 0. \end{aligned}$$

On note r le nombre de valeurs propres non nulles de $F_2'F_1'$, c'est-à-dire le rang de $F_2'F_1'$. Ainsi l'image par F_1' de la base orthonormale $(\varphi_1^I, \cdots, \varphi_r^I, \varphi_{r+1}^I, \cdots, \varphi_p^I)$ de $\mathbb{R}^p$ muni de la métrique D_{f_I} donne une famille orthogonale que l'on peut normaliser soit $(\frac{F_1'\varphi_1^I}{\sqrt{\lambda_1}}, \cdots, \frac{F_1'\varphi_r^I}{\sqrt{\lambda_r}})$, ce qui donne une famille orthonormale de $\mathbb{R}^q$ muni de la métrique D_{f_J} constituée de vecteurs propres de $F_1'F_2'$.

On en déduit que le rang de $F_1'F_2'$ est supérieure à r. Par symétrie entre les deux analyses, on en déduit que $F_1'F_2'$ et $F_2'F_1'$ ont même rang et donc les mêmes valeurs propres non nulles.

Par conséquent pour toute valeur propre non nulle, on a

$$F_1'\varphi_\alpha^I = \sqrt{\lambda_\alpha}\,\varphi_\alpha^J,$$

d'où les formules de transition.

Pour une valeur propre nulle, $\lambda_\alpha = 0$, le calcul de la norme de $F_1'\varphi_\alpha^I$ montre que

$$F_1'\varphi_\alpha^I = 0.$$

Les formules de transition sont encore satisfaites.

□

Remarque 3.9. — Il existe diverses formulations des relations de transition. Par exemple si la valeur propre λ_α est non nulle, on peut écrire

$$\psi_\alpha^J = \frac{1}{\sqrt{\lambda_\alpha}} F_1' \psi_\alpha^I.$$

On en déduit

$$\forall j \in J, \;\; \psi_\alpha^j = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i \in I} f_i^j \psi_\alpha^i$$

De même en inversant i et j, on a aussi

$$\forall i \in I, \;\; \psi_\alpha^i = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j \in J} f_j^i \psi_\alpha^j.$$

On a aussi les même relations de transition pour les facteurs.

Représentation simultanée. En AFC, on effectue une représentation simultanée des modalités $i \in I$ et $j \in J$. Plus précisément, sur chaque axe α, on représente $i \in I$ par le point d'abscisse ψ_α^i et $j \in J$ par le point d'abscisse ψ_α^j. Autrement dit, on superpose les représentations des nuages $\mathcal{N}(I)$ et $\mathcal{N}(J)$ dans leurs systèmes d'axes respectifs. D'après les formules de transitions, il en résulte qu'au facteur $\frac{1}{\sqrt{\lambda_\alpha}}$ près, le point j est le barycentre des points i affectés des poids f_i^j. De même le point i est le barycentre des points j affectés des poids f_j^i.

5. Inerties

5.1. Inertie totale. — On a le résultat suivant :

Proposition 3.10 – Inertie totale

Les nuages $\mathcal{N}(I)$ et $\mathcal{N}(J)$ ont même inertie totale égale à

$$I_T = \sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-f_if_j)^2}{f_if_j}.$$

Démonstration. — On a

$$\begin{aligned} I_T &= \sum_{j\in J} f_j\|f_I^j - f_I\|^2_{D(1/f_I)},\\ &= \sum_{j\in J} f_j \sum_{i\in I}\frac{1}{f_i}(f_i^j - f_i)^2,\\ &= \sum_{j\in J}\sum_{i\in I}\frac{f_j}{f_i}(\frac{f_{i,j}}{f_j} - f_i)^2,\\ &= \sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-f_if_j)^2}{f_if_j}. \end{aligned}$$

□

Remarque 3.11. — En écrivant :

$$(f_{i,j}-f_if_j)^2 = f_{i,j}^2 - 2f_if_jf_{i,j} + f_i^2f_j^2,$$

et en remarquant que

$$\frac{-2f_if_jf_{i,j}+f_i^2f_j^2}{f_if_j} = -2f_{i,j} + f_if_j,$$

on en déduit que

$$\sum_{i\in I}\sum_{j\in J}\frac{-2f_if_jf_{i,j}+f_i^2f_j^2}{f_if_j} = -2\sum_{i\in I}\sum_{j\in J} f_{i,j} + \sum_{i\in I} f_i\sum_{j\in J} f_j = -2+1 = -1.$$

Par conséquent, on a

$$I_T = \sum_{i\in I}\sum_{j\in J}\frac{f_{i,j}^2}{f_if_j} - 1.$$

5.2. Interprétation de l'inertie totale dans le cas d'un tableau de contingence. — On suppose que K est un tableau de contingence, et plus précisément que I (resp. J) est l'ensemble des modalités d'une variable qualitative X (resp. Y). Ainsi K donne les effectifs de co-occurence des couples de modalités (i,j) sur un échantillon de taille k. Donc F est un estimateur de la mesure de probabilité théorique $p_{I,J}$ (loi jointe de (X,Y)). On sait alors que asymptotiquement, i.e. pour k tendant vers l'infini, on a

$$k\sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-p_{ij})^2}{p_{ij}} \to \chi^2_{pq-1},$$

où $p = \mathrm{Card}(I)$ et $q = \mathrm{Card}(J)$.

Lorsque l'on teste l'hypothèse :

$$H_0 : p_{IJ} = p_Ip_J,$$

H_0 représente l'hypothèse d'indépendance des variables aléatoires X et Y, on est amené à estimer les lois marginales p_I par f_I et p_J par f_J. Pour p_I, on estime $p-1$ paramètres puisque la somme des p_i vaut 1, de même pour p_J on estime $q-1$ paramètres. Il en résulte que

$$k\sum_{i\in I}\sum_{j\in J}\frac{(f_{i,j}-f_if_j)^2}{f_if_j} \to \chi^2_{\mu},$$

avec

$$\mu = pq - 1 - (p-1) - (q-1) = (p-1)(q-1).$$

On en conclut que la quantité

$$kI_T = k \sum_{i \in I} \sum_{j \in J} \frac{(f_{i,j} - f_i f_j)^2}{f_i f_j}$$

permet de tester l'hypothèse d'indépendance des variables X et Y.

Dans le cas d'indépendance, kI_T aura tendance à être faible ($kI_T \leq c_\alpha$), et par conséquent, puisque I_T est la somme des valeurs propres, plus les valeurs propres sont faibles moins les facteurs sont interprétables.

Si X et Y ne sont pas indépendants, l'AFC permet de voir comment f_{IJ} s'écarte de l'indépendance, les axes factoriels associés aux plus grandes valeurs propres traduisant les liaisons entre X et Y.

5.3. Décomposition de l'inertie, Contributions. — On exprime I_T en fonction des composantes principales des deux ACP, on note r le nombre de valeurs propres non nulles, on a

$$I_T = \sum_{\alpha=1}^{r} \sum_{j \in J} f_j (\psi_\alpha^j)^2 = \sum_{\alpha=1}^{r} \sum_{i \in I} f_i (\psi_\alpha^i)^2.$$

De plus la norme d'une composante principale valant la valeur propre, on a

$$\lambda_\alpha = \sum_{j \in J} f_j (\psi_\alpha^j)^2 = \sum_{i \in I} f_i (\psi_\alpha^i)^2.$$

En intervertissant les sommes

$$I_T = \sum_{j \in J} f_j \sum_{\alpha=1}^{r} (\psi_\alpha^j)^2 = \sum_{j \in J} f_j \rho^2(j),$$

où $\rho^2(j)$ est la distance au carré entre f_I^j et f_I.

De même

$$I_T = \sum_{i \in I} f_i \sum_{\alpha=1}^{r} (\psi_\alpha^i)^2 = \sum_{i \in I} f_i \rho^2(i),$$

où $\rho^2(i)$ est la distance au carré entre f_J^i et f_J. On en déduit les définitions suivantes :

Définition 3.12

La contribution de j et i à l'inertie de l'axe α sont respectivement :

$$CTR_\alpha(j) = \frac{f_j (\psi_\alpha^j)^2}{\lambda_\alpha} \text{ et } CTR_\alpha(i) = \frac{f_i (\psi_\alpha^i)^2}{\lambda_\alpha}.$$

La contribution de l'axe α à l'inertie de j et de i sont

$$COR_\alpha(j) = \frac{(\psi_\alpha^j)^2}{\rho^2(j)} = \cos^2(\theta_{j,\alpha}) \text{ et } COR_\alpha(i) = \frac{(\psi_\alpha^i)^2}{\rho^2(i)} = \cos^2(\theta_{i,\alpha}),$$

où $\theta_{i,\alpha}$ et $\theta_{j,\alpha}$ désignent respectivement les angles formés entre $f_J^i - f_J$ et u_J^α d'une part et entre $f_I^j - f_I$ et u_I^α d'autre part.

Exemple 3.13. — Montrer que les relations suivantes sont vérifiées :

$$\cos^2(\theta_{j,\alpha}) = corr^2(f_j^I, \varphi_\alpha^I) \text{ et } \cos^2(\theta_{i,\alpha}) = corr^2(f_i^J, \varphi_\alpha^J),$$

où $corr^2(f_j^I, \varphi_\alpha^I)$ et $corr^2(f_i^J, \varphi_\alpha^J)$ sont calculées respectivement avec les mesures de probabilté f_i et f_J.

6. Principe d'équivalence distributionnelle

Proposition 3.14 – Principe d'équivalence distributionnelle

Si deux lignes i_1 et i_2 (resp. colonnes j_1 et j_2) du tableau f_{IJ} ou k_{IJ} sont proportionnelles, alors on ne change pas les résultats de l'analyse des correspondances en remplaçant ces deux lignes (resp. colonnes) par leur somme i_0 (resp. j_0) affectée de la somme de leurs poids :

$$\forall j \in J, \quad f_{i_0 j} = f_{i_1 j} + f_{i_2 j}.$$

Démonstration. — On suppose que les deux lignes i_1 et i_2 du tableau f_{IJ} ou k_{IJ} sont proportionnelles, alors il existe un réel a tel que

$$\forall j \in J, \quad f_{i_1 j} = a \times f_{i_2 j},$$

on en déduit que

$$f_{i_1 \cdot} = \sum_{j \in J} f_{i_1 j} = a \times \sum_{j \in J} f_{i_2 j} = a f_{i_2 \cdot}.$$

Ainsi dans le nuage $\mathcal{N}(I)$, les deux profils lignes i_1 et i_2 sont confondus :

$$\forall j \in J, \quad f_j^{i_1} = \frac{f_{i_1 j}}{f_{i_1 \cdot}} = \frac{a f_{i_2 j}}{a f_{i_2 \cdot}} = f_j^{i_2}.$$

Par conséquent l'ACP du nuage $\mathcal{N}(I)$ n'est pas modifié si l'on réunit les deux individus i_1 et i_2 en un individu i_0 affecté du poids $f_{i_0 \cdot} = f_{i_1 \cdot} + f_{i_2 \cdot}$:

$$f_J^{i_0} = f_J^{i_1} = f_J^{i_2}.$$

Ainsi pour tout $j \in J$,

$$\begin{aligned} f_{i_0 j} &= f_{i_0 \cdot} f_j^{i_0}, \\ &= f_{i_1 \cdot} f_j^{i_0} + f_{i_2 \cdot} f_j^{i_0}, \\ &= f_{i_1 \cdot} f_j^{i_1} + f_{i_2 \cdot} f_j^{i_2}, \\ &= f_{i_1 j} + f_{i_2 j}. \end{aligned}$$

Pour le nuage $\mathcal{N}(J)$ initial, les distances entre les colonnes j et j' sont

$$d^2(j, j') = \sum_{i \in I} \frac{1}{f_{i \cdot}} (f_i^j - f_i^{j'})^2,$$

et pour le nuage en tenant compte du regroupement des lignes, on a

$$d^2(j, j') = \sum_{i \in I \setminus \{i_1, i_2\}} \frac{1}{f_{i \cdot}} (f_i^j - f_i^{j'})^2 + \frac{1}{f_{i_0 \cdot}} (f_{i_0}^j - f_{i_0}^{j'})^2.$$

Or

$$\begin{aligned} \frac{1}{f_{i_0 \cdot}} (f_{i_0}^j - f_{i_0}^{j'})^2 &= f_{i_0 \cdot} (\frac{f_{i_0}^j - f_{i_0}^{j'}}{f_{i_0 \cdot}})^2, \\ &= (f_{i_1 \cdot} + f_{i_2 \cdot})(\frac{f_j^{i_0}}{f_{\cdot j}} - \frac{f_{j'}^{i_0}}{f_{\cdot j'}})^2, \\ &= f_{i_1 \cdot} (\frac{f_j^{i_0}}{f_{\cdot j}} - \frac{f_{j'}^{i_0}}{f_{\cdot j'}})^2 + f_{i_2 \cdot} (\frac{f_j^{i_0}}{f_{\cdot j}} - \frac{f_{j'}^{i_0}}{f_{\cdot j'}})^2, \\ &= f_{i_1 \cdot} (\frac{f_j^{i_1}}{f_{\cdot j}} - \frac{f_{j'}^{i_1}}{f_{\cdot j'}})^2 + f_{i_2 \cdot} (\frac{f_j^{i_2}}{f_{\cdot j}} - \frac{f_{j'}^{i_2}}{f_{\cdot j'}})^2. \end{aligned}$$

Par conséquent les distances entre les individus j et j' sont les mêmes dans les situations, donc l'ACP de $\mathcal{N}(J)$ est identique dans les deux situations.

□

Remarque 3.15. — Cette propriété garantit une invariance des résultats vis à vis du choix de la momenclature pour la construction des modalités d'une variable, sous réserve de regrouper des modalités aux profils similaires.

7. Tableau de Burt

Définition 3.16 – Tableau de Burt pour un tableau de contingence

On appelle tableau de Burt associé à un tableau de contingence k_{IJ} le tableau B

$$\forall (j, j') \in J^2, \;\; b_{jj'} = \sum_{i \in I} \frac{k_{ij} k_{ij'}}{k_{i\cdot}}.$$

Matriciellement, on a

$$B = K' D_{1/k_I} K.$$

Proposition 3.17

L'AFC de B donne les mêmes axes factoriels u_J^α et les mêmes facteurs de variance φ_α^J que ceux obtenus dans l'AFC de k_{IJ}. Les valeurs propres de l'AFC de B sont les carrés des valeurs propres de L'AFC de k_{IJ}.

Démonstration. — Pour l'AFC de B, on remarque que par interversion de somme

$$\sum_{j'=1}^{q} b_{jj'} = \sum_{i \in I} k_{ij} \sum_{j'=1}^{q} \frac{k_{ij'}}{k_{i\cdot}} = k(\cdot, j) \text{ et } \sum_{j=1}^{q} \sum_{j'=1}^{q} b_{jj'} = k.$$

Ainsi ce qui joue le rôle de F dans l'AFC de B est la matrice notée G

$$G = \frac{1}{k} B.$$

La matrice G est symétrique donc les deux marges sont égales. On note g_J cette marge commune, on a d'après le calcul précédent

$$\forall j \in J, \;\; g_{\cdot j} = g_{j\cdot} = f_{\cdot j}.$$

Donc $g_J = f_J$. Les matrices profil ligne et profil colonne, G_1 et G_2 sont

$$G_1 = G D_{1/f_J} = F' D_{1/f_I} F D_{1/f_J} = F_2 F_1 \text{ et } G_2 = G' D_{1/f_J} = G_1.$$

Par conséquent l'AFC de B revient à diagonaliser $G_1 G_2 = (F_2 F_1)^2$. On conclut avec des résultats classique de diagonalisation. □

Remarque 3.18. — Si l'on veut représenter l'ensemble I, il suffit de rajouter f_{IJ} en supplémentaire au tableau $g_{II} = G$. On obtient la même représentation que dans l'AFC de f_{IJ}.

Soit $G = (g_{jj'})_{j,j' \in J}$ le tableau défini par

$$G = F' D_{1/f_I} F.$$

On a

$$\forall (j, j') \in J^2, \;\; g_{jj'} = \sum_{i \in I} \frac{f_{ij} f_{ij'}}{f_{i\cdot}}.$$

CHAPITRE 4

ANALYSE DES CORRESPONDANCES MULTIPLES

1. Notations-Tableau disjonctif complet-tableau de Burt

1.1. Notations et définitions. — On note :

Q	:	ensemble de questions ou de variables qualitatives,
I	:	ensemble des individus qui ont répondu aux questions, avec $n = \mid I \mid$,
J	:	ensemble de toutes les modalités de réponse à toutes les questions, avec $p = \mid J \mid$,
J_q	:	ensemble de toutes les modalités de réponse à la question q,
k_{IJ}	:	tableau de taille $n \times p$ défini par

$$k(i,j) = \begin{cases} 1 & \text{si l'individu } i \text{ a adopté la modalité } j \text{ de } J, \\ 0 & \text{sinon.} \end{cases}$$

Définition 4.1

Le tableau k_{IJ} est dit *disjonctif* si chaque individu choisit au plus une modalité par question (deux modalités d'une même question s'excluent mutuellement). Le tableau k_{IJ} est dit *complet* si chaque individu choisit au moins une modalité par question.

Proposition 4.2

Un tableau k_{IJ} est disjonctif complet (TDC) si et seulement si :

$$\sum_{j \in J_q} k(i,j) = 1 \text{ pour toute question } q \in Q \text{ et tout individu } i \in I.$$

Démonstration. — En effet si le tableau k_{IJ} est disjonctif complet, alors soit q une question, alors tout individu i doit choisir une modalité et une seule parmi les modalités de la question q. Ainsi la réponse de l'individu i est une suite de 0 avec un seul 1, on en déduit que la somme $\sum_{j \in J_q} k(i,j) = 1$.

Réciproquement, si pour toute question q et tout individu i, on a l'égalité $\sum_{j \in J_q} k(i,j) = 1$, alors comme $k(i,j)$ est un entier qui vaut 0 ou 1, on en déduit que pout tout individu i, il existe un seul entier j_0 de J_q tel que $k(i,j_0) = 1$ et $k(i,j) = 0$ pour tout j de J_q différent de j_0. Donc l'individu i a choisi la modalité j_0 de J_q et n'a pas choisi les autres. Le tableau est donc disjonctif complet. $\square$

Définition 4.3 – ACM

Une analyse des correspondances multiples d'un tableau k_{IJ} disjonctif, complet est une analyse factorielle des correspondances du tableau k_{IJ}.

1.2. Propriétés des tableaux disjonctifs complets. —

Proposition 4.4

Pour tout individu $i \in I$, toute modalité $j \in J$ et toute question $q \in Q$, on a :

$$\begin{aligned} k(i) &= \operatorname{Card} Q, \\ k(j) &= \sum_{i \in I} k(i,j) = \text{ nombre d'individus ayant choisi la modalité } j, \\ \sum_{j \in J_q} k(j) &= n, \\ k &= n \operatorname{Card} Q. \end{aligned}$$

Démonstration. — On a

$$k(i) = \sum_{j \in J} k(i,j) = \sum_{q \in Q} \sum_{j \in J_q} k(i,j) = \sum_{q \in Q} 1 = \operatorname{Card} Q.$$

Par définition on a

$$k(j) = \sum_{i \in I} k(i,j) = \text{ nombre d'individus ayant choisi la modalité } j.$$

Puis on a

$$\sum_{j \in J_q} k(j) = \sum_{i \in I} \sum_{j \in J_q} k(i,j) = \sum_{i \in I} 1 = n.$$

Enfin

$$k = \sum_{i \in I} \sum_{j \in J} k(i,j) = \sum_{i \in I} k(i) = n \operatorname{Card} Q.$$

□

Exemple 4.5. — On suppose que l'on a 3 questions avec 2 modalités pour la première, 3 pour la seconde et 4 pour la troisième. Si l'on interroge n individus, le tableau K sera une matrice de format $n \times 9$, car 9 est le nombre de modalités pour toutes les questions, et sera du type

	J_1	J_2	J_3	total
1				
$\vdots$				
i	0 1	0 0 1	0 0 1 0	$k(i) = \operatorname{Card} Q = 3$
$\vdots$				
n				
total			$k(j)$	$k = 3n$

Le nuage $\mathcal{N}(I)$ est un ensemble de points de $\mathbb{R}^9$. Or d'après le tableau, la somme des deux premières coordonnées donne 1, puis celle des trois suivantes 1 encore et enfin des quatre dernières 1 aussi. Ce qui signifie que le nuage $\mathcal{N}(I)$ est inclus dans un sous espace affine de dimension $9 - 3 = 6$.

Dans le cas général, le nuage $\mathcal{N}(I)$ est inclus dans un sous espace affine de dimension $card(J) - card(Q)$.

2. Tableau de Burt

Définition 4.6 – Tableau de Burt pour un tableau disjonctif complet

Le tableau de Burt associé à un tableau disjonctif complet k_{IJ}, noté B_{JJ}, est défini pour tout $j, j' \in J$ par :

$$B(j,j') = \sum_{i\in I} k(i,j)\,k(i,j') = \text{ nombre d'individus qui ont choisi les modalités } j \text{ et } j'.$$

Si $j, j' \in J_q$, alors

$$B(j,j') = \begin{cases} 0 & \text{si } j \neq j' \\ k(j) & \text{sinon.} \end{cases}$$

Matriciellement, on a

$$B = K'K.$$

Proposition 4.7

Pour toute modalité $j \in J$ et toute question $q \in Q$, on a :

$$\begin{aligned} \sum_{j'\in J_q} B(j,j') &= k(j), \\ B(j) &= \sum_{j'\in J} B(j,j') = k(j)\ \mathrm{Card}\, Q, \\ B &= n(\mathrm{Card}\, Q)^2. \end{aligned}$$

Démonstration. — On a

$$\begin{aligned} \sum_{j'\in J_q} B(j,j') &= \sum_{j'\in J_q} \sum_{i\in I} k(i,j)\,k(i,j'), \\ &= \sum_{i\in I} k(i,j) \sum_{j'\in J_q} k(i,j'), \\ &= \sum_{i\in I} k(i,j), \\ &= k(j). \end{aligned}$$

Puis

$$\begin{aligned} B(j) &= \sum_{j'\in J} B(j,j'), \\ &= \sum_{q\in Q} \sum_{j'\in J_q} B(j,j'), \\ &= \sum_{q\in Q} k(j), \\ &= k(j)\ \mathrm{Card}\, Q. \end{aligned}$$

Enfin

$$B = \sum_{j\in J, j'\in J} B(j,j') = \sum_{j\in J} k(j)\ \mathrm{Card}\, Q = k\ \mathrm{Card}\, Q = n(\mathrm{Card}\, Q)^2.$$

□

Exemple 4.8. — On reprend l'exemple précédent, le tableau de Burt est alors un tableau de tableaux de contingence où toutes les questions sont croisées deux à deux. Dans ce cas, le tableau de Burt est une matrice 9×9 du type :

	J_1	J_2	J_3	total
J_1	$\begin{pmatrix} k(1) & 0 \\ 0 & k(2) \end{pmatrix}$	$\begin{pmatrix} B(1,3) & B(1,4) & B(1,5) \\ B(2,3) & B(2,4) & B(2,5) \end{pmatrix}$	$\begin{pmatrix} B(1,6) & B(1,7) & B(1,8) & B(1,9) \\ B(2,6) & B(2,7) & B(2,8) & B(2,9) \end{pmatrix}$	$3k(1)$ $3k(2)$
J_2	$\begin{pmatrix} B(3,1) & B(3,2) \\ B(4,1) & B(4,2) \\ B(5,1) & B(5,2) \end{pmatrix}$	$\begin{pmatrix} k(3)) & 0 & 0 \\ 0 & k(4) & 0 \\ 0 & 0 & k(5) \end{pmatrix}$		
J_3			$\begin{pmatrix} k(6)) & 0 & 0 & 0 \\ 0 & k(7) & 0 & 0 \\ 0 & 0 & k(8) & 0 \\ 0 & 0 & 0 & k(9) \end{pmatrix}$	$B(j) = 3k(j)$
total				$B = 9n$

On remarque que lorsqu'une question est croisée avec elle-même, le tableau de contingence correspond à une matrice diagonale.

3. Propriétés de l'AFC d'un questionnaire

Proposition 4.9

Soit F^I_α (resp. G^J_α) les projections des profils-lignes (resp. profils-colonnes) sur l'axe de rang α issu de l'AFC de k_{IJ}. Soit $F^J_{B\alpha}$ (resp. $G^J_{B\alpha}$) les projections des profils-lignes (resp. profils-colonnes) sur l'axe de rang α issu de l'AFC de B_{JJ}. On a :

$$F^J_{B\alpha} = G^J_{B\alpha} = \sqrt{\lambda_\alpha}\, G^J_\alpha.$$

Par ailleurs, pour tout $i \in I$, on a :

$$F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j\in J} \frac{k(i,j)}{\operatorname{Card} Q} G_\alpha(j).$$

En notant $q(i)$ la modalité j de la question $q \in Q$ choisie par l'individu i, on obtient :

$$F_\alpha(i) = \frac{1}{\operatorname{Card} Q} \sum_{q\in Q} \frac{G_\alpha(q(i))}{\sqrt{\lambda_\alpha}}.$$

Remarque 4.10. — La dernière égalité de cette proposition exprime que $F_\alpha(i)$ est égal à la moyenne des $\dfrac{G^{q(i)}_\alpha}{\sqrt{\lambda_\alpha}}$, coordonnées "normalisées" des modalités qui ont été choisies par l'individu i. Autrement dit encore, sur chaque axe, la représentation de chaque individu coïncide avec la moyenne des modalités qu'il a choisies à $1/\sqrt{\lambda_\alpha}$ près.

Démonstration. — On note F_1 la matrices des profils colonnes et F_2 des profils lignes de k_{IJ} et B_1 la matrices des profils colonnes et B_2 des profils lignes de B_{JJ}. Le tableau étant disjonctif et complet, on a

$$F_1 = FD_{1/f_J} \text{ et } F_2 = F'D_{1/f_J} = nF'.$$

Puisque le tableau de Burt est une matrice symétrique, on en déduit que $B_1 = B_2$. De plus, on constate que

$$\frac{B(j)}{B} = \frac{k(j)}{n \operatorname{Card} Q} = \frac{k(j)}{k}.$$

La marge selon J du tableau k_{IJ} est égale à la marge selon J du tableau B_{JJ} . Donc les métriques des AFC de k_{IJ} et B_{JJ} dans l'espace $\mathbb{R}^J$ sont identiques. On en déduit que

$$B_1 = B_2 = \frac{B}{n \operatorname{Card} Q^2} D_{1/f_J} = \frac{K'K}{n \operatorname{Card} Q^2} D_{1/f_J} = \frac{k^2}{n \operatorname{Card} Q^2} F'FD_{1/f_J} = nF'FD_{1/f_J} = F_2F_1.$$

Ainsi pour réaliser l'AFC de k_{IJ} on diagonalise F_2F_1 et pour réaliser l'AFC de B_{JJ} on diagonalise $B_1B_2 = (F_2F_1)^2$. Par conséquent tout vecteur propre de F_2F_1 est vecteur propre de $(F_2F_1)^2$ et les valeurs propres de $(F_2F_1)^2$ sont celles de F_2F_1 élevées au carré. De plus les métriques étant identiques, les vecteurs unitaires représentant les axes factoriels sont les mêmes dans les deux AFC. Pour les composantes principales du nuages des profils colonnes de k_{IJ}, ce sont des vecteurs propres de $F_1'F_2' = (F_2F_1)'$, la métriques étant identiques on a que $F^J_{B\alpha}$ et G^J_α sont colinéaires et comme

$$\|F^J_{B\alpha}\|_{D_{f_J}} = \lambda_\alpha \text{ et } \|G^J_\alpha\|_{D_{f_J}} = \sqrt{\lambda_\alpha}.$$

On en déduit

$$F^J_{B\alpha} = G^J_{B\alpha} = \sqrt{\lambda_\alpha}\, G^J_\alpha.$$

Les formules de transition entraînent que, pour tout $i \in I$, on a :

$$F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} F_2' G^J_\alpha = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j\in J} \frac{k(i,j)}{k(i)} G_\alpha(j) = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{j\in J} \frac{k(i,j)}{\mathrm{Card}\, Q} G_\alpha(j).$$

Avec les notations proposées on a donc

$$F_\alpha(i) = \frac{1}{\mathrm{Card}\, Q} \sum_{q\in Q} \frac{G_\alpha(q(i))}{\sqrt{\lambda_\alpha}}.$$

□

Proposition 4.11

Lors de l'AFC de k_{IJ}, le centre de gravité des profils f^j_I pour $j \in J_q$ est confondu avec le centre de gravité global.

$$\sum_{j\in J_q} f_{.j} G_\alpha(j) = 0.$$

Démonstration. — Du fait de la structure en blocs des tableaux k_{IJ} et B_{JJ}, on a pour toute question $q \in Q$:

$$\sum_{j\in J_q} f_{.j} G_\alpha(j) = 0.$$

□

Remarque 4.12. — On a le même résultat pour les profils (lignes ou colonnes) du tableau (symétrique) B_{JJ}.

Remarque 4.13. — En pratique, on effectue l'AFC de B_{JJ} et on met le tableau k_{IJ} en supplémentaire. On a alors :

$$\begin{aligned} G_{B\alpha}(j) &= \sqrt{\lambda_\alpha}\, G_\alpha(j) = \sum_{i\in I} \frac{k(i,j)}{k(j)} F_\alpha(i) \\ G_{B\alpha}(j) &= \sum_{q(i)=j} \frac{F_\alpha(i)}{k(j)}. \end{aligned}$$

Autrement dit, pour tout axe factoriel, chaque modalité $j \in J$ est représentée par le centre de gravité des individus l'ayant choisie.

4. Contributions en ACM

On a une expression pour l'inertie totale.

Proposition 4.14

Dans le cadre d'une ACM, l'inertie totale est donnée par :

$$I_T = \frac{\mathrm{Card}\, J}{\mathrm{Card}\, Q} - 1.$$

Remarque 4.15. — Ce résultat ne dépend que du nombre de questions et du nombre de modalités. Cette situation est très différente de l'AFC où l'inertie totale est le chi deux à un facteur près et donc très liée aux données, par contre ce résultat est plus proche d'une ACP normée où l'inertie totale est le nombre de variable.

Démonstration. — L'inertie totale est donnée par

$$I_T = \sum_{j\in J} f_{.j} \|f_I^j - f_I\|^2_{D_{1/f_I}}$$

On pose

$$\rho^2(j) = \|f_I^j - f_I\|^2_{D_{1/f_I}}$$

On a alors

$$I_T = \sum_{q\in Q}\sum_{j\in J_q} f_{.j}\rho^2(j),$$

On note $p(j)$ la proportion des individus ayant adopté la modalité j, on a

$$p_j = \frac{k(j)}{n}.$$

On démontre alors que

$$\rho^2(j) = \frac{1-p_j}{p_j}.$$

En effet en utilisant l'égalité $k(i,j)^2 = k(i,j)$, on obtient :

$$\begin{aligned}
\rho^2(j) &= \sum_{i=1}^{n} \frac{1}{f_i}(f_i^j - f_i)^2, \\
&= n\sum_{i=1}^{n} (\frac{k(i,j)}{k(j)} - \frac{1}{n})^2, \\
&= n\sum_{i=1}^{n} \frac{k(i,j)}{k(j)^2} - 2\frac{k(i,j)}{nk(j)} + \frac{1}{n^2}, \\
&= \frac{n}{k(j)} - 2 + 1, \\
&= \frac{1}{p_j} - 1.
\end{aligned}$$

On pose

$$CR(j) = f_{.j}\rho^2(j), \text{ et } CR(J_q) = \sum_{j\in J_q} f_{.j}\rho^2(j).$$

Comme $f_{.j} = \frac{k(j)}{k} = \frac{p_j}{\mathrm{Card}\, Q}$, on a

$$CR(j) = \frac{1-p_j}{\mathrm{Card}\, Q},$$

d'où

$$CR(J_q) = \frac{\text{Card}\, J_q - 1}{\text{Card}\, Q},$$

et on en conclut que

$$I_T = \frac{\text{Card}\, J}{\text{Card}\, Q} - 1.$$

□

Remarque 4.16. — La contribution d'une modalité est d'autant plus forte qu'elle est rare. La contribution d'une question dépend du nombre de modalités. Pour une question avec beaucoup de modalités, chaque modalité est représenté dans un espace de dimension plus élevé ce qui la comparaison avec des questions avec peu de modalités possible.

Proposition 4.17 – Décompositions en fonction des axes

On a

$$I_T = \sum_\alpha \lambda_\alpha = \sum_\alpha \sum_{q\in Q} \sum_{j\in J_q} f_{\cdot j} G_\alpha^2(j),$$

on pose

$$CR_\alpha(j) = f_{\cdot j} G_\alpha^2(j), \quad CR_\alpha(J_q) = \sum_{j\in J_q} f_{\cdot j} G_\alpha^2(j).$$

On pose

$$CTR_\alpha(q) = \frac{CR_\alpha(J_q)}{\lambda_\alpha}$$

est la contribution relative de J_q à l'inertie de l'axe α. On peut poser

$$COR_\alpha(q) = \frac{CR_\alpha(J_q)}{CR(J_q)}, \quad QLT(q) = \sum_\alpha COR_\alpha(q),$$

$$INR(q) = \frac{CR(J_q)}{CR(J)} = \frac{CR(J_q)}{I_T}.$$

Règles d'interprétation

1. Proximité entre individus : deux individus se ressemblent s'ils ont choisi les mêmes modalités.
2. Proximité entre deux modalités de variables différentes : ces modalités correspondent aux points moyens des individus les ayant choisies et sont proches parce qu'elles concernent les mêmes individus ou des individus semblables.
3. Proximité entre deux modalités d'une même variable : par construction, elles s'excluent. Si elles sont proches, c'est que les groupes des individus les ayant choisies se ressemblent.

APPENDICE A

ESPACE AFFINE

1. Définitions

Définition A.1 – Espace affine

Soit E un espace vectoriel, on dit que $\mathcal{E}$ est un espace affine de direction E si il existe une application f de $\mathcal{E} \times \mathcal{E}$ dans E notée

$$\forall (A, B) \in \mathcal{E} \times \mathcal{E}, \quad f((A, B)) = \overrightarrow{AB},$$

vérifiant les deux conditions suivantes

- A1 : Relation de Chasles

$$\forall (A, B, C) \in \mathcal{E} \times \mathcal{E} \times \mathcal{E}, \quad \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC},$$

- A2 : Pour tout $A \in \mathcal{E}$, l'application f_A définie de $\mathcal{E}$ dans E par

$$\forall M \in \mathcal{E}, \quad f_A(M) = \overrightarrow{AM} \text{ est une bijection .}$$

Les éléments de $\mathcal{E}$ sont appelés points et ceux de E vecteurs. On appelle dimension de $\mathcal{E}$ la dimension de E.

Remarque A.2. — Pour tout entier n non nul, $\mathbb{R}^n$ est un espace affine de direction $\mathbb{R}^n$ espace vectoriel. Ainsi la notation $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ peut être vu comme un vecteur de $\mathbb{R}^n$ ou un point de l'espace affine $\mathbb{R}^n$.

Notations : Soit $A \in \mathcal{E}$ et $u \in E$, $A + u$ désigne l'unique point B de $\mathcal{E}$ tel que $\overrightarrow{AB} = u$. Ainsi

$$\forall (A, B) \in \mathcal{E} \times \mathcal{E}, \quad \forall u \in E, \quad \overrightarrow{AB} = u \Longleftrightarrow B = A + u \Longleftrightarrow B - A = u.$$

Définition A.3

On considère $\mathcal{E}$ un espace affine de direction E, on dit que $\mathcal{F}$ est un sous-espace affine si il existe un point A de $\mathcal{E}$ et un sous-espace vectoriel F de E tels que

$$\mathcal{F} = A + F = \{M \in \mathcal{E}, \, \exists u \in F, M = A + u\}.$$

La dimension de $\mathcal{F}$ est celle de F.

Exemple A.4. — Une droite affine de $\mathcal{E}$ est un sous-espace affine de dimension 1. Dans ce cas $F = \text{Vect}(u)$ où u est non nul, soit A un point de la droite affine, on note $\mathcal{D}_{A,u}$ la droite affine passant par A de direction

Vect(u). On dit encore que u est un vecteur directeur de la droite affine $\mathcal{D}_{A,u}$, on a

$$\mathcal{D}_{A,u} = \{M \in \mathcal{E},\ \ \exists t \in \mathbb{R},\ \ M = A + t\,u\}.$$

On peut aussi définir une droite affine par deux points distincts A et B, alors la droite affine passant par A et B est $\mathcal{D}_{A,\overrightarrow{AB}}$

Exemple A.5. — Un plan affine de $\mathcal{E}$ est un sous-espace affine de dimension 2. Dans ce cas $F = \mathrm{Vect}(u,v)$ où u et v sont des vecteurs non colinéaires, soit A un point du plan affine, on note $\mathcal{P}_{A,(u,v)}$ le plan affine passant par A de direction Vect(u, v). On a

$$\mathcal{P}_{A,(u,v)} = \{M \in \mathcal{E},\ \ \exists (t,s) \in \mathbb{R}^2,\ \ M = A + t\,u + s\,v\}.$$

On peut aussi définir un plan affine par trois points non alignés A, B et C, alors le plan affine passant par A, B et C est $\mathcal{P}_{A,(\overrightarrow{AB},\overrightarrow{AC})}$

Exemple A.6. — On considère le système linéaire

$$AX = b \text{ où } A \in \mathcal{M}_{n,p}(\mathbb{R}),\ \ X = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p,\ \ b \in \mathbb{R}^n.$$

On suppose qu'il existe une solution particulière X_0, alors l'ensemble des solutions du système linéaire est le sous-espace affine $X_0 + \mathrm{Ker}\,A$ de $\mathbb{R}^p$, de dimension $\dim \mathrm{Ker}\,A = p - rg(A)$, où $rg(A)$ est le rang de A.

Exemple A.7. — Un hyperplan affine de $\mathcal{E}$ est un sous-espace affine de dimension $dimE - 1$.

Remarque A.8. — Lorsque l'on fixe un point O dans un espace affine $\mathcal{E}$ de direction E, on vectorialise l'espace affine, c'est-à-dire à l'aide de la fonction f_O on construit une structure d'espace vectoriel sur $\mathcal{E}$, tout point M de $\mathcal{E}$ est assimilé au vecteur $\overrightarrow{OM}$.

2. Barycentre

Définition A.9 – Barycentre

On considère $\mathcal{E}$ un espace affine de direction E, soit $M_1, \cdots, M_n$ n points de $\mathcal{E}$, et pour tout $1 \leq i \leq n$, on affecte à chaque point M_i un coefficient ou poids p_i qui est un réel. Soit O une origine,

- si $\displaystyle\sum_{i=1}^{n} p_i = 0$, alors le vecteur $\displaystyle\sum_{i=1}^{n} p_i\overrightarrow{OM_i}$ est indépendant de O.
- si $\displaystyle\sum_{i=1}^{n} p_i = p \neq 0$, alors le point G défini par

$$G = \frac{1}{p}\sum_{i=1}^{n} p_i M_i = O + \frac{1}{p}\sum_{i=1}^{n} p_i\overrightarrow{OM_i} \text{ est indépendant de O.}$$

On dit que g est le barycentre des $(M_i, p_i)_{1 \leq i \leq n}$.

Exemple A.10. — Le milieu de deux points A et B est le barycentre de A et B affectés des poids $1/2$ et $1/2$.

Définition A.11

On considère $\mathcal{E}$ un espace affine de direction E, soit $\mathcal{M} = \{M_1, \cdots, M_n\}$ un ensemble de n points de $\mathcal{E}$, on note $< \mathcal{M} >$ l'ensemble des barycentres des points de $\mathcal{M}$ affectés de poids quelconques. Alors $< \mathcal{M} >$ est un sous-espace affine de $\mathcal{E}$. On dit que $< \mathcal{M} >$ est le sous-espace affine engendré par $\mathcal{M}$. C'est le plus petit sous-espace affine contenant $\mathcal{M}$.

Proposition A.12

Le sous-espace affine engendre par $\mathcal{M} = \{M_1, \cdots, M_n\}$ est associé au sous-espace vectoriel $\mathrm{Vect}(\overrightarrow{M_1M_2}, \cdots, \overrightarrow{M_1M_n})$. *la dimension de $< \mathcal{M} >$ est au plus $n-1$.*

3. Applications affines

Définition A.13 – Applications affines

On considère $\mathcal{E}$ un espace affine de direction E, soit f une application de $\mathcal{E}$ dans $\mathcal{E}$. On dit que f est une application affine si il existe un point O de $\mathcal{E}$ tel que l'application $\vec{f}$ de E dans E qui à tout vecteur u de E associé le vecteur $f(O)f(O+u)$ est linéaire. On appelle $\vec{f}$ l'application linéaire associée à f.

Remarque A.14. — Une application affine f est caractérisée par sa valeur en un point et son application linéaire associée.

Exemple A.15. — Une translation de vecteur u est une application affine telle que

$$\forall M \in \mathcal{E}, \quad f(M) = M + u.$$

Exemple A.16. — Une projection orthogonale affine sur le sous-espace affine $\mathcal{F}$ est une application affine telle qu'il existe un point O de $\mathcal{F}$ vérifiant

$$\forall M \in \mathcal{E}, \quad f(M) = O + p(OM) \text{ où } p \text{ est la projection orthogonale linéaire sur F.}$$

Proposition A.17

On considère $\mathcal{E}$ un espace affine de direction E, soit f une application de $\mathcal{E}$ dans $\mathcal{E}$. l'application f est une application affine si et seulement si f conserve les barycentres c'est-à-dire pour tout entier n

$$\forall (x_i, t_i) \in \mathcal{E} \times \mathbb{R}, \ \textit{avec} \ \sum_{i=1}^{n} t_i = 1, \quad f(\sum_{i=1}^{n} t_i x_i) = \sum_{i=1}^{n} t_i f(x_i).$$

APPENDICE B

ENDOMORPHISME SYMÉTRIQUE

Définition B.1 – Endomorphisme symétrique

Soit E un espace euclidien muni d'une métrique M, et f un endomorphisme de E, on dit que f est un endomorphisme symétrique si pour tous x et y de E, on a l'égalité

$$< x, f(y) >_M = < f(x), y >_M .$$

Théorème B.2 – Théorème spectral

Tout endomorphisme symétrique est diagonalisable. De plus il existe une base orthonormale de E constituées de vecteurs propres de f.

La matrice d'un endomorphisme symétrique dans une base orthomormale quelconque de E est une matrice symétrique à coefficients réels. On en déduit la version matricielle du théorème spectral

Proposition B.3

Soit A une matrice symétrique à coefficients réels de $\mathcal{M}(\mathbb{R})$, alors A est diagonalisable, il existe donc une matrice diagonale D et P une matrice inversible tels que

$$A = PDP^{-1}.$$

De plus il est possible de choisir P orthogonale dans ce cas, l'égalité devient

$$A = PDP'.$$