UNIVERSITE D'ANTANANARIVO
DOMAINE SCIENCES ET TECHNOLOGIES
MENTION MATHEMATIQUES ET INFORMATIQUE
Mémoire pour l'obtention du Diplôme de Master
PARCOURS : MATHEMATIQUES APPLIQUÉES
SPECIALITE : COMBINATOIRE ET OPTIMISATION

MÉTHODES D'ANALYSE DES DONNÉES ET APPLICATION

Présenté par : **NIAVOMALALA Ravoniaina Harison**
Soutenu le : 30 Octobre 2020
Devant le commission d'examen formée de :

Président de Jury : Monsieur **RANDRIANARIVONY Arthur**
Professeur titulaire
Rapporteur : Monsieur **RAKOTONDRALAMBO Joseph**
Maître de conférences
Examinateur : Monsieur **RANDRIAMAHALEO Fanilo Rajaofetra**
Maître de conférences

REMERCIEMENTS

On remercie Dieu Le Tout Puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Avant de commencer le développement de cette expérience professionnelle, il me parait tout naturel de commencer par remercier les personnes qui m'ont permis d'effectuer ce travail ainsi que ceux qui m'ont permis d'en faire un moment agréable et profitable.

Je tiens à remercier également mon encadreur
Monsieur RAKOTONDRALAMBO Joseph pour le temps qu'il a consacré et pour les précieuses informations qu'il m'a prodiguées avec intérêt et compréhension.

J'adresse aussi mes vifs remerciements aux membres du jury :
Monsieur RANDRIANARIVONY Arthur président et l'examinateur
Monsieur RANDRIAMAHALEO Fanilo Rajaofetra pour avoir bien voulu examiner et juger ce travail.

Enfin, j'adresse mes plus sincères remerciements à ma famille : mes parents, tous mes proches et mes amis, qui m'ont accompagné, aidé, soutenu et encouragé tout au long de la réalisation de ce mémoire.

Table des matières

1 Analyse en Composantes Principales(ACP) 8

1.1 Introduction . . . 8
1.2 Qu'est-ce que l'ACP ? . . . 8
1.3 Domaines d'application . . . 9
1.4 Inertie d'un nuage de points . . . 12
1.4.1 Distance entre deux individus (ou variables) . . . 12
1.4.2 Inertie totale du nuage de points . . . 13
1.4.3 Inertie expliquée par un sous espace F. . . . 14
1.4.4 Décomposition de l'inertie totale . . . 14
1.5 Recherche du maximum . . . 15
1.5.1 Recherche des axes suivants . . . 17
1.5.2 Les composantes principales . . . 17
1.6 ACP dans l'espace des variables . . . 19
1.6.1 Facteurs principaux . . . 20
1.7 Les représentations graphiques. . . . 20
1.7.1 Résultats relatifs aux individus . . . 20
1.7.2 Résultats relatifs aux variables . . . 21
1.8 Aspects pratiques . . . 21
1.8.1 Interprétation . . . 21
1.8.2 Nombre d'axe à retenir . . . 22
1.9 Application . . . 22
1.9.1 Résultats généraux . . . 23

2 Analyse factorielle des correspondances (AFC) 26

2.1 Introduction . . . 26
2.2 Données . . . 26
2.3 Objectifs . . . 27
2.4 Principes de l'AFC . . . 27
2.4.1 Représentation géométrique des profils . . . 28
2.4.2 ACP sur un nuage de profils . . . 29
2.4.3 Interprétation . . . 34
2.5 Application . . . 35

3 Analyse des Correspondances Multiples(ACM) 37

3.1 Introduction . . . 37
3.2 Objectifs . . . 37
3.3 Définitions et notations . . . 38
3.3.1 Le Tableau disjonctif complet . . . 38
3.3.2 Tableau Burt . . . 38
3.3.3 Matrice des Fréquences . . . 40

3.4 AFC sur le tableau disjonctif complet . . . 40
3.5 Représentation Géométrique . . . 41
3.5.1 Etude des Individus . . . 41
3.5.2 Etude des modalités . . . 43
3.5.3 Coordonnnées factorielles des individus et des modalités . . . 45
3.5.4 Facteurs principaux . . . 45
3.5.5 Composantes principales . . . 45
3.5.6 Relations barycentriques . . . 46
3.6 Interprétation . . . 46
3.6.1 Nombre d'axes à retenir . . . 46
3.6.2 Contribution et Qualité . . . 46
3.7 Exemple . . . 46

4 Analyse Factorielle Discriminante 50
4.1 Données et Notations . . . 50
4.2 Axes, facteurs et variables discriminantes . . . 52
4.3 Exemple (Cas de deux groupes) . . . 54

5 Classification, segmentation 56
5.1 Introduction . . . 56
5.2 Les objectifs . . . 56
5.3 Tableau de données . . . 56
5.4 Mesures d'éloignement . . . 57
5.5 La classification hiérarchique . . . 58
5.5.1 Classification ascendante hiérarchique . . . 58
5.5.2 Classification Hiérarchique Descendante (CHD) . . . 61
5.6 Exemple . . . 62
5.7 Agrégation autour de centres mobiles . . . 66
5.8 EXEMPLE . . . 66
5.8.1 Variantes . . . 67
5.9 Combinaison . . . 67
5.10 La Classification des Données Qualitatives . . . 68
5.10.1 Caractères de natures différentes : . . . 69
5.11 Classification spectrale . . . 69
5.11.1 Construction de la matrice de similarités S . . . 70
5.11.2 Partitionnement du graphe . . . 70
5.11.3 Algorithme de classification spectrale . . . 70

A DERIVEES MATRICIELLES 72

B Décomposition de l'inertie Totale 73

C Méthode des multiplicateurs de Lagrange 74

Bibliographie 75

Liste des tableaux

1.1 Représentation matricielle des tableaux de données 8
1.2 Matrice des poids des individus 9
1.3 Matrice de Variance-covariance 10

2.1 Tableau de contingences 27
2.2 Tableau des profils des lignes et Colonnes 28
2.3 Tableau de Référence sur ACP 30
2.4 Tableau de présence/absence 35

3.1 Tableau des données sous forme de codage condensé. 37
3.2 Transformation des données en TDC sur l'ACM 38
3.3 Représentation des données sous forme du tableau de Burt. 38
3.4 Mise en fréquences du tableau disjonctif complet. 40
3.5 Représentation des Profls-lignes 40
3.6 Représentation des Profls-colonnes 41

4.1 Tableaux des données 50

5.1 Matrice des données à classifier 56

Table des figures

1.1 Représentation graphique des données centrées réduites 12

3.1 Représentation Graphique sur les deux axes factoriels 49

4.1 Représentation graphique de la décomposition de Huygens 52
4.2 Cas de deux groupes 54

5.1 Production de Crevettes en tonnes 62
5.2 Dendogramme selon la méthode du lien minimum 64
5.3 Dendogramme selon la méthode de Ward 65

Introduction

L'analyse de données, qu'est-ce que c'est ?

L'analyse de données est un ensemble de méthodes statistiques appliquées à un jeu de données dans le but d'extraire des informations pertinentes ; on appelle cette extraction fouille de données. Le but est de dégager des tendances, des profils, de détecter des comportements ou de trouver des liens, des règles. Il existe deux grands types d'analyse de données : analyse descriptive et analyse prédictive.

L'analyse descriptive a pour but de résumer les données en leur assignant une nouvelle représentation, de synthétiser en faisant ressortir ce qui est dissimulé par le volume. On peut classer les individus dans des catégories, trouver les individus les plus proches ou les plus éloignés entre eux ;mais aussi trouver les exceptions ou les cas atypiques. On peut également voir si des variables sont proches,expliquer une variable en fonction des autres ou encore repérer les variables les plus influentes.

L'analyse prédictive consiste à analyser les données actuelles afin de faire des hypothèses sur des comportements futurs. On se sert des données que l'on possède déjà pour extrapoler et deviner le comportement de nouveaux individus mais également l'évolution des individus déjà présents.

On peut distinguer deux " familles " de méthodes :analyses factorielles et classification.

Les analyses factorielles consistent à transformer le tableau de données initial en un nouveau tableau contenant la même information, mais sous forme hiérarchisée. Il est composé d'axes factoriels. Le premier axe factoriel correspond à la combinaison linéaire des variables initiales qui différencie au maximum les individus entre eux. Il est de variance maximum. Les axes factoriels sont indépendants les uns des autres et classés en fonction de leur variance. En général il suffit d'un petit nombre d'axes factoriels (trois ou quatre) pour rendre compte de l'essentiel de l'information contenue dans le tableau initial. L'interprétation de ces axes factoriels permet de mettre en évidence la forme des interrelations entre les variables étudiées et les ressemblances et dissemblances entre les individus relativement à ces variables. Les deux méthodes les plus communément utilisées sont l'analyse en composantes principales (adaptée pour des données hétérogènes combinant des variables exprimées dans des échelles de mesure différentes, ou encore pour des variables exprimées en pourcentages) et l'analyse des correspondances (adaptée pour des tableaux de contingence ou de **variables qualitatives**).

Les classifications permettent d'élaborer des typologies et de regrouper les individus par classes en fonction de leurs ressemblances par rapport à l'ensemble des variables. Un critère souvent utilisé du point de vue technique est de chercher la classification qui minimise la variance intraclasse (variabilité entre les individus d'une même classe) et maximise la variance interclasse (la variabilité entre les classes). Les méthodes les plus classiques sont la classification ascendante hiérarchique et la classification par nuées dynamiques.

Chapitre 1

Analyse en Composantes Principales(ACP)

1.1 Introduction

L'ACP est une des plus anciennes méthodes factorielles. Elle a été conçue par Karl Peason (1901) et intégrée à la statistique par Harold Hotelling (1933). Elle est utilisée lorsqu'on observe sur n individus, p variables quantitatives $X^1, X^2, ..., X^p$ présentant des liaisons multiples que l'on veut analyser. Ces observations sont regroupées dans un tableau (matrice) rectangulaire X ayant n lignes (individus) et p colonnes (variables) :

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^p \\ x_2^1 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^p \end{pmatrix}$$

TABLE 1.1 – Représentation matricielle des tableaux de données

Un individu est représenté par $e_i' = (x_i^1, x_i^2, \cdots, x_i^p)$.
Pour avoir une image de l'ensemble des individus, on se place dans un espace affine de dimension p en choisissant comme origine un point particulier de $\mathbb{R}^p$, par exemple le vecteur dont toutes les coordonnées sont nulles. Alors, chaque individu sera représentée par un point dans cet espace. L'ensemble des points qui représentent les individus est appelé "nuage des individus".
De même dans $\mathbb{R}^n$, chaque variable pourra être représentée par un point de l'espace affine correspondant. L'ensemble des points qui représentent les variables est appelé "nuage des variables".

On constate que, ces espaces étant de dimension supérieure en général à 2, on ne peut visualiser ces représentations.

1.2 Qu'est-ce que l'ACP ?

L'Analyse en Composantes Principales peut être considérée comme une méthode de projection qui permet de projeter les observations depuis l'espace à p dimensions des p variables

vers un espace à k dimensions ($k < p$) tel qu'un maximum d'informations soit conservé (l'information est ici mesurée au travers de la variance totale du nuage de points) sur les premières dimensions. Si l'information associée aux 2 ou 3 premiers axes représente un pourcentage suffisant de la variabilité totale du nuage de points, on pourra représenter les observations sur un graphique à 2 ou 3 dimensions, facilitant ainsi grandement l'interprétation.
Autrement dit, on cherche à définir k nouvelles variables combinaisons linéaires des p variables initiales qui feront perdre le moins d'information possible.

Remarque 1 *L'ACP sur un tableau de données tel que $p > n$ est impossible.*

1.3 Domaines d'application

L'Analyse en Composantes Principales (ACP) est l'une des méthodes d'analyse de données multivariées les plus utilisées. Elle permet d'explorer des jeux de données multidimensionnels constitués de variables quantitatives. Elle est largement utilisée en biostatistique, marketing, sciences sociales et bien d'autres domaines.

Définition 1 *On associe aux individus un poids p_i tel que $p_1 + \cdots + p_n = 1$ que l'on représente par la matrice diagonale de taille n*

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

TABLE 1.2 – Matrice des poids des individus

Remarque 2 *Le cas le plus fréquent est de considérer que tous les individus ont la même importance : $p_i = 1/n$.*

Point moyen c'est le vecteur g des moyennes arithmétiques de chaque variable :

$$g' = \left(\bar{x}^1, \bar{x}^2, \cdots, \bar{x}^p\right) = \sum_{i=1}^{n} p_i e_i$$

On peut écrire sous forme matricielle

$$g = X' D_p \mathbf{1}_n$$

Tableau centré il est obtenu en centrant les variables autour de leur moyenne $y_i^j = x_i^j - \bar{x}^j$, c'est-à-dire y^j=x$^j - \bar{x}^j \mathbf{1}_n$ ou,en notation matricielle,

$$Y = X - \mathbf{1}_n g'$$

Proposition 1

$$Y = \left(I_n - \mathbf{1}_n \mathbf{1}_n' D_p\right) X$$

Preuve : *Rappelons que $g = X' D_p \mathbf{1}_n$.*
Autrement dit,

$$g' = \left(X' D_p \mathbf{1}_n\right)' = \left(D_p \mathbf{1}_n\right)' \left(X'\right)' = \mathbf{1}_n' D_p' X$$

or D_p est une matrice diagonale ($D_p' = D_p$).
En effet, $g' = \mathbf{1}_n' D_p X$
Par conséquent,

$$Y = X - \mathbf{1}_n \mathbf{1}_n' D_p X = (I_n - \mathbf{1}_n \mathbf{1}_n' D_p)\, X$$

Matrice de variance-covariance c'est une matrice carrée de dimension p

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

TABLE 1.3 – Matrice de Variance-covariance

où σ_{kl} est la covariance des variables x^k et x^l et σ_j^2 est la variance de la variable x^j.

Proposition 2

$$V = X' D_p X - gg' = Y' D_p Y$$

Preuve : D'une part,

$$Y = \begin{pmatrix} x_1^1 - \bar{x}^1 & x_1^2 - \bar{x}^2 & \cdots & x_1^p - \bar{x}^p \\ x_2^1 - \bar{x}^1 & x_2^2 - \bar{x}^2 & \cdots & x_2^p - \bar{x}^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 - \bar{x}^1 & x_n^2 - \bar{x}^2 & \cdots & x_n^p - \bar{x}^p \end{pmatrix}$$

alors,

$$Y' = \begin{pmatrix} x_1^1 - \bar{x}^1 & x_2^1 - \bar{x}^1 & \cdots & x_n^1 - \bar{x}^1 \\ x_1^2 - \bar{x}^2 & x_2^2 - \bar{x}^2 & \cdots & x_n^2 - \bar{x}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^p - \bar{x}^p & x_2^p - \bar{x}^p & \cdots & x_n^p - \bar{x}^p \end{pmatrix}$$

Par suite, calculons $Y' D_p = \begin{pmatrix} x_1^1 - \bar{x}^1 & x_2^1 - \bar{x}^1 & \cdots & x_n^1 - \bar{x}^1 \\ x_1^2 - \bar{x}^2 & x_2^2 - \bar{x}^2 & \cdots & x_n^2 - \bar{x}^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^p - \bar{x}^p & x_2^p - \bar{x}^p & \cdots & x_n^p - \bar{x}^p \end{pmatrix} \begin{pmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{pmatrix}$

$$= \begin{pmatrix} p_1(x_1^1 - \bar{x}^1) & p_2(x_2^1 - \bar{x}^1) & \cdots & p_n(x_n^1 - \bar{x}^1) \\ p_1(x_1^2 - \bar{x}^2) & p_2(x_2^2 - \bar{x}^2) & \cdots & p_n(x_n^2 - \bar{x}^2) \\ \vdots & \vdots & \ddots & \vdots \\ p_1(x_1^p - \bar{x}^p) & p_2(x_2^p - \bar{x}^p) & \cdots & p_n(x_n^p - \bar{x}^p) \end{pmatrix}$$

D'autre part,

$$Y' D_p Y = \begin{pmatrix} p_1(x_1^1 - \bar{x}^1) & \cdots & p_n(x_n^1 - \bar{x}^1) \\ p_1(x_1^2 - \bar{x}^2) & \cdots & p_n(x_n^2 - \bar{x}^2) \\ \vdots & \ddots & \vdots \\ p_1(x_1^p - \bar{x}^p) & \cdots & p_n(x_n^p - \bar{x}^p) \end{pmatrix} \begin{pmatrix} x_1^1 - \bar{x}^1 & x_1^2 - \bar{x}^2 & \cdots & x_1^p - \bar{x}^p \\ x_2^1 - \bar{x}^1 & x_2^2 - \bar{x}^2 & \cdots & x_2^p - \bar{x}^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 - \bar{x}^1 & x_n^2 - \bar{x}^2 & \cdots & x_n^p - \bar{x}^p \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n p_i(x_i^1 - \bar{x}^1)(x_i^1 - \bar{x}^1) & \cdots & \sum_{i=1}^n p_i(x_i^1 - \bar{x}^1)(x_i^p - \bar{x}^p) \\ \sum_{i=1}^n p_i(x_i^2 - \bar{x}^2)(x_i^1 - \bar{x}^1) & \cdots & \sum_{i=1}^n p_i(x_i^2 - \bar{x}^2)(x_i^p - \bar{x}^p) \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n p_i(x_i^p - \bar{x}^p)(x_i^1 - \bar{x}^1) & \cdots & \sum_{i=1}^n p_i(x_i^p - \bar{x}^p)(x_i^p - \bar{x}^p) \end{pmatrix}$$

Donc, $Y'D_pY = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$

D'où,

$$Y'D_pY = V$$

Matrice de corrélation Si l'on note $r_{kl} = \frac{\sigma_{kl}}{\sigma_k \sigma_l}$, c'est la matrice carrée d'ordre p

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{pmatrix}$$

Proposition 3

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

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

Preuve : On a

$$V = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

alors,

$$D_{1/\sigma}V = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_p \end{pmatrix} \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1 & \frac{\sigma_{12}}{\sigma_2} & \cdots & \frac{\sigma_{1p}}{\sigma_p} \\ \frac{\sigma_{21}}{\sigma_1} & \sigma_2 & \cdots & \frac{\sigma_{2p}}{\sigma_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_1} & \frac{\sigma_{p2}}{\sigma_2} & \cdots & \sigma_p \end{pmatrix}$$

Ensuite, $D_{1/\sigma}VD_{1/\sigma} = \begin{pmatrix} \sigma_1 & \frac{\sigma_{12}}{\sigma_2} & \cdots & \frac{\sigma_{1p}}{\sigma_p} \\ \frac{\sigma_{21}}{\sigma_1} & \sigma_2 & \cdots & \frac{\sigma_{2p}}{\sigma_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{p1}}{\sigma_1} & \frac{\sigma_{p2}}{\sigma_2} & \cdots & \sigma_p \end{pmatrix} \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_p \end{pmatrix}$

D'oú, $D_{1/\sigma}VD_{1/\sigma} = R$

Les données centrées réduites forment un tableau Z contenant les données $z_i^j = \frac{y_i^j}{\sigma_j}$, c'est-à-dire $\mathrm{z}^j = \frac{Y^j}{\sigma_j}$ [1]

1. Pour que les distances soient indépendantes des unités de mesure et ne pas privilégier les variables dispersées.

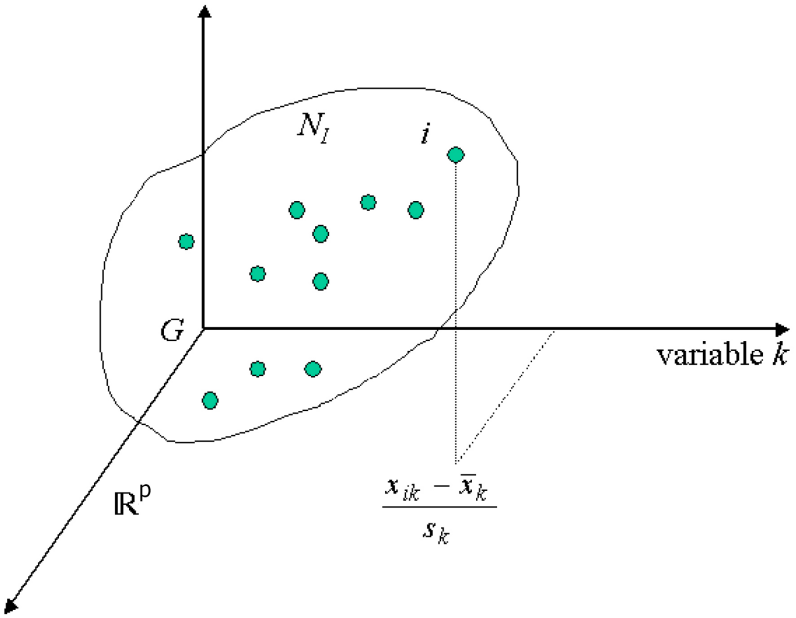


FIGURE 1.1 – Représentation graphique des données centrées réduites

1.4 Inertie d'un nuage de points

Soit $\mathcal{M} = (x_i, p_i)$ le nuage de points. On note $\mathcal{N} = (y_i, p_i)$ le nuage centré où on a ramené le centre de gravité à l'origine du repère.

1.4.1 Distance entre deux individus (ou variables)

Généralisation simple On donne un poids $m_i > 0$ à la variable i

$$d^2(u, v) = \sum_{i=1}^{p} m_i (u_i - v_i)^2$$

Cela revient à multiplier la coordonnée i par $\sqrt{m_i}$
Plus précisement, on munit l'espace $\mathbb{R}^p$ d'une métrique M (matrice $p \times p$ symétrique définie positive).

- Un produit scalaire : $< x, y >_M = {}^t x M y$
- Une norme : $\|x\|_M = \sqrt{< x, x >_M}$
- Une distance : $d_M(x, y) = \|x - y\|_M$

Remarque 3 *En pratique, on utilise le plus souvent l'une des métriques suivantes :*

— *$M = I_d$, la distance est la distance euclidienne usuelle, et on parle d'ACP canonique ou simple. Elle s'utilise lorsque les variables sont homogènes (même dimension) et de même ordre de grandeur.*

— *$M = D_{1/S^2}$ où D_{1/S^2} est la matrice diagonale des inverses des variances $D_{1/S^2} = D_{1/S} D_{1/S}$. Le choix de cette métrique revient à diviser chaque variable (colonne) par son écart-type. On parle alors d'ACP normée. Ici la distance ne dépend plus des unités de mesure puisque x_i^j / s_j est une grandeur sans dimension. Cette métrique donne à chaque caractère la même importance quelle que soit sa dispersion. Elle s'utilise lorsque les variables ne sont pas homogènes, ou ne sont pas de même ordre de grandeur.*

1.4.2 Inertie totale du nuage de points

L'information pour un nuage de points est définie par la notion d'inertie.

Définition 2 (Inertie) *On appelle inertie d'un nuage de point $X = (x_i^j)$ la quantité :*

$$I = \sum_{i=1}^{n} p_i d_M^2(G, M_i) = \sum_{i=1}^{n} p_i \|e_i - g\|_M^2 = \sum_{i=1}^{n} p_i \|y_i\|_M^2 \tag{1.1}$$

où $\sum_{i=1}^{n} p_i = 1$ et G est le centre de gravité du nuage de points.
L'inertie du nuage $\mathcal{M}$ est évidemment égale à l'inertie du nuage centré $\mathcal{N}$. Dans la suite du chapitre, on supposera que le nuage est centré.

Définition 3 *L'inertie en un point a du nuage de points est*

$$I_a = \sum_{i=1}^{n} p_i d_M^2(a, M_i) = \sum_{i=1}^{n} p_i \|e_i - a\|_M^2 = \sum_{i=1}^{n} p_i (e_i - a)' M (e_i - a) \tag{1.2}$$

Proposition 4

$$I = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - e_j\|_M^2 \tag{1.3}$$

Preuve : *Notons $K = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - e_j\|_M^2$*
On a $\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - e_j\|_M^2 = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - g + g - e_j\|_M^2$
Par suite,
$K = \frac{1}{2}(\sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - g\|_M^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|g - e_j\|_M^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j < e_i - g, g - e_j >_M)$
En effet,
$K = \frac{1}{2}(\sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - g\|_M^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|g - e_j\|_M^2 + 2 < \sum_{i=1}^{n} p_i (e_i - g), \sum_{j=1}^{n} p_j (g - e_j) >_M)$
D"une part, G est le centre de gravité alors $\sum_{i=1}^{n} p_i (e_i - g) = 0$
Par conséquent,

$$K = \frac{1}{2}(\sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - g\|_M^2 + \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|g - e_j\|_M^2)$$

Donc,

$$K = \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - g\|_M^2$$

Autrement dit,

$$K = \sum_{i=1}^{n} p_i \|e_i - g\|_M^2 \underbrace{\sum_{j=1}^{n} p_j}_{=1}$$

D'où,

$$K = \sum_{i=1}^{n} p_i \|e_i - g\|_M^2$$

On conclut que

$$I = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j \|e_i - e_j\|_M^2$$

1.4.3 Inertie expliquée par un sous espace F.

Définition 4 *On appelle inertie du nuage des individus $\mathcal{N}$ expliquée ou portée par le sous-espace F de $\mathbb{R}^p$, l'inertie du nuage projeté sur F, c'est-à-dire :*

$$I_F(\mathcal{N}) = \sum_{i=1}^{n} p_i d_M^2(O, \hat{y}_i^F) = \sum_{i=1}^{n} p_i \|\hat{y}_i^F\|_M^2 \tag{1.4}$$

où $\hat{y}_i^F$ est la projection orthogonale de y_i sur F.

1.4.4 Décomposition de l'inertie totale

Si on note $F^\perp$ le complémentaire orthogonal de F dans $\mathbb{R}^p$, tout élément de $\mathbb{R}^p$ se décompose de manière unique sous la forme $y = \hat{y}^F + \hat{y}^{F^\perp}$ tel que $< \hat{y}^F, \hat{y}^{F^\perp} >_M = 0$

$$I = \sum_{i=1}^{n} p_i \|y_i\|_M^2 = \sum_{i=1}^{n} p_i \|\hat{y}_i^F + \hat{y}_i^{F^\perp}\|_M^2$$

Par suite,

$$I = \sum_{i=1}^{n} p_i \|\hat{y}_i^F\|_M^2 + \sum_{i=1}^{n} p_i \|\hat{y}_i^{F^\perp}\|_M^2 + 2\sum_{i=1}^{n} p_i < \hat{y}_i^F, \hat{y}_i^{F^\perp} >_M$$

On en déduit que :

$$I = I_F + I_{F^\perp} \text{ (Théorème d'Huygeens)} \tag{1.5}$$

La quantité $I_{F^\perp}$ peut donc être considérée comme une mesure de la déformation du nuage lors de la projection sur F :

$$I_{F^\perp} = \sum_{i=1}^{n} p_i \|y_i - \hat{y}_i^F\|_M^2 \tag{1.6}$$

Remarque 4 *L'inertie totale se décompose pour tout F sous espace de $\mathbb{R}^p$ comme la somme de l'inertie totale du nuage projeté sur F et la déformation du nuage $\mathcal{N}$ par projection orthogonale sur F.*

Proposition 5

$$I = tr(MV) = tr(VM) \tag{1.7}$$

Preuve : D'une part,

$$I = \sum_{i=1}^{n} p_i \|y_i\|_M^2 \text{ or } \|y\|_M^2 = < y, y >_M = {}^t y M y$$

Par suite,

$$I = \sum_{i=1}^{n} p_i < y_i, y_i >_M = \sum_{i=1}^{n} p_i tr(< y_i, y_i >_M)$$

car $< y_i, y_i >_M$ est un scalaire.

D'autre part,

$$I = \sum_{i=1}^{n} p_i tr({}^t y_i M y_i)$$

Comme $tr(ABC) = tr(CAB) = tr(BCA)$

On a,

$$I = \sum_{i=1}^{n} p_i tr(y_i {}^t y_i M)$$

En effet,

$$I = tr\left(\underbrace{(\sum_{i=1}^{n} p_i y_i {}^t y_i)}_{V} M \right)$$

Par conséquent,

$$I = tr(VM) = tr(MV)$$

car $tr(AB) = tr(BA)$

Remarque 5 *i. Si* $M = I_d$,

$$I = Tr(V)$$

Autrement dit, $I = \sum_{j=1}^{p} s_j^2$

ii. Si $M = D_{1/S^2}$, $I = p$. *Dans ce cas, l'inertie est égale au nombre de variables et ne dépend pas de leurs valeurs.*

iii. La terminologie vient de la mécanique. Si on considère un système de n points matériels de masse ponctuelle $\frac{1}{n}$ *dans* $\mathbb{R}^p$ *de coordonnées* ${}^t x_{i.}$ *alors la quantité I est l'inertie au sens mécanique du terme.*

1.5 Recherche du maximum

Rappelons que l'objectif principal est d'obtenir une représentation fidèle du nuage des individus de $\mathbb{R}^p$(resp. des variables dans $\mathbb{R}^n$) en le projetant sur un espace de faible dimension. Le choix de l'espace de projection s'effectue selon le critère de l'inertie . C'est-à-dire, on cherche le sous-espace de dimension k portant l'inertie maximale du nuage. Rappelons que si u est un vecteur M-normé ($\|u\|_M = 1$) , et Δ_u est la droite vectorielle engendrée par u, la projection orthogonale de y_i sur Δ_u est $\hat{y}_i^u =< y_i, u >_M = {}^t y_i M u$ et l'inertie expliquée par Δ_u est donnée par

$$I_{\Delta_u} = \sum_{i=1}^{n} p_i \|\hat{y}_i^u\|_M^2 = \sum_{i=1}^{n} p_i < y_i, u >_M^2 = \sum_{i=1}^{n} p_i ({}^t y_i M u)^2$$

$$= \sum_{i=1}^{n} p_i ({}^t u M y_i {}^t y_i M u) = {}^t u M \underbrace{\sum_{i=1}^{n} p_i (y_i {}^t y_i)}_{V} M u$$

En effet,

$$I_{\Delta_u} = {}^t u M V M u \tag{1.8}$$

De plus, si on décompose l'espace $\mathbb{R}^p$ comme la somme de sous-espaces de dimension 1 et orthogonaux entre eux :

$$\Delta_1 \perp \Delta_2 \perp \cdots \perp \Delta_p$$

alors

$$I = I_{\Delta_1} + I_{\Delta_2} + \cdots + I_{\Delta_p} \tag{1.9}$$

Définition 5 *Les axes* $\Delta_{u_1}, \Delta_{u_2}, \cdots, \Delta_{u_p}$ *sont appelés axes principaux d'inertie de l'ACP.*

Le problème à résoudre : trouver u_1 tel que $I_{\Delta_{u_1}}$ soit maximum avec la contrainte $\|u_1\|_M^2 = 1$ est le problème de la recherche d'un optimum d'une fonction de plusieurs variables liées par une contrainte (les inconnues sont les composantes de u_1). Pour chercher les optimums d'une fonction $f(t_1, t_2 \cdots, t_p)$ de p variables liées par une relation $l(t_1, t_2, ..., t_p) = cte$, on applique la Méthode des multiplicateurs de Lagrange.
Calculons les dérivées partielles de la fonction

$$f(t_1, t_2 \cdots, t_p) - \lambda(l(t_1, t_2, ..., t_p) - cte)$$

Autrement dit, dans le cas de la recherche de u_1, il faut calculer les dérivées partielles de :

$$g(u_1) = g(u_{11}, u_{12}, \cdots, u_{1p}) = {}^{\mathrm{t}}u_1 MVMu_1 - \lambda_1({}^{\mathrm{t}}u_1 Mu_1 - 1)$$

En utilisant la dérivée matricielle, on obtient :

$$\frac{\partial g(u_1)}{\partial u_1} = \frac{\partial {}^{\mathrm{t}}u_1}{\partial u_1} MVMu_1 + {}^{\mathrm{t}}u_1 MVM\frac{\partial u_1}{\partial u_1} - \lambda_1(\frac{\partial {}^{\mathrm{t}}u_1}{\partial u_1} Mu_1 + {}^{\mathrm{t}}u_1 M\frac{\partial u_1}{\partial u_1})$$

Remarquant que, $\frac{\partial {}^{\mathrm{t}}u_1}{\partial u_1} = \begin{bmatrix} \frac{\partial {}^{\mathrm{t}}u_1}{\partial u_{11}} \\ \vdots \\ \frac{\partial {}^{\mathrm{t}}u_1}{\partial u_{1j}} \\ \vdots \\ \frac{\partial {}^{\mathrm{t}}u_1}{\partial u_{1p}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & 1 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{bmatrix} = I_p$

On peut remarquer aussi que les éléments des deux vecteurs ${}^{\mathrm{t}}u_1$ et u_1 sont égaux ligne à ligne, puisque chacun est la transposée de l'autre et qu'ils sont de dimension 1×1 .
En effet,

$$\frac{\partial g(u_1)}{\partial u_1} = 2MVMu_1 - 2\lambda_1 Mu_1$$

et

$$\frac{\partial g(u_1)}{\partial \lambda_1} = {}^{\mathrm{t}}u_1 Mu_1 - 1$$

Le système à résoudre est :

$$\text{(Condition nécessaire du 1}^{\text{er}}\text{ ordre)} \begin{cases} 2MVMu_1 - 2\lambda_1 Mu_1 = 0 & (1) \\ {}^{\mathrm{t}}u_1 Mu_1 - 1 = 0 & (2) \end{cases}$$

Comme M est une matrice inversible
alors l'équation matricielle (1) donne

$$VMu_1 = \lambda_1 u_1 \qquad (3)$$

on déduit que u_1 est vecteur propre de la matrice VM associé à la valeur propre λ_1. De plus, partant de (1) , $MVMu_1 = \lambda_1 Mu_1$. En multipliant à gauche par ${}^{\mathrm{t}}u_1$ les deux membres de l'équation.

On obtient :

$$\underbrace{{}^{\mathrm{t}}u_1 MVMu_1}_{I_{\Delta_{u_1}}} = \lambda_1 \underbrace{{}^{\mathrm{t}}u_1 Mu_1}_{=1(2)}$$

Cela signifie que la valeur propre λ_1 est la plus grande valeur propre de la matrice VM.

Corollaire 1 *Le sous-espace E_k de dimension k portant l'inertie maximale est engendré par les k vecteurs propres de VM associés aux k plus grandes valeurs propres.*

1.5.1 Recherche des axes suivants

Théorème 1 *1. Il existe une base M-orthonormée $(u_1, u_2, \cdots, u_p)$ de vecteurs propres de la matrice VM associés aux valeurs propres (réelles positives) rangées par ordre décroissant $\lambda_1 \geqslant \cdots \geqslant \lambda_p \geqslant 0$.*

2. Les vecteurs $u_1, u_2, ..., u_k$ engendrent respectivement les axes principaux d'inertie de l'ACP et on a pour tout $j \in \{1, \cdots, p\}$,

$$I_{\Delta_{u_j}} = \lambda_j$$

3. Pour tout $k < p$, le sous espace vectoriel E_k engendré par les k premiers vecteurs $u_1, u_2, \cdots, u_k$, est un sous espace vectoriel principal de dimension k, et l'inertie expliquée par E_k est donnée par $I_{E_k} = \lambda_1 + \cdots + \lambda_k$.

Preuve :

1. Rappelons que M et V sont symétriques.
   Par suite,$< x, VMy >_M = {}^t xMVMy =< VMy, x >_M = {}^t yMVMx$.
   Cela revient à dire que VM est M-symétrique.
   De plus, $< VMx, y >_M = {}^t(VMx)My = {}^t x {}^t M {}^t VMy =< x, VMy >_M$. Par conséquent, les valeurs propres de VM sont réelles et positives, et VM admet une base M-orthonormée de vecteurs propres.
2. Rappelons que le premier axe principal d'inertie est engendré par le vecteur propre u_1 associé à λ_1 la plus grande valeur propre de VM. Pour les autres axes, on admet l'hypothèse que

   $$\lambda_k = \max\{< u, VMu >_M ; \|u\|_M = 1, < u, u_i >_M = 0, i \in \{1, 2, \cdots, k-1\}\}$$

   Ensuite $I_{\Delta_{u_k}} = {}^t u_k MVMu_k = {}^t u_k M(VMu_k)$
   Comme u_k est un vecteur propre de VM,
   alors, $I_{\Delta_{u_k}} = {}^t u_k M(\lambda_k u_k)$
   Autrement dit, $I_{\Delta_{u_k}} = \lambda_k {}^t u_k Mu_k$ or $\|u_k\|_M = 1$
   Par conséquent, $I_{\Delta_{u_k}} = \lambda_k$
3. D'après la **corollaire 1** E_k se décompose de manière unique

   $$E_k = \Delta_{u_1} \oplus \Delta_{u_2} \oplus \cdots \oplus \Delta_{u_k}$$

   D'où,

   $$I_{E_k} = \lambda_1 + \cdots + \lambda_k$$

1.5.2 Les composantes principales

Coordonnées des individus

Supposons que

$$e_i - g = \sum_{j=1}^{p} c_i^j u_j$$

Alors , $< e_i - g, u_k >_M = \sum_{j=1}^{p} c_i^j < u_j, u_k >_M = c_i^k$.
La coordonnée de l'individu centré $e_i - g$ sur l'axe principal u_k est donc donné par la projection M-orthogonale $c_i^k =< y_i, u_k >_M = {}^t(e_i - g)Mu_k$

Définition 6 *Le vecteur de* $\mathbb{R}^n$

$$c^j = \begin{pmatrix} c_1^j \\ c_2^j \\ \vdots \\ c_n^j \end{pmatrix} = YMu_j \text{ est appelé } j\text{-ème composante principale.}$$

Chaque c^j *contient les coordonnées des projections* M*-orthogonales des individus centrés sur l'axe défini par les* u_j*.*

Proposition 6 1. *la variance de* c^j *est* λ_j*.*

2. *Pour tout* $j \neq l$*,*
$$cov(c^j, c^l) = 0$$
3. *Les Composantes principales sont des combinaisons linéaires des variables de départ* y_i*.*
4. *Les Composantes principales* $c^1, \cdots, c^p$ *sont des vecteurs propres de la matrice* $YMY'D_p$*, de valeurs propres* $\lambda_1, \lambda_2, \cdots, \lambda_p$*.*

Preuve :

1. Par définition ,$Var(c^j) = {}^t c^j D_p c^j$ or $c^j = YMu_j$
   En effet,
   $$Var(c^j) = {}^t(YMu_j) D_p YMu_j$$
   Par suite,
   $$Var(c^j) = {}^t u_j M {}^t Y D_p Y M u_j$$
   Comme ${}^t Y D_p Y = V$
   Par conséquent,
   $$Var(c^j) = {}^t u_j M V M u_j$$
   D'une part, u_j est un vecteur propre de VM
   Il s'ensuit que
   $$Var(c^j) = {}^t u_j M(\lambda_j u_j) = \lambda_j {}^t u_j M u_j$$
   D'autre part, $\|u_j\|_M = 1$
   D'où
   $$Var(c^j) = \lambda_j$$
2. Sachant que $cov(c^j, c^l) = {}^t c^j D_p c^l = {}^t u_j M V M u_l = {}^t u_j M(\lambda_l u_l) = \lambda_l \underbrace{({}^t u_j M u_l)}_{<u_j, u_l>_M}$
   or si $j \neq l, < u_j, u_l >_M = 0$
   D'où,
   $$cov(c^j, c^l) = 0$$
   On conclut que les composantes principales ne sont pas corrélées entre elles.
3. Par définition, $c^j = YMu_j$
   D'une part, $Y = \begin{bmatrix} y^1 & y^2 & \cdots & y^p \end{bmatrix}$
   et $M = \begin{pmatrix} m_{11} & m_{12} & \cdot & m_{1p} \\ m_{12} & m_{22} & \cdots & m_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ m_{1p} & m_{2p} & \cdots & m_{pp} \end{pmatrix}, u_j = \begin{pmatrix} u_{j1} \\ u_{j2} \\ \vdots \\ u_{jp} \end{pmatrix}$

Par suite,

$$c^j = \begin{bmatrix} y^1 & y^2 & \cdots & y^p \end{bmatrix} \begin{pmatrix} m_{11} & m_{12} & \cdot & m_{1p} \\ m_{12} & m_{22} & \cdots & m_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ m_{1p} & m_{2p} & \cdots & m_{pp} \end{pmatrix} \begin{pmatrix} u_{j1} \\ u_{j2} \\ \vdots \\ u_{jp} \end{pmatrix}$$

En effet,

$$c^j = \begin{bmatrix} y^1 & y^2 & \cdots & y^p \end{bmatrix} \begin{pmatrix} \sum_{k=1}^{p} m_{1k}u_{jk} \\ \sum_{k=1}^{p} m_{2k}u_{jk} \\ \vdots \\ \sum_{k=1}^{p} m_{pk}u_{jk} \end{pmatrix}$$

Donc,

$$c^j = \sum_{l=1}^{p}(\sum_{k=1}^{p} m_{lk}u_{jk})y^l$$

4. Sachant que u_j est un vecteur propre de VM associé à la valeur propre λ_j .
   Autrement dit, $VMu_j = \lambda_j u_j$
   D'autre part, $V = Y'D_pY$
   Par suite,

$$Y'D_p\underbrace{YMu_j}_{c^j} = \lambda_j u_j$$

   En multipliant membre à membre par YM à gauche

$$YMY'D_pc^j = \lambda_j \underbrace{YMu_j}_{c^j}$$

   Par conséquent,

$$YMY'D_pc^j = \lambda_j c^j$$

1.6 ACP dans l'espace des variables

On peut envisager le problème de la représentation des variables de façon complètement symétrique de celui des individus. Les raisonnements se font dans $\mathbb{R}^n$ au lieu de $\mathbb{R}^p$.
Chaque variable X^j peut alors être représentée par un vecteur de $\mathbb{R}^n$ appelé espace vectoriel des variables.

$$x^j = \begin{pmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{pmatrix} = j\text{-ème colonne de } X$$

Pour construire cette ACP, on a besoin de définir :

— Le tableau de données : Il s'agit du tableau (p, n) obtenu en mettant les vecteurs $y^1, y^2, \cdots, y^p$ sous forme de vecteurs lignes, et en mettant ces lignes l'une en dessous de l'autre. Il est clair que le tableau obtenu est Y'.
— Une métrique sur l'espace des variables $\mathbb{R}^n$: on a déjà vu qu'un choix naturel est de prendre D_p.
— La matrice du poids : on va ici choisir la matrice M.

1.6.1 Facteurs principaux

Définition 7 *On associe à u_k le facteur principal $v_k = \begin{pmatrix} v_{1k} \\ v_{2k} \\ \vdots \\ v_{pk} \end{pmatrix} = Mu_k$ de taille p.*

C'est un vecteur propre de MV car $MVv_k = MVMu_k = M(\lambda_k u_k) = \lambda_k v_k$.

Remarque 6 *En pratique, on calcule les v_k par diagonalisation de MV, puis on obtient les Composantes principales.*
On voit que la matrice des v_{jk} sert de matrice de passage entre la nouvelle base et l'ancienne.

$$c_i^k = \sum_{l=1}^{p} y_i^l v_{lk}, \ \boldsymbol{c}_k = \sum_{l=1}^{p} y^l v_{lk}, \ \boldsymbol{c}_k = Yv_k$$

Réciproquement, les u_{kj} forment une matrice de passage entre l'ancienne base et la nouvelle.

$$y_i^j = \sum_{k=1}^{p} c_i^k u_{kj}, \ \boldsymbol{y}^j = \sum_{k=1}^{p} \boldsymbol{c}_k u_{kj}, \ Y = \sum_{k=1}^{p} \boldsymbol{c}_k {}^t u_k$$

1.7 Les représentations graphiques.

1.7.1 Résultats relatifs aux individus

Définition 8 *Pour tout $i, k \leq p$, la projection du nuage $\mathcal{N}$ sur le plan principal $(\Delta_{u_i}, \Delta_{u_k})$ est appelé carte des individus.*

Rappelons que l'inertie totale du nuage $\mathcal{N}$ des individus vaut $I = \sum_{i=1}^{n} p_i \|y_i\|_M^2 = tr(VM) = \sum_{i=1}^{p} \lambda_i$.

Définition 9 *On sait que $var(\boldsymbol{c}_k) = \lambda_k = \sum_{i=1}^{n} p_i (c_i^k)^2$. La contribution de l'individu i à la composante k est donc*

$$\frac{I_k(i)}{I_k}$$

où $I_k(i)$ est la part d'inertie portée par Δ_{u_k} et I_k l'inertie totale suivant l'axe Δ_{u_k}.
Rappelons que $I_k = I_{\Delta_{u_k}} = \lambda_k$ et $I_k(i) = p_i (c_i^k)^2$
On en déduit que la contribution de l'individu i au k-ième axe principal est

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

Généralement, la contribution de l'individu i à l'inertie du nuage des individus est donc le rapport

$$\frac{\sum_{k=1}^{p} p_i (c_i^k)^2}{\sum_{k=1}^{p} \lambda_k} = \frac{p_i \|y_i\|_M^2}{I}$$

Définition 10 *La qualité globale de la représentation du nuage $\mathcal{N}$ sur le sous espace principal E_k engendré par $(u_1, \cdots, u_k)$ est mesurée par le pourcentage d'inertie expliquée par E_k*

$$\frac{I_{E_k}}{I} = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\sum_{i=1}^{p} \lambda_i}$$ [2]

2. Plus cette qualité est proche de 1, plus le nuage de points initial est "concentré" autour de E_k.

Définition 11 ***Qualité locale de la représentation*** *de l'individu i sur un axe principal. Elle est mesurée par le cosinus carré de l'angle que fait y_i et l'axe principal k.*

$$\cos(\widehat{y_i, u_k}) = \frac{< y_i, u_k >_M}{\|y_i\|_M \underbrace{\|u_k\|_M}_{1}} = \frac{< y_i, u_k >_M}{\|y_i\|_M}$$

or $< y_i, u_k >_M = < e_i - g, u_k >_M = c_i^k$
En effet,

$$\cos^2(\widehat{y_i, u_k}) = \frac{(c_i^k)^2}{\sum_{k=1}^{p}(c_i^k)^2}$$

On conclut que, la qualité de représentation de l'individu i sur l'espace principal E_q est

$$\cos^2(\widehat{y_i, E_q}) = \frac{\sum_{k=1}^{q}(c_i^k)^2}{\sum_{k=1}^{p}(c_i^k)^2}$$

Remarque 7 *On peut considérer que plus $p_i(c_i^k)^2$ est grand, plus l'individu i est important sur l'axe k.*

1.7.2 Résultats relatifs aux variables

Définition 12 *Pour tout $k, l \leq r$ (correspondant aux valeurs propres non nulles), la projection du nuage $\mathcal{V}$ sur le plan principal $(\Delta_{v_k}, \Delta_{v_l})$ est appelé* ***carte des variables****.*

Qualité de la représentation du nuage des variables :

L'inertie totale du nuage vaut

$$I(\mathcal{V}) = I(\mathcal{N}) = I = \sum_{i=1}^{r} \lambda_i$$

La qualité globale de la représentation du nuage $\mathcal{V}$ sur le sous espace principal F_k est mesurée par $\frac{\lambda_1+\lambda_2+\cdots+\lambda_k}{\sum_{i=1}^{p}\lambda_i}$.
En effet, ***la qualité de la représentation*** *de la variable y^i sur l'axe principal engendré par v_k est mesurée par :*

$$\cos^2(\widehat{y^i, v_k}) = \frac{< y^i, v_k >_M^2}{s_i^2} = r^2(y^i, v_k)$$

où $r(y^j, v_k)$ est le coefficient de corrélation linéaire entre y^j et v_k.

1.8 Aspects pratiques

1.8.1 Interprétation

A partir des relations données précédemment, nous pouvons définir quelques règles pour l'interprétation :

- Un individu sera du côté des variables pour lesquelles il a de fortes valeurs, inversement il sera du côté opposé des variables pour lesquelles il a de faibles valeurs.
- Plus les valeurs d'un individu sont fortes pour une variable plus il sera éloigné de l'origine suivant l'axe factoriel décrivant le mieux cette variable.
- Deux individus à une même extrémité d'un axe (i.e. éloignés de l'origine) sont proches (i.e. se ressemblent).

— Deux variables très corrélées positivement sont du même côté sur un axe.
— Il n'est pas possible d'interpréter la position d'un individu par rapport à une seule variable, et réciproquement, il n'est pas possible d'interpréter la position d'une variable par rapport à un seul individu. Les interprétations doivent se faire de manière globale.

Autrement dit, la représentation des variables dans l'espace des k facteurs permet d'interpréter visuellement les corrélations entre les variables .

En effet, qu'il s'agisse de la représentation des observations ou des variables dans l'espace des facteurs, deux points très éloignés dans un espace à k dimensions peuvent apparaître proches dans un espace à 2 dimensions en fonction de la direction utilisée pour la projection. D'une part, quand toutes les variables ont le même signe de corrélation avec la première composante principale, cette composante est alors appelée « facteur de taille », la seconde « facteur de forme » ; cela s'appelle **Effet de Taille** .

1.8.2 Nombre d'axe à retenir

Il ya des méthodes pour choisir les axes :

KAISER ne s'intéresse en général qu'aux axes dont les valeurs propres sont supérieures à la moyenne (qui vaut 1 en ACP normée).

Coude ou de Cattell utilise le résultat suivant : lorsque des variables sont peu corrélées, les valeurs propres de la matrice d'inertie décroissent régulièrement et l'ACP présente alors peu d'intérêt. A l'inverse, lorsqu'il existe une structure sur les données, on observe des ruptures dans la décroissance des valeurs propres. On cherchera donc à ne retenir que les axes correspondant aux valeurs qui précèdent la décroissance régulière.

Analytiquement, cela revient à chercher un point d'inflexion dans la décroissance des valeurs propres, et de ne pas aller au-delà dans l'analyse.

Critère du Scree-test s'intéresse sur les axes correspondant à des différences secondes >0 (un peu large).

1.9 Application

Considérons une donnée sous forme de tableau concernant la production de crevette par type de pecherie de Madagascar en 2018 obtenues par les bateaux SANTIG-DU, POSEIDON II, AGIOS SPYRIDON , AFRODITI , MELAKY 7 , MELAKY 8, MELAKY 2, MELAKY 3. On divise en deux parties le type de pêche : Crevette entiers et Poissons

	Crev. Entiers	Poissons
SANTIG-DU	16.6308	17.9489
POSEIDON II	14.0892	7.1728
AGIOS SPYRIDON	22.5171	20.04
AFRODITI	21.78164	27.76
MELAKY 7	11.8066	18.4195
MELAKY 8	11.2254	19.4965
MELAKY 2	8.5105	13.8635
MELAKY 3	9.245	10.97

1.9.1 Résultats généraux

	Eigenvalue	% variance
1	1.62033	81.017
2	0.37967	18.983

Corrélation entre les variables

	Crev. Entiers	Poissons
Crev. Entiers	0	0.10082
Poissons	0.62033	0

Résultats relatifs aux individus

Scores des individus

	PC1	PC2
SANTIG-DU	0.30948	0.27808
POSEIDON II	-0.90412	1.7032
AGIOS SPYRIDON	1.1007	1.1495
AFRODITI	1.7067	-0.4156
MELAKY 7	-0.14602	-0.83464
MELAKY 8	-0.11079	-1.1549
MELAKY 2	-0.88801	-0.70498
MELAKY 3	-1.0679	-0.02071

Contributions des individus La contribution relative d'un individu i à la formation de la composante principale α est l'inertie relative de cet individu sur l'axe factoriel k est :

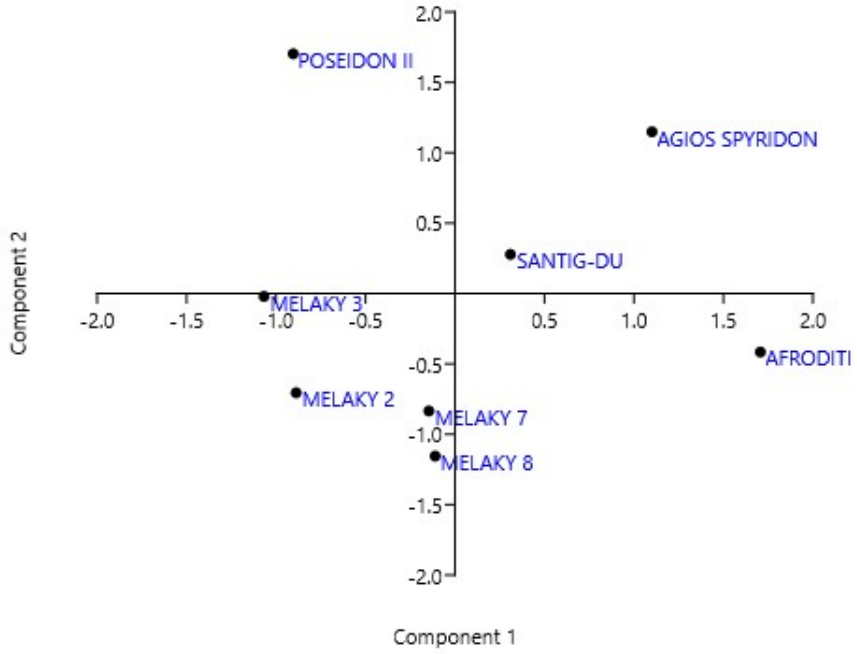
$$CTR_\alpha(i) = \frac{(\text{Score de i sur l'axe } \alpha)^2}{n\lambda_\alpha}$$

%	PC1	PC2
SANTIG-DU	0.74	2.55
POSEIDON II	6.31	95.51
AGIOS SPYRIDON	9.35	43.5
AFRODITI	22.47	5.69
MELAKY 7	0.16	22.94
MELAKY 8	0.09	43.91
MELAKY 2	6.08	16.36
MELAKY 3	8.8	0.01

Qualités de la représentation des individus La qualité de la représentation d'un individu i par la composante principale α est définie par :

$$QLT_\alpha(i) = \frac{(\text{Score de i sur l'axe } \alpha)^2}{\sum_l (\text{Score de i sur l'axe } l)^2}$$

Cosinus carré	PC1	PC2
SANTIG-DU	0.55	0.45
POSEIDON II	0.22	0.78
AGIOS SPYRIDON	0.49	0.52
AFRODITI	0.94	0.06
MELAKY 7	0.03	0.97
MELAKY 8	0.009	0.99
MELAKY 2	0.61	0.39
MELAKY 3	0.99	0.00006



On a ainsi, l'axe horizontal représente le résultat d'ensemble des bateaux : si on prend leur coordonnée sur l'axe horizontal, on obtient le même classement que si on prend leur moyenne générale.

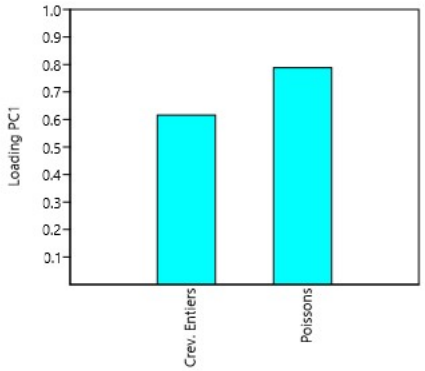
Axe 1	
+	-
AGIOS SPYRIDON,SANTIG-DU,AFRODITI	POSEIDON II,MELAKY 3,MELAKY 2 MELAKY 7,MELAKY 8

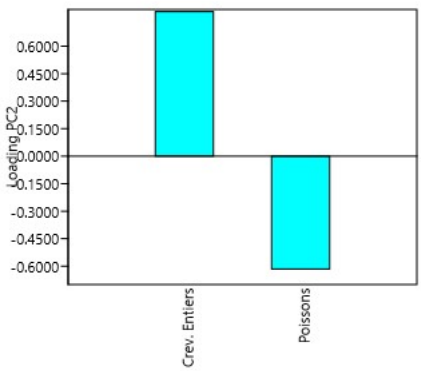
Par ailleurs, le bateau « le plus haut » sur le graphique, celui qui a la coordonnée la plus élevée sur l'axe vertical, est POSEIDON II dont les résultats sont les plus contrastés en faveur.

Axe 2	
+	-
POSEIDON II,AGIOS SPYRIDON,SANTIG-DU	AFRODITI, MELAKY 7,MELAKY 8 MELAKY 3,MELAKY 2

Résultats relatifs aux variables

Corrélation	PC1	PC2
Crevette entiers	0.61562	0.78804
Poissons	0.78804	-0.615262





Axe 1	
+	-
Crevette entiers Poissons	

Axe 2	
+	-
Crevette entiers	Poissons

On a ici clairement l'Effet de Taille, puisque toutes les variables ont le même signe (positif) de corrélation avec la première composante principale.

Chapitre 2

Analyse factorielle des correspondances (AFC)

2.1 Introduction

L'analyse factorielle des correspondances (AFC) est une méthode statistique d'analyse des données mise au point à partir des années 1960 par Jean-Paul Benzécri et son équipe d'abord à la faculté des sciences de Rennes puis à celle de Jussieu à Paris. Elle se rattache à la famille des analyses factorielles qui regroupe différentes méthodes d'analyses de grands tableaux rectangulaires de données, visant toutes à identifier et à hiérarchiser des facteurs corrélés aux données placées en colonnes.
Les méthodes d'analyse factorielle des correspondances (AFC) tout comme celles d'analyse en composantes principales (ACP) s'utilisent pour décrire et hiérarchiser les relations statistiques qui peuvent exister entre des individus placés en ligne et des variables placées en colonne dans un tableau rectangulaire de données. L'une et l'autre de ces deux méthodes considèrent le tableau de données comme un nuage de points dans un espace mathématique ayant autant de dimensions qu'il y a de colonnes dans le tableau de données ; elles cherchent à le projeter sur des axes ou des plans (appelés factoriels) de façon que l'on puisse en visualiser et étudier au mieux la forme et donc rechercher globalement des corrélations. La spécificité de l'AFC est qu'elle considère en même temps un nuage de point représentant les lignes (individus) et un autre représentant les colonnes (variables).

2.2 Données

Soit X une variable qualitative. On dispose d'un échantillon de n individus sur lesquels la variable est mesurée.

Définition 13 (Modalités ou catégories) *les valeurs que peut prendre une variable qualitative ; si la variable a m modalités (valeurs possibles), on note x_i, $1 \leq i \leq m$, ces modalités, ou plus simplement i.*

Définition 14 (Effectif) *le nombre d'occurrences de la modalité i dans l'échantillon ; on le note n_i tel que $\sum_{i=1}^{m} n_i = n$.*

Définition 15 (Profil) *c'est l'ensemble des valeurs n_i/n ; la somme du profil sur les modalités est* 1.

Soient X et Y deux variables qualitatives à r et s modalités (notées $l_1, l_2, \cdots, l_r$ et l'autre par $d_1, d_2, \cdots, d_s$) respectivement décrivant un ensemble de n individus.

Définition 16 *le tableau de contingence est une matrice à r lignes et s colonnes renfermant les effectifs n_{ij} d'individus. La matrice peut s'écrire*

$$N = \begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1s} \\ n_{21} & n_{22} & \cdots & n_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ n_{r1} & n_{r2} & \cdots & n_{rs} \end{pmatrix}$$

TABLE 2.1 – Tableau de contingences

L'opération consistant à établir un tel tableau est appelée un "tri croisé" dans le domaine de l'enquête. Les effectifs marginaux sont :

Marge ligne l'effectif total de la modalité i de X, c'est-à-dire,

$$n_{i+} = \sum_{j=1}^{s} n_{ij}$$

On définit aussi le profil marginal des lignes n_{i+}/n.

Marge colonne c'est la somme $n_{+j} = \sum_{i=1}^{r} n_{ij}$, c'est-à-dire l'effectif total de la modalité j de Y.
On définit aussi le profil marginal des colonnes n_{+j}/n.

Remarque 8
— *Le tableau des profils-lignes n_{ij}/n_{i+}, qui représente la fréquence de la modalité j conditionnellement à $X = i$; la somme de chaque ligne est ramenée à* 100% .
— *De même pour le tableau des profils-colonnes, n_{ij}/n_{+j} représente la fréquence de la modalité i conditionnellement à $Y = j$ tel que leur somme vaut* 100%.

2.3 Objectifs

Le but de l'Analyse Factorielle des Correspondances consiste à représenter un maximum de l'inertie totale sur le premier axe factoriel, un maximum de l'inertie résiduelle sur le second axe, et ainsi de suite jusqu'à la dernière dimension.Ainsi, permet de résumer et de visualiser l'information contenue dans le tableau de contingence formé par les deux variables catégorielles.

2.4 Principes de l'AFC

Le principe de ces méthodes est de partir sans a priori sur les données et de les décrire en analysant la hiérarchisation de l'information présente dans les données.
Rappelons que notre tableau de données est un tableau de contingence N à r lignes et s colonnes.
Notons $D_r = diag(n_{1+}, n_{2+}, \cdots, n_{r+})$ et $D_s = diag(n_{+1}, n_{+2}, \cdots, n_{+s})$ les matrices diagonales des effectifs marginaux des variables X et Y .
Le tableau des profils des lignes n_{ij}/n_{i+} est donné par $T_r = D_r^{-1}N$.
Celui des profils des colonnes n_{ij}/n_{+j} par $T_s = ND_s^{-1}$.

Remarque 9 *Lorqu'on réalise l'AFC d'une table de contingence comportant r lignes et s colonnes, la dimension de l'espace dans lequel se trouve l'ensemble des résultats est* $\inf(r-1, s-1)$.

	1	$\cdots$	j	$\cdots$	s
1					
$\vdots$			$\vdots$		
i		$\cdots$	$\frac{n_{ij}}{n_{i+}}$	$\cdots$	
$\vdots$			$\vdots$		
r					

	1	$\cdots$	j	$\cdots$	s
1					
$\vdots$			$\vdots$		
i		$\cdots$	$\frac{n_{ij}}{n_{+j}}$	$\cdots$	
$\vdots$			$\vdots$		
r					

TABLE 2.2 – Tableau des profils des lignes et Colonnes

2.4.1 Représentation géométrique des profils

Nuages des profils-lignes

Les profils-lignes forment un nuage de r points de $\mathbb{R}^s$. Le i-ème profil-ligne est alors muni du poids $f_{i+} = \frac{n_{i+}}{n}$. On définit alors la matrice des poids $\frac{1}{n}D_r$.

Définition 17 *On appelle nuage des profils-lignes $\mathcal{M}_r$ l'ensemble des r points L_i de $\mathbb{R}^s$ munis de leurs poids f_{i+} : $\mathcal{M}_r = \{(L_i, f_{i+}); i = 1, \cdots, r\}$.*

Propriétés 1 *1. Le centre de gravité g_r du nuage $\mathcal{M}_r$ est donné par :*

$$g_r = \begin{pmatrix} \frac{n_{+1}}{n} \\ \frac{n_{+2}}{n} \\ \vdots \\ \frac{n_{+s}}{n} \end{pmatrix} = \begin{pmatrix} f_{+1} \\ f_{+2} \\ \vdots \\ f_{+s} \end{pmatrix}$$

Autrement dit, c'est le profil marginal des colonnes.

2. Les points L_i de $\mathcal{M}_r$, ainsi que leur centre de gravité g_r, appartiennent à un sous-espace affine de $\mathbb{R}^s$, à savoir l'hyperplan $\mathcal{H}_{s-1}$ de dimension $s-1$ défini par :

$$\mathcal{H}_{s-1} = \{(x_1, x_2, \cdots, x_s) \in \mathbb{R}^s; \sum_{i=1}^{s} x_i = 1\}$$

Preuve

1. Notons $g_r = \sum_{i=1}^{r} f_{i+} L_{i+}$. Ainsi pour tout $j \in \{1, \cdots, s\}$,

$$g_r(j) = \sum_{i=1}^{r} f_{i+} L_i(j) = \sum_{i=1}^{r} \frac{n_{i+}}{n} \frac{n_{ij}}{n_{i+}} = \sum_{i=1}^{r} \frac{n_{ij}}{n} = \frac{n_{+j}}{n}$$

D'où,

$$g_r = \begin{pmatrix} g_r(1) \\ \vdots \\ g_r(s) \end{pmatrix} = \begin{pmatrix} f_{+1} \\ \vdots \\ f_{+s} \end{pmatrix}$$

2. Pour tout $i \in \{1, \cdots, r\}$,

$$\sum_{j=1}^{s} L_i(j) = \sum_{j=1}^{s} \frac{n_{ij}}{n_{i+}} = \frac{n_{i+}}{n_{i+}} = 1$$

On en déduit que chaque L_i est dans $\mathcal{H}_{s-1}$.
D'autre part, g_r est une combinaison linéaire des L_i. Il s'ensuit que,g_r est aussi dans $\mathcal{H}_{s-1}$.

Nuages des profils-colonnes

Les deux variables X et Y jouant des rôles symétriques, ce qui vient d'être fait pour les profils-lignes peut aussi être fait pour les profils-colonnes. Chaque profil-colonne C_j est un point dans l'espace $\mathbb{R}^r$. L'ensemble des profils-colonnes forme donc un nuage de s points dans $\mathbb{R}^r$. Chaque point est affecté d'un poids égal à sa fréquence marginale $\frac{n_{+j}}{n}$, et la matrice des poids est donc $\frac{1}{n}D_s$.

Définition 18 *On appelle nuage des profils-colonnes $\mathcal{M}_s$, l'ensemble des s points C_j de $\mathbb{R}^r$ munis de leurs poids f_{+j} : $\mathcal{M}_s = \{(C_j, f_{+j}); j = 1, \cdots, s\}$*

De manière similaire que dans la profil-ligne, le centre de gravité à pour coordonnée

$$g_s = \begin{pmatrix} \frac{n_{1+}}{n} \\ \frac{n_{2+}}{n} \\ \vdots \\ \frac{n_{r+}}{n} \end{pmatrix} = \begin{pmatrix} f_{1+} \\ f_{2+} \\ \vdots \\ f_{r+} \end{pmatrix}$$

De plus, les points C_j de $\mathcal{M}_s$, ainsi que leur centre de gravité g_s, appartiennent à un sous-espace affine de $\mathbb{R}^r$, à savoir l'hyperplan $\mathcal{H}_{r-1}$ de dimension $r-1$ défini par :

$$\mathcal{H}_{r-1} = \{(x_1, x_2, \cdots, x_r) \in \mathbb{R}^r; \sum_{i=1}^{r} x_i = 1\}$$

2.4.2 ACP sur un nuage de profils

Métrique du χ^2

Profils-lignes La distance choisie entre deux profils-lignes L_i et L_j est la métrique du chi2 définie par :

$$d^2_{\chi^2}(L_i, L_l) =< L_i - L_l, L_i - L_l >_M = \sum_{j=1}^{s} \frac{1}{f_{+j}} (L_i(j) - L_l(j))^2 = \sum_{j=1}^{s} \frac{n}{n_{+j}} \left(\frac{n_{ij}}{n_{i+}} - \frac{n_{lj}}{n_{l+}} \right)^2$$

où la matrice M est la matrice diagonale définie par $M = nD_s^{-1}$. Intuitivement, la pondération n/n_{+j} permet de donner des importances comparables aux différentes « variables ».

De façon plus fondamentale, cette distance a la propriété d'équivalence distributionnelle, si deux colonnes j et k de N ont le même profil, il est logique de les regrouper en une seule d'effectif $n_{ij} + n_{ik}$; on a alors quand $\frac{n_{ij}}{n_{+j}} = \frac{n_{ik}}{n_{+k}}$

$$\frac{n}{n_{+j}} \left(\frac{n_{ij}}{n_{i+}} - \frac{n_{+j}}{n} \right)^2 + \frac{n}{n_{+k}} \left(\frac{n_{ik}}{n_{i+}} - \frac{n_{+k}}{n} \right)^2 = \frac{n}{n_{+j} + n_{+k}} \left(\frac{n_{ij} + n_{ik}}{n_{i+}} - \frac{n_{+j} + n_{+k}}{n} \right)^2$$

La distance entre les profils-ligne est donc inchangée.

Profils-colonnes On définit la distance entre deux profils-colonnes C_j et C_k comme

$$d^2_{\chi^2}(C_j, C_k) = \sum_{i=1}^{r} \frac{1}{f_{i+}} (C_j(i) - C_k(i))^2 = \sum_{i=1}^{r} \frac{n}{n_{i+}} \left(\frac{n_{ij}}{n_{+j}} - \frac{n_{ik}}{n_{+k}} \right)^2$$

soit une métrique de matrice $M = nD_r^{-1}$.
Les propriétés sont les mêmes que les profils-lignes.

Propriétés 2 *L'inertie totale du nuage des profils-lignes par rapport à g_r est*

$$I_{g_r} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{1}{n_{i+}n_{+j}} \left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2 \tag{2.1}$$

Preuve : Par définition, $I_{g_r} = \sum_{i=1}^{r} f_{i+} d^2_{\chi^2}(L_i, g_r)$
Comme,$f_{i+} = \frac{n_{i+}}{n}$, $L_i(j) = \frac{n_{ij}}{n_{i+}}$ et $g_r(j) = \frac{n_{+j}}{n}$
Par suite,

$$I_{g_r} = \sum_{i=1}^{r} \frac{n_{i+}}{n} \sum_{j=1}^{s} \frac{n}{n_{+j}} \left(\frac{n_{ij}}{n_{i+}} - \frac{n_{+j}}{n}\right)^2$$

Mettant en facteur $\frac{1}{n_{i+}}$ dans le terme carré,
on en déduit que,

$$I_{g_r} = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{1}{n_{i+}n_{+j}} \left(n_{ij} - \frac{n_{i+}n_{+j}}{n}\right)^2$$

On conclut que cette inertie mesure donc l'écart à l'indépendance.
On obtient de même façon l'inertie des profils-colonnes.

ACP des deux nuages profils Par analogie avec les notations du chapitre sur l'ACP, on a :

	Données	Métrique	Poids
Profils-lignes	$X = T_r = D_r^{-1}N$	$M = nD_s^{-1}$	$D = \frac{1}{n}D_r$
Profils-colonnes	$X = T_s' = D_s^{-1}N'$	$M = nD_r^{-1}$	$D = \frac{1}{n}D_s$

TABLE 2.3 – Tableau de Référence sur ACP

ACP Profils-lignes

Matrice des variances-covariances

$$V = X'D_pX - gg' = Y'D_pY$$

Autrement dit,

$$V = (D_r^{-1}N)'(\frac{1}{n}D_r)(D_r^{-1}N) - g_rg_r'$$

Par suite,

$$V = \frac{1}{n}N'D_r^{-1}N - g_rg_r'$$

Alors,$VM = \frac{1}{n}N'D_r^{-1}NM - g_rg_r'M = \frac{1}{n}N'D_r^{-1}N(nD_s^{-1}) - g_rg_r'(nD_s^{-1})$
Par conséquent,

$$VM = N'D_r^{-1}ND_s^{-1} - ng_rg_r'D_s^{-1}$$

Propriétés 3
1. *g_r est un vecteur propre de VM associé à la valeur propre 0,et de $N'D_r^{-1}ND_s^{-1}$ associé à la valeur propre 1.*
2. *Les autres vecteurs propres sont orthogonaux à g_r, et sont associés aux mêmes valeurs propres pour les deux matrices.*

Preuve :

1. On a ,

$$Mg_r = nD_s^{-1}\begin{pmatrix} \frac{n_{+1}}{n} \\ \frac{n_{+2}}{n} \\ \vdots \\ \frac{n_{+s}}{n} \end{pmatrix} = diag(\frac{1}{n_{+1}}, \frac{1}{n_{+2}}, \cdots, \frac{1}{n_{+s}}) \begin{pmatrix} n_{+1} \\ n_{+2} \\ \vdots \\ n_{+s} \end{pmatrix} = \mathbf{1}_s$$

Il s'ensuit que, $g_r' M g_r = \begin{pmatrix} f_{+1} & f_{+2} & \cdots & f_{+s} \end{pmatrix} \mathbf{1}_s = \sum_{j=1}^{s} f_{+j} = 1$
Par suite,$VMg_r = \frac{1}{n} N' D_r^{-1} N \underbrace{Mg_r}_{\mathbf{1}_s} - g_r \underbrace{g_r' M g_r}_{1}$

$$= \frac{1}{n} N' \underbrace{D_r^{-1}}_{=diag(\frac{1}{n_{1+}}, \cdots, \frac{1}{n_{r+}})} \begin{pmatrix} \sum_{j=1}^{s} n_{1j} \\ \vdots \\ \sum_{j=1}^{s} n_{rj} \end{pmatrix} - g_r$$

or $N' = \begin{pmatrix} n_{11} & n_{21} & \cdots & n_{r1} \\ n_{12} & n_{22} & \cdots & n_{r2} \\ \vdots & \vdots & \ddots & \vdots \\ n_{1s} & n_{2s} & \cdots & n_{rs} \end{pmatrix}$
donc,

$$VMg_r = \frac{1}{n} \begin{pmatrix} n_{11} & n_{21} & \cdots & n_{r1} \\ n_{12} & n_{22} & \cdots & n_{r2} \\ \vdots & \vdots & \ddots & \vdots \\ n_{1s} & n_{2s} & \cdots & n_{rs} \end{pmatrix} \mathbf{1}_r - g_r$$

Par conséquent,

$$VMg_r = \underbrace{\frac{1}{n} \begin{pmatrix} \sum_{i=1}^{r} n_{i1} \\ \vdots \\ \sum_{i=1}^{r} n_{is} \end{pmatrix}}_{g_r} - g_r$$

D'où,

$$VMg_r = 0$$

D'autre part, $N' D_r^{-1} N D_s^{-1} = VM + g_r g_r' M$
c'est-à-dire,

$$N' D_r^{-1} N D_s^{-1} g_r = \underbrace{VMg_r}_{0} + g_r \underbrace{g_r' M g_r}_{1}$$

En effet,

$$N' D_r^{-1} N D_s^{-1} g_r = g_r$$

2. Soit u des vecteurs propres orthogonaux à g_r
Calculant, $VMu = N' D_r^{-1} N D_s^{-1} u - g_r g_r' M u$ or g_r et u orthogonaux alors $g_r' M u = 0$
Il s'ensuit que

$$VMu = N' D_r^{-1} N D_s^{-1} u = \lambda u$$

Facteurs principaux

Notons $u_k, k \in \{1, \cdots, r-1\}$ les vecteurs propres autres que g_r de VM et $N'D_r^{-1}ND_s^{-1}$ tel que $VMu_k = N'D_r^{-1}ND_s^{-1}u_k = \lambda_k u_k$
D'après le résultat d'ACP,le facteur principal $v_k = Mu_k$ est un vecteur propre de MV

Propriétés 4 *Les facteurs principaux sont des vecteurs propres de $D_s^{-1}N'D_r^{-1}N$.*

Preuve :
Pour chaque axe principal k,

$$D_s^{-1}N'D_r^{-1}Nv_k = D_s^{-1}N'D_r^{-1}N(Mu_k) = nD_s^{-1}\underbrace{N'D_r^{-1}ND_s^{-1}u_k}_{\lambda_k u_k}$$

or $M = nD_s^{-1}$.
D'où,

$$D_s^{-1}N'D_r^{-1}Nv_k = M\lambda_k u_k = \lambda_k v_k$$

Composantes principales

Les composantes principales donnent les coordonnées des profils-lignes sur chaque axe :

$$c^k(i) =< L_i, u_k >_{\chi^2} = nL_i'D_s^{-1}u_k = n\begin{pmatrix} \frac{n_{i1}}{n_{i+}} & \cdots & \frac{n_{is}}{n_{i+}} \end{pmatrix} diag(\frac{1}{n_{+1}}, \cdots, \frac{1}{n_{+s}})u_k$$

Autrement dit,

$$c^k(i) = n\sum_{j=1}^{s}\frac{n_{ij}}{n_{i+}n_{+j}}u_k(j)$$

Par analogie de l'ACP, $c^k = YMu_k = XMu_k = Xv_k = nD_r^{-1}ND_s^{-1}u_k$

Propriétés 5 *Les composantes principales sont des vecteurs propres de $D_r^{-1}ND_s^{-1}N'$ associés aux valeurs propres de VM.*

Preuve :
On a, $D_r^{-1}ND_s^{-1}N'c^k = D_r^{-1}ND_s^{-1}N'Xv_k = D_r^{-1}ND_s^{-1}N'D_r^{-1}Nv_k$
or précédemment, $D_s^{-1}N'D_r^{-1}Nv_k = M\lambda_k u_k = \lambda_k v_k$
Par suite, $D_r^{-1}ND_s^{-1}N'c^k = D_r^{-1}N\lambda_k v_k$
Par conséquent, $D_r^{-1}ND_s^{-1}N'c^k = \lambda_k D_r^{-1}Nv_k = \lambda_k Xv_k$
D'où,

$$D_r^{-1}ND_s^{-1}N'c^k = \lambda_k c^k$$

ACP Profils-colonnes

Symétriquement comme dans Profils-lignes, on échange les indices r et s et on transpose N. On en déduit que,

$$V = \frac{1}{n}ND_s^{-1}N' - g_s g_s' \text{ ainsi } VM = ND_s^{-1}N'D_r^{-1} - ng_s g_s' D_r^{-1}$$

De plus, montrons que g_s est un vecteur propre de VM associé à la valeur propre 0, et que diagonaliser VM revient à diagonaliser la matrice $ND_s^{-1}N'D_r^{-1}$.

Propriétés 6 *On note $C = ND_s^{-1}N'D_r^{-1}$.*
$D_r c^k$ est un vecteur propre de C associé à la valeur propre λ_k non nulle.

Preuve : Rappelons $D_r c^k = D_r(nD_r^{-1}ND_s^{-1}u_k) = nND_s^{-1}u_k$.
D'une part,

$$C(D_r c^k) = ND_s^{-1}N'D_r^{-1}(nND_s^{-1}u_k) = nND_s^{-1}(N'D_r^{-1}ND_s^{-1}u_k)$$

Or u_k est un vecteur propre de $N'D_r^{-1}ND_s^{-1}$ associé à λ_k non nulle
Alors $N'D_r^{-1}ND_s^{-1}u_k = \lambda_k u_k$
Par conséquent,

$$C(D_r c^k) = nND_s^{-1}\lambda_k u_k$$

On en déduit que

$$C(D_r c^k) = \lambda_k(D_r c^k)$$

Facteurs principaux

Notons par a_k les vecteurs principaux de l'ACP des profils-colonnes correspondant aux valeurs propres non nulles.
D'après l'ACP, le facteur principal $b_k = Ma_k = nD_r^{-1}a_k$.
On admet que , les facteurs principaux sont les vecteurs propres de $D_r^{-1}ND_s^{-1}N'$.

Composantes principales

Notons $\tilde{c}^k$ les composantes principales de l'ACP des profils-colonnes. Ses coordonnées sur l'axe de vecteur directeur a_k :

$$\tilde{c}^k(j) =< a_k, C_i >_{\chi^2} = nC_i'D_r^{-1}a_k = n\sum_{i=1}^{r}\frac{n_{ij}}{n_{i+}n_{+j}}a_k(i)$$

Autrement dit,

$$\tilde{c}^k = nD_s^{-1}N'D_r^{-1}a_k$$

Nous rappelons que les composantes principales c^k et $\tilde{c}^k$ sont centrées, et de variance λ_k.

Formules de transition Précédament $N'D_r^{-1}ND_s^{-1}$ et $ND_s^{-1}N'D_r^{-1}$ ont mêmes valeurs propres non nulles λ_k. Leurs vecteurs propres sont reliés par les relations suivantes :

Théorème 2 *Soit $p = rang(N'D_r^{-1}ND_s^{-1}) = rang(ND_s^{-1}N'D_r^{-1})$. Pour tout $k \leq p$, il existe une relation dite de transition, entre les vecteurs propres u_k et a_k :*

$$\begin{cases} u_k = \frac{1}{\sqrt{\lambda_k}}T_r'a_k \\ a_k = \frac{1}{\sqrt{\lambda_k}}T_s u_k \end{cases} \tag{2.2}$$

Preuve : Rappelons que $a_k = \frac{D_r c^k}{\|D_r c^k\|_{\chi^2}}$
or $\|D_r c^k\|_{\chi^2}^{\ 2} = n(D_r c^k)'D_r^{-1}D_r c^k = n(c^k)'D_r c^k = n\sum_{i=1}^{r} n_{i+}c^k(i)^2 = n^2 var(c^k) = n^2\lambda_k$. On admet que $a_k = \frac{D_r c^k}{n\sqrt{\lambda_k}}$
D'une part, $c^k = nD_r^{-1}ND_s^{-1}u_k$ ainsi $D_r c^k = nND_s^{-1}u_k$.
Par suite, $\frac{D_r c^k}{n\sqrt{\lambda_k}} = \frac{ND_s^{-1}u_k}{\sqrt{\lambda_k}}$
Par conséquent,

$$a_k = \frac{1}{\sqrt{\lambda_k}}T_s u_k$$

D'autre part, $N'D_r^{-1}a_k = \frac{1}{\sqrt{\lambda_k}}N'D_r^{-1}ND_s^{-1}u_k$
Sachant que $N'D_r^{-1}ND_s^{-1}u_k = \lambda_k u_k$ car u_k est un vecteur propre de $N'D_r^{-1}ND_s^{-1}$ associé à λ_k.

En effet, $T_r' a_k = \frac{1}{\sqrt{\lambda_k}} \lambda_k u_k = \sqrt{\lambda_k} u_k$
D'où,

$$u_k = \frac{1}{\sqrt{\lambda_k}} T_r' a_k$$

Théorème 3 *Pour tout $k \leq p$,*

$$c^k = \frac{1}{\sqrt{\lambda_k}} D_r^{-1} N \tilde{c}^k$$

$$\tilde{c}^k = \frac{1}{\sqrt{\lambda_k}} D_s^{-1} N' c^k$$

Preuve : On a $c^k = n D_r^{-1} N D_s^{-1} u_k$ or $u_k = \frac{1}{\sqrt{\lambda_k}} N' D_r^{-1} a_k$
Par suite,

$$c^k = n D_r^{-1} N D_s^{-1} \frac{1}{\sqrt{\lambda_k}} N' D_r^{-1} a_k$$

Autrement dit,

$$c^k = \frac{1}{\sqrt{\lambda_k}} D_r^{-1} N (n D_s^{-1} N' D_r^{-1} a_k)$$

Comme $D_s^{-1} N' D_r^{-1} a_k = \tilde{c}^k$
Il s'ensuit que

$$c^k = \frac{1}{\sqrt{\lambda_k}} D_r^{-1} N \tilde{c}^k$$

Réciproquement, $\tilde{c}^k = n D_s^{-1} N' D_r^{-1} a_k = n D_s^{-1} N' D_r^{-1} \frac{1}{\sqrt{\lambda_k}} T_s u_k = \frac{1}{\sqrt{\lambda_k}} n D_s^{-1} N' D_r^{-1} N D_s^{-1} u_k$ C'est-à- dire, $\tilde{c}^k = \frac{1}{\sqrt{\lambda_k}} D_s^{-1} N' (n D_r^{-1} N D_s^{-1} u_k)$
Précédemment, $c^k = n D_r^{-1} N D_s^{-1} u_k$
On conclut que,

$$\tilde{c}^k = \frac{1}{\sqrt{\lambda_k}} D_s^{-1} N' c^k$$

Comparaison lignes-colonnes

Les deux analyses conduisent aux mêmes valeurs propres et les facteurs principaux de l'une sont les composantes principales de l'autre (à un facteur près).

2.4.3 Interprétation

Par analogie avec les notations du chapitre sur l'ACP

Contribution

La Contribution relative du profil-ligne L_i au k-ième axe est

$$\frac{p_i (c_i^k)^2}{\lambda_k} = \frac{f_{i+} (c^k(i))^2}{\lambda_k}$$

Pour les mêmes raisons, la contribution relative du profil-colonne C_j au k-ième axe est

$$\frac{p_j (c_j^k)^2}{\lambda_k} = \frac{f_{+j} (\tilde{c}^k(j))^2}{\lambda_k}$$

Qualité de Représentation

L'AFC est une ACP et on peut donc mesurer la qualité de la représentation de la modalité i

— Qualité de la représentation du profil-ligne L_i au k-ième axe est

$$\cos^2(\widehat{L_i, u_k}) = \frac{(c^k(i))^2}{\sum_{k=1}^{s}(c^k(i))^2}$$

— Qualité de la représentation du profil-colonne C_j au k-ième axe est

$$\cos^2(\widehat{C_j, u_k}) = \frac{(\tilde{c}^k(j))^2}{\sum_{k=1}^{r}(\tilde{c}^k(j))^2}$$

2.5 Application

On a relevé sur $n = 10$ individus deux variables qualitatives, la variable X à 4 modalités {A,B,C,D} et la variable Y à trois modalités {I,II,III}. Les résultats sont regroupés dans la table 2.4 qui donne sous forme d'une ★ les modalités relevées sur un individu.

Ind	A	B	C	D	I	II	III
1	★				★		
2	★						★
3			★			★	
4			★				★
5		★			★		
6	★				★		
7	★					★	
8			★			★	
9				★			★
10			★		★		

TABLE 2.4 – Tableau de présence/absence

Tableau de contingence

	I	II	III
A	2	1	1
B	1	0	0
C	1	2	1
D	0	0	1

Marges de lignes et colonnes

A	B	C	D
4	1	4	1

I	II	III
4	3	3

Valeurs propres le nombre de valeurs propres est $\min(4-1; 3-1) = 2$

Axis	Eigenvalue	% of total	Cumulative
1	0.3125	65.217	65.217
2	0.166667	34.783	100

Coordonnées, contribution et qualité des lignes

	Coordonnées		Contributions		$\cos^2$	
	Axis 1	Axis 2	Axis 1	Axis 2	Axis 1	Axis 2
A	0.172516	0.109109	0.038	0.029	0.71	0.29
B	1.0351	0.654654	0.343	0.257	0.71	0.29
C	-0.0862582	-0.436436	0.01	0.457	0.038	0.96
D	-1.38013	0.654654	0.61	0.257	0.82	0.18

Coordonnées, contribution et qualité des colonnes

	Coordonnées		Contributions		$\cos^2$	
	Comp1	Comp2	Axis 1	Axis 2	Axis 1	Axis 2
I	0.578638	0.267261	0.43	0.17	0.82	0.18
II	4.7×10^{-16}	-0.62361	6.6×10^{-32}	0.7	5.6×10^{-31}	0.999
III	-0.771517	0.267261	0.57	0.13	0.89	0.11

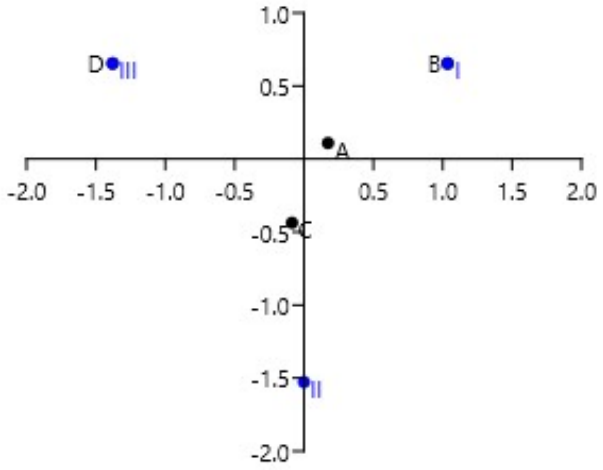
L'axe 1 oppose I au III du point de vue de leur profil.

Comme en ACP, pour chaque tableau, la somme des contributions d'une colonne égale 1 (ou 100%). Ce qui entraîne une moyenne des contributions égale à 1/r pour les modalités ligne, et 1/s pour les modalités colonne. Dans notre exemple, r=4 et s=3, donc les moyennes des contributions sont égales à 0,25 et 0.333.

	-	+
Axe 1	D, III	B,I
Axe 2	C, II	B, D

L'axe 1 oppose B, I aux D, III du point de vue de leur profil.
De même, les sommes des cosinus carrés pour une ligne égalent 1. Donc la moyenne des cosinus carrés égale 1/nombre d'axes total = 1/2=0.5.

	-	+
Axe 1	D, III	B,A,I
Axe 2	C, II	



Chapitre 3

Analyse des Correspondances Multiples(ACM)

3.1 Introduction

L'analyse factorielle des correspondances, vue dans le chapitre précédent, s'applique à des situations où les individus statistiques sont décrits par deux variables nominales. Mais il est fréquent que l'on dispose d'individus décrits par plusieurs (deux ou plus) variables nominales ou ordinales. C'est notamment le cas lorsque nos données sont les résultats d'une enquête basée sur des questions fermées. Une extension de l'AFC à ces situations a donc été proposée. Elle est généralement appelée Analyse des Correspondances Multiples ou ACM.

Nous nous plaçons donc dans la situation où nous disposons de n individus statistiques, décrits par p variables nominales ou ordinales $X^1, X^2, \cdots, X^p$. L'ACM vise à mettre en évidence :

- les relations entre les modalités des différentes variables ;
- éventuellement, les relations entre individus statistiques ;
- les relations entre les variables, telles qu'elles apparaissent à partir des relations entre modalités.

On considère donc dans cette section p variables qualitatives observées simultanément sur n individus de poids identiques $\frac{1}{n}$.

		VARIABLES				
		1	$\cdots$	j	$\cdots$	p
	1	k_{11}	$\cdots$	k_{1j}	$\cdots$	k_{1p}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
INDIVIDUS	i	k_{i1}	$\cdots$	k_{ij}	$\cdots$	k_{ip}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	n	k_{n1}	$\cdots$	k_{nj}	$\cdots$	k_{np}

TABLE 3.1 – Tableau des données sous forme de codage condensé.

3.2 Objectifs

On veut réaliser l'AFC au cas de $p \geq 2$ variables $X^1, X^2, \cdots, X^p$ à $m_1, m_2, \cdots, m_p$ modalités. Ceci est particulièrement utile pour l'exploration d'enquêtes où les questions sont à

réponses multiples.
Soit $k_{ij} \in X^j$ avec X^j qui est l'ensemble des modalités de la $j^{\text{ème}}$variable.

3.3 Définitions et notations

3.3.1 Le Tableau disjonctif complet

Un Tableau disjonctif est tel que chaque ligne correspond à un individu et chaque colonne à une modalité. Les observations x_{ij} sont codées 1 si l'individu i a la modalité j et 0 sinon.

Particulièrement, le tableau disjonctif à la variable X^j est la matrice à n lignes et m_j colonnes. On le note par X_j.

Remarque 10 *1. On vérifie facilement que le tableau de contingence des variables X^j et X^k est*

$$N_{jk} = X_j' X_k$$

2. On admet que les effectifs marginaux de la variable X^j est donnée par la matrice diagonale $D_j = X_j' X_j$.

Le Tableau disjonctif complet est la matrice $X = (X_1|X_2|\cdots|X_p)$ constitue de n lignes et $m = m_1 + m_2 + \cdots + m_p$ colonnes ($m_j = card(X_j)$).

$X=$

	1	$\cdots$	j	$\cdots$	m
1	x_{11}	$\cdots$	x_{1j}	$\cdots$	x_{1m}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
i	x_{i1}	$\cdots$	x_{ij}	$\cdots$	x_{im}
$\vdots$	$\vdots$		$\vdots$		$\vdots$
n	x_{n1}	$\cdots$	x_{nj}	$\cdots$	x_{nm}

TABLE 3.2 – Transformation des données en TDC sur l'ACM

$$\begin{cases} x_{ij} = 1 & \text{si l'individu } i \text{ possède la modalité } j \\ x_{ij} = 0 & \text{sinon} \end{cases}$$

3.3.2 Tableau Burt

On appelle tableau de Burt le tableau $B = (B_{j,k})_{j,k=1,\cdots,p} = X'X$. Autrement dit, formé de tableaux de contingence et de matrices d'effectifs marginaux.

	1	$\cdots$	s'	$\cdots$	m
1					
$\vdots$			$\vdots$		
s		$\cdots$	$b_{ss'}$	$\cdots$	
$\vdots$			$\vdots$		
m					

TABLE 3.3 – Représentation des données sous forme du tableau de Burt.

Avec :

— $b_{ss'} = \sum_{i=1}^{n} x_{is}x_{is'}$ représente le nombre d'individus qui possèdent la modalité s et s'.

— b_{ss} le nombre d'individus qui possèdent la modalité s situé sur la diagonale noté n_s.

Remarque 11 *si on considère les données du tableau disjonctif X comme des observations de variables qualitatives, alors le tableau de Burt représente la variance de X à un facteur multiplicatif près.*

Propriétés 7 *1. B est une matrice symétrique.*

2. La somme des lignes (resp. des colonnes) de B est pn_s, $s = 1, \cdots, m$.

3. La somme des éléments de B est np^2.

Preuve :

Calculons le transposé de B

$$B' = (X'X)' \text{ or } (AB)' = B'A'$$

En effet,

$$B' = X'(X')' = X'X$$

D'où B symétrique.
D'une part, Calculons la somme de la ligne s de B

$$\sum_{s'=1}^{m} b_{ss'} = \sum_{s'=1}^{m}\sum_{i=1}^{n} x_{is}x_{is'}$$

Par suite,

$$\sum_{s'=1}^{m} b_{ss'} = \sum_{i=1}^{n} x_{is}\sum_{s'=1}^{m} x_{is'}$$

Comme $\sum_{s'=1}^{m} x_{is'} = p$,
alors

$$\sum_{s'=1}^{m} b_{ss'} = \sum_{i=1}^{n} x_{is}p = p\sum_{i=1}^{n} x_{is}$$

or $\sum_{i=1}^{n} x_{is} = n_s$
D'où

$$\sum_{s'=1}^{m} b_{ss'} = pn_s$$

D'autre part, la somme des éléments de B est

$$\sum_{s=1}^{m}\sum_{s'=1}^{m} b_{ss'} = \sum_{s=1}^{m} pn_s$$

Autrement dit,

$$\sum_{s=1}^{m}\sum_{s'=1}^{m} b_{ss'} = \sum_{j=1}^{p} p\underbrace{\sum_{s=1}^{m_j} n_s}_{=n}$$

Par conséquent,

$$\sum_{s=1}^{m}\sum_{s'=1}^{m} b_{ss'} = \sum_{j=1}^{p} pn$$

On conclut que la somme des éléments de B est np^2.

3.3.3 Matrice des Fréquences

Remarquant que la somme des éléments de chaque ligne de X est égale à p. D'autre part, la somme colonne correspondent au nombre d'individus possédant la modalité j est égale à l'effectif marginal de la modalité correspondante noté n_j. Autrement dit, le total général vaut np.

On en déduit de ce tableau disjonctif complet une matrice de fréquence $f_{ij} = \frac{x_{ij}}{np}$.

$F=$

	1	$\cdots$	j	$\cdots$	m	
1	$\frac{x_{11}}{np}$	$\cdots$	$\frac{x_{1j}}{np}$	$\cdots$	$\frac{x_{1m}}{np}$	$\frac{1}{n}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
i	$\frac{x_{i1}}{np}$	$\cdots$	$\frac{x_{ij}}{np}$	$\cdots$	$\frac{x_{im}}{np}$	$\frac{1}{n}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
n	$\frac{x_{n1}}{np}$	$\cdots$	$\frac{x_{nj}}{np}$	$\cdots$	$\frac{x_{nm}}{np}$	$\frac{1}{n}$
	$\frac{n_1}{np}$	$\cdots$	$\frac{n_j}{np}$	$\cdots$	$\frac{n_m}{np}$	

TABLE 3.4 – Mise en fréquences du tableau disjonctif complet.

On admet que le poids de l'individu i est

$$f_{i+} = \sum_{j=1}^{m} f_{ij} = \sum_{j=1}^{m} \frac{x_{ij}}{np} = \frac{1}{np} \sum_{j=1}^{m} x_{ij} = \frac{1}{n} \text{ (constant)}$$

De même manière pour la modalité j

$$f_{+j} = \sum_{i=1}^{n} f_{ij} = \sum_{i=1}^{n} \frac{x_{ij}}{np} = \frac{1}{np} \sum_{i=1}^{n} x_{ij} = \frac{n_j}{np}$$

Comme dans AFC, notons $D_r = diag(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n})$ et $D_m = diag(\frac{n_1}{np}, \frac{n_2}{np}, \cdots, \frac{n_m}{np})$

3.4 AFC sur le tableau disjonctif complet

En effet,le tableau des profils-lignes est $l_{ij} = \frac{f_{ij}}{f_{i+}} = \frac{\frac{x_{ij}}{np}}{\frac{1}{n}} = \frac{x_{ij}}{p}$

$L=$

	1	$\cdots$	j	$\cdots$	m
1	$\frac{x_{11}}{p}$	$\cdots$	$\frac{x_{1j}}{p}$	$\cdots$	$\frac{x_{1m}}{p}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
i	$\frac{x_{i1}}{p}$	$\cdots$	$\frac{x_{ij}}{p}$	$\cdots$	$\frac{x_{im}}{p}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
n	$\frac{x_{n1}}{p}$	$\cdots$	$\frac{x_{nj}}{p}$	$\cdots$	$\frac{x_{nm}}{p}$

$= \frac{1}{p} X$

TABLE 3.5 – Représentation des Profls-lignes

Par raison de symétrie, la matrice des profils-colonnes est $c_{ij} = \frac{f_{ij}}{f_{+j}} = \frac{\frac{x_{ij}}{np}}{\frac{n_j}{np}} = \frac{x_{ij}}{n_j}$

$C=$

	1	$\cdots$	j	$\cdots$	m
1	$\frac{x_{11}}{n_1}$	$\cdots$	$\frac{x_{1j}}{n_j}$	$\cdots$	$\frac{x_{1m}}{n_m}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
i	$\frac{x_{i1}}{n_1}$	$\cdots$	$\frac{x_{ij}}{n_j}$	$\cdots$	$\frac{x_{im}}{n_m}$
$\vdots$	$\vdots$		$\vdots$		$\vdots$
n	$\frac{x_{n1}}{n_1}$	$\cdots$	$\frac{x_{nj}}{n_j}$	$\cdots$	$\frac{x_{nm}}{n_m}$

TABLE 3.6 – Représentation des Profls-colonnes

Précédemment, les effectifs marginaux de la variable X^j est D_j. Le tableau des profils colonnes est donc XD^{-1} où D est la matrice diagonale par blocs $D = diag(D_1, D_2, \cdots, D_p)$

3.5 Représentation Géométrique

3.5.1 Etude des Individus

En ACM on utilise la distance du χ^2 pour comparer deux individus décrits par deux points de $\mathbb{R}^m$.

Distance du χ^2 entre deux individus i et i'

$$d^2(i,i') = \sum_{j=1}^{m} \frac{1}{f_{+j}} \left(\frac{x_{ij}}{p} - \frac{x_{i'j}}{p} \right)^2 = \sum_{j=1}^{m} \frac{np}{n_j} \left(\frac{x_{ij}}{p} - \frac{x_{i'j}}{p} \right)^2 \text{ (métrique } D_m^{-1})$$

Par suite,

$$d^2(i,i') = \frac{n}{p} \sum_{j=1}^{m} \frac{1}{n_j} (x_{ij} - x_{i'j})^2$$

On admet que deux individus sont proches s'ils possèdent les mêmes modalités, sachant que l'on donne plus de "poids" dans cette distance au fait que ces deux individus ont en commun une modalité rare (n_j petit).

Centre de gravité G_l

$$G_l = \begin{pmatrix} g_1 \\ \vdots \\ g_j \\ \vdots \\ g_m \end{pmatrix} \text{ où } g_j = \sum_{i=1}^{n} f_{i+} l_{ij} = \sum_{i=1}^{n} \frac{1}{n} \frac{x_{ij}}{p} = \frac{n_j}{np}$$

Donc,

$$G_l = \begin{pmatrix} \frac{n_1}{np} \\ \vdots \\ \frac{n_j}{np} \\ \vdots \\ \frac{n_m}{np} \end{pmatrix}$$

Propriétés 8

$$I(L) = \frac{m}{p} - 1 \tag{3.1}$$

Preuve : Par définition,

$$I(L) = \sum_{i=1}^{n} f_{i+} d^2(i, G_l)$$

D'une part,

$$d^2(i, G_l) = \sum_{j=1}^{m} \frac{1}{f_{+j}} \left(\frac{x_{ij}}{p} - \frac{n_j}{np} \right)^2$$

Par suite,

$$d^2(i, G_l) = \sum_{j=1}^{m} \frac{np}{n_j} \left(\frac{x_{ij}}{p} - \frac{n_j}{np} \right)^2$$

$$d^2(i, G_l) = \sum_{j=1}^{m} \frac{1}{npn_j} \left(nx_{ij} - n_j \right)^2$$

D'aprés l'identité remarquable,

$$d^2(i, G_l) = \frac{1}{np} \left[\sum_{j=1}^{m} \frac{n^2 x_{ij}^2}{n_j} + \sum_{j=1}^{m} \frac{n_j^2}{n_j} - 2 \sum_{j=1}^{m} \frac{n x_{ij} n_j}{n_j} \right]$$

En simplifiant par n_j,

$$d^2(i, G_l) = \frac{1}{np} \left[\sum_{j=1}^{m} \frac{n^2 x_{ij}^2}{n_j} + \underbrace{\sum_{j=1}^{m} n_j}_{=np} - 2n \underbrace{\sum_{j=1}^{m} x_{ij}}_{=p} \right] \quad or\ x_{ij}^2 = x_{ij}$$

En effet,

$$d^2(i, G_l) = \frac{n}{p} \sum_{j=1}^{m} \frac{x_{ij}}{n_j} - 1$$

Donc,

$$I(L) = \sum_{i=1}^{n} \frac{1}{n} \left[\frac{n}{p} \sum_{j=1}^{m} \frac{x_{ij}}{n_j} - 1 \right]$$

Autrement dit,,

$$I(L) = \sum_{i=1}^{n} \frac{1}{p} \sum_{j=1}^{m} \frac{x_{ij}}{n_j} - \sum_{i=1}^{n} \frac{1}{n}$$

Par conséquent,

$$I(L) = \frac{1}{p}\sum_{j=1}^{m}\underbrace{\sum_{i=1}^{n}\frac{x_{ij}}{n_j}}_{=1} - 1$$

D'où,

$$I(L) = \frac{m}{p} - 1 \tag{3.2}$$

3.5.2 Etude des modalités

De même manière que l'étude sur les individus mais l'étude se fait sur $\mathbb{R}^n$. En utilisant la distance χ^2 avec la métrique D_r^{-1}.

Distance du χ^2 entre deux modalités j et j'

$$d^2(j,j') = \sum_{i=1}^{n}\frac{1}{f_{i+}}\left(\frac{x_{ij}}{n_j} - \frac{x_{ij'}}{n_{j'}}\right)^2$$

C'est-à-dire,

$$d^2(j,j') = n\sum_{i=1}^{n}\left(\frac{x_{ij}}{n_j} - \frac{x_{ij'}}{n_{j'}}\right)^2$$

On admet que deux modalités sont proches si elles sont possédées par les mêmes individus.

Centre de gravité G_c

$$G_c = \begin{pmatrix} g_1 \\ \vdots \\ g_i \\ \vdots \\ g_n \end{pmatrix} \text{ où } g_i = \sum_{j=1}^{m} f_{+j}c_{ij} = \sum_{j=1}^{m}\frac{n_j}{np}\frac{x_{ij}}{n_j} = \frac{1}{np}\sum_{j=1}^{m}x_{ij} = \frac{1}{n}$$

Donc,

$$G_c = \begin{pmatrix} \frac{1}{n} \\ \vdots \\ \frac{1}{n} \\ \vdots \\ \frac{1}{n} \end{pmatrix}$$

Inertie d'une modalité j

$$I(j) = f_{+j}d^2(j,G_l)$$

D'autre part,

$$d^2(j,G_l) = n\sum_{i=1}^{n}\left(\frac{x_{ij}}{n_j} - \frac{1}{n}\right)^2$$

Par suite,

$$d^2(j,G_l) = n\sum_{i=1}^{n}\left(\frac{x_{ij}^2}{n_j^2} - 2\frac{1}{n}\frac{x_{ij}}{n_j} + \frac{1}{n^2}\right)$$

En distribuant la somme,

$$d^2(j, G_l) = n \sum_{i=1}^{n} \frac{x_{ij}^2}{n_j^2} - 2 \underbrace{\sum_{i=1}^{n} \frac{x_{ij}}{n_j}}_{=1} + \underbrace{\sum_{i=1}^{n} \frac{1}{n}}_{=1} \text{ or } x_{ij}^2 = x_{ij}$$

En effet,

$$d^2(j, G_l) = \frac{n}{n_j^2} \underbrace{\sum_{i=1}^{n} x_{ij}}_{=n_j} - 1$$

Donc,

$$d^2(j, G_l) = \frac{n}{n_j} - 1$$

Il s'ensuit que

$$I(j) = \frac{n_j}{np} \left(\frac{n}{n_j} - 1 \right)$$

D'où

$$I(j) = \frac{1}{p} - \frac{n_j}{np} \tag{3.3}$$

Inertie d'une variable X_j

Une variable X_j est l'ensemble des modalités avec $card(X_j) = m_j$ Par conséquent,

$$\begin{aligned} I(X_j) &= \sum_{k=1}^{m_j} I(k) \\ &= \sum_{k=1}^{m_j} \left(\frac{1}{p} - \frac{n_k}{np} \right) \\ &= \sum_{k=1}^{m_j} \frac{1}{p} - \sum_{k=1}^{m_j} \frac{n_k}{np} \\ &= \frac{m_j}{p} - \frac{1}{np} \underbrace{\sum_{k=1}^{m_j} n_k}_{=n} \end{aligned}$$

$$\textit{On conclut que, } I(X_j) = \frac{m_j - 1}{p} \tag{3.4}$$

Propriétés 9

$$I(C) = I(L)$$

Preuve : Par définition $I(C)$ est l'inertie total des variables
Alors,

$$I(C) = \sum_{j=1}^{p} I(X_j)$$

Autrement dit,

$$I(C) = \sum_{j=1}^{p} \frac{m_j - 1}{p}$$

Par suite,

$$I(C) = \frac{1}{p}\underbrace{\sum_{j=1}^{p} m_j}_{=m} - \frac{1}{p}\underbrace{\sum_{j=1}^{p} 1}_{=p}$$

On en déduit que

$$I(C) = \frac{m}{p} - 1$$

D'où,

$$I(C) = I(L)$$

3.5.3 Coordonnnées factorielles des individus et des modalités

Effectuer une ACM consiste à appliquer l'AFC au TDC , c'est-à-dire, effectuer une ACP pondérée des nuages des point-individus et des point-modalités (centrés). On reprend donc les résultats du cours d'AFC.

3.5.4 Facteurs principaux

Les facteurs principaux sont les vecteurs propres de

$$(XD^{-1})'(\frac{1}{p}X) = \frac{1}{p}D^{-1}X'X = \frac{1}{p}D^{-1}B$$

et on a donc pour chaque axe principal k

$$\frac{1}{p}D^{-1}Bv_k = \mu_k v_k$$

Remarque 12 *Si* $n > \sum_{j=1}^{p} m_j$, *le rang de* X *est* $\sum_{j=1}^{p} m_j - p + 1$ *et le nombre de valeurs propres non trivialement égales à 0 ou 1 est* $\sum_{j=1}^{p} m_j - p$.

3.5.5 Composantes principales

Soit c^k le vecteur à n composantes des coordonnées des n individus sur l'axe factoriel associé à la valeur propre μ_k.

D'après les résultats sur l'AFC, on a

$$c^k = \frac{1}{\sqrt{\mu}_k}\frac{1}{p}Xv_k$$

En effet,

$$c^k(i) = \frac{1}{\sqrt{\mu}_k}\frac{1}{p}\sum_{j=1}^{m} x_{ij}v_k(i)$$

Autrement dit,

$$c^k(i) = \frac{1}{\sqrt{\mu}_k}\frac{1}{p}\sum_{j\in M_i} v_k(i)$$

où M_i est l'ensemble des modalités prises par l'individu i.

3.5.6 Relations barycentriques

$v_k = \frac{1}{\sqrt{\mu_k}} D^{-1} X c^k$,c'est-à-dire, $v_k(j) = \frac{1}{\sqrt{\mu_k}} \frac{1}{n_j} \sum_{i=1}^{n} x_{ij} c_j^k$
Il s'ensuit que

$$v_k(j) = \frac{1}{\sqrt{\mu}_k} \frac{1}{n_j} \sum_{i \in A_j} c_j^k$$

où A_j est l'ensemble des individus qui possèdent la modalité j.

3.6 Interprétation

3.6.1 Nombre d'axes à retenir

On garde les axes tels que les valeurs propres sont supérieures à la moyenne des valeurs propres. Autrement dit, $\mu_k > \frac{1}{p}$.

3.6.2 Contribution et Qualité

On définit la contribution et la Qualité de la représentation comme dans AFC tel que le poids d'une modalité devienne $\frac{n_j}{np}$.

3.7 Exemple

On interroge $n = 10$ individus son état civil, on observe qu'on a trois variables qualitatives :

	Age du client	Situation familiale	Profession
ind1	plus de 50 ans	célibataire	employé
ind2	moins de 23 ans	célibataire	employé
ind3	de 23 à 40 ans	veuf	employé
ind4	de 23 à 40 ans	divorcé	employé
ind5	moins de 23 ans	célibataire	employé
ind6	de 23 à 40 ans	célibataire	employé
ind7	plus de 50 ans	marié	cadre
ind8	plus de 50 ans	marié	cadre
ind9	de 40 à 50 ans	célibataire	employé
ind10	plus de 50 ans	célibataire	employé

Tableau disjoint complet

Posons :
Age1=moins de 23 ans, Age2=de 23 à 40 ans, Age3=de 40 à 50 ans, Age4=plus de 50 ans
S1=célibataire, S2=veuf, S3=divorcé, S4=marié
Prof1=employé, Prof2=cadre

	Age1	Age2	Age3	Age4	S1	S2	S3	S4	Prof1	Prof2
ind1	0	0	0	1	1	0	0	0	1	0
ind2	1	0	0	0	1	0	0	0	1	0
ind3	0	1	0	0	0	1	0	0	1	0
ind4	0	1	0	0	0	0	1	0	1	0
ind5	1	0	0	0	1	0	0	0	1	0
ind6	0	1	0	0	1	0	0	0	1	0
ind7	0	0	0	1	0	0	0	1	0	1
ind8	0	0	0	1	0	0	0	1	0	1
ind9	0	0	1	0	1	0	0	0	1	0
ind10	0	0	0	1	1	0	0	0	1	0

Tableau de Burt

	Age1	Age2	Age3	Age4	S1	S2	S3	S4	Prof1	Prof2
Age1	2	0	0	0	2	2	1	1	2	3
Age2	0	3	0	0	3	2	0	0	2	2
Age3	0	0	1	0	1	2	0	0	1	1
Age4	0	0	0	4	1	2	0	0	1	1
S1	2	3	1	1	6	0	0	0	2	2
S2	2	2	2	2	0	1	0	0	2	2
S3	1	0	0	0	0	0	1	0	0	1
S4	1	0	0	0	0	0	0	2	0	1
Prof1	2	2	1	1	2	2	0	0	8	0
Prof2	3	2	1	1	2	2	1	1	0	2

Valeurs propres

Axis	Eigenvalue	% of total	Cumulative
1	0.838175	35.922	35.922
2	0.559434	23.976	59.898
3	0.333333	14.286	74.183
4	0.333333	14.286	88.469
5	0.193442	8.2904	96.759
6	0.0756149	3.2406	100
7	5.65549E-33	2.4238E-31	100
8	9.84674E-35	4.22E-33	100
9	5.22893E-35	2.241E-33	100

Axes à conserver

On garde les facteurs dont les valeurs propres sont supérieures à la moyenne des valeurs propres. Autrement dit, $\lambda_k > \frac{1}{p} = \frac{1}{3}$.

D'après le tableau si-dessus, 2 facteurs ont une valeur propre supérieure à la moyenne.

Analyse des individus

	Coordonnées		Contributions		$\cos^2$	
	Axis1	Axis2	Axis1	Axis2	Axis1	Axis2
ind1	-0.0509094	-0.372635	0.0003	0.025	0.018	0.98
ind2	0.51082	-0.81446	0.031	0.119	0.28	0.66
ind3	0.745779	1.26054	0.066	0.284	0.26	0.74
ind4	0.745779	1.26054	0.066	0.284	0.26	0.74
ind5	0.51082	-0.81446	0.031	0.119	0.28	0.66
ind6	0.582633	0.214066	0.041	0.008	0.88	0.12
ind7	-1.75242	0.226747	0.366	0.009	0.98	0.016
ind8	-1.75242	0.226747	0.366	0.009	0.98	0.016
ind9	0.51082	-0.81446	0.031	0.119	0.28	0.66
ind10	-0.0509094	-0.372635	0.0003	0.025	0.018	0.98

Analyses des modalités

	Coordonnées		Contributions		$\cos^2$	
	Axis1	Axis2	Axis1	Axis2	Axis1	Axis2
Age1	0.557957	-1.08892	0.025	0.14	0.21	0.79
Age2	0.755197	1.21895	0.068	0.27	0.28	0.72
Age3	0.557957	-1.08892	0.012	0.07	0.21	0.79
Age4	-0.984865	-0.0975246	0.15	0.002	0.99	0.01
S1	0.366509	-0.662827	0.032	0.16	0.23	0.77
S2	0.814597	1.68532	0.026	0.17	0.19	0.81
S3	0.814597	1.68532	0.026	0.17	0.19	0.81
S4	-1.91412	0.303157	0.29	0.011	0.98	0.024
Prof1	0.478531	-0.0757892	0.073	0.003	0.98	0.02
Prof2	-1.91412	0.303157	0.29	0.011	0.98	0.024

Nous interprèterons seulement les deux premières dimensions : c'est suffisant ici. De plus, l'interpétation de toute autre dimension se fait selon le même principe. Le principe général est de repérer les modalités ayant des contributions importantes aux axes et de regarder ensuite leur positionnement sur le graphique.

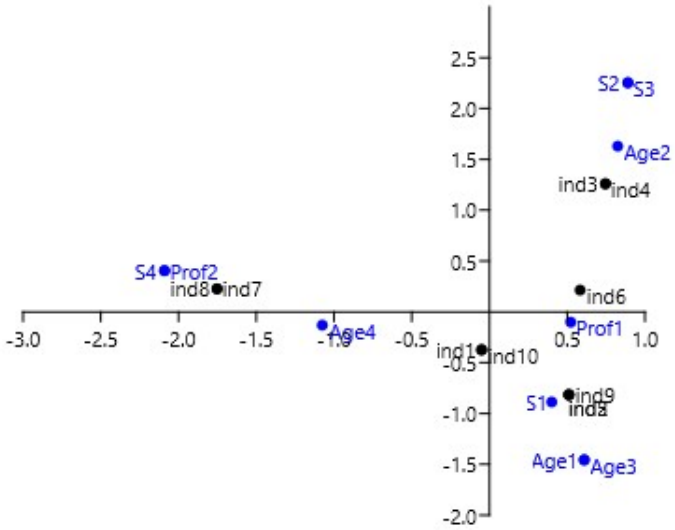
Sur l'axe 1, ces contributions sont celles du marié (pratiquement 29%), plus de 50 ans (près de 15%) et cadre (29%). Les mariés âgés plus de 50 ans sont le plus souvent titulaires sur la cadre.

Sur l'axe 2, les contributions les plus importantes sont celles du célibataire (16%), et des veuf (17%), et divorcé (17%), âgés moins de 23 ans (14%) et de 23 à 40 ans(27%). En observant le graphique, on voit que l'axe 2 discrimine veuf, divorcé, Âgés de 23 à 40 ans en haut et du célibataire,Âgés moins de 23 ans en bas.

De manière similaire pour les individus. Sur l'axe 1, on s'intéresse au individu 7 et 8 (plus de 36%).

Sur l'axe 2, les individus qui ont les plus importances sont les individus 2, 5 et 9 (plus de 11%), les individus 3 et 4 (plus de 28%). En observant le graphique, on voit que l'axe 2 discrimine les individus 3, 4 en haut et les individus 2, 5 et 9 en bas.

FIGURE 3.1 – Représentation Graphique sur les deux axes factoriels



Chapitre 4

Analyse Factorielle Discriminante

L'analyse factorielle discriminante (AFD) ou simplement analyse discriminante est une méthode géométrique et essentiellement descriptive qui vise à décrire, expliquer et prédire l'appartenance à des groupes prédéfinis (classes) d'un ensemble d'observations à partir d'une série de variables prédictives. Comme dans l'ACP et les autres méthodes factorielles on cherche un espace dans lequel on va projeter le nuage de points ; ici on veut mettre en évidence les groupes, autrement dit préserver les distances à l'intérieur des groupes et entre les centres de gravité des groupes.

4.1 Données et Notations

On se place dans le cadre de la modélisation d'une variable Y qualitative à m modalités $Y_1, Y_2, \cdots, Y_m$ à partir de p variables explicatives $X_1, X_2, \cdots, X_p$ quantitatives. On suppose que l'on dispose d'un échantillon de taille n pour lequel les p variables explicatives et la variable à expliquer ont été mesurées simultanément. Chaque individu est affecté d'un poids p_i.

En notant x_{ij} la valeur de la $j$ème variable explicative mesurée sur le $i$ème individu, on obtient ainsi la matrice de données de dimension $n \times p$ suivante :

		1	$\cdots$	j	$\cdots$	p
X=	1					
	$\vdots$			$\vdots$		
	i		$\cdots$	x_{ij}	$\cdots$	
	$\vdots$			$\vdots$		
	n					

TABLE 4.1 – Tableaux des données

Notons :
— $x_i = (x_{i1}, \cdots, x_{ip})' \in \mathbb{R}^p$ décrivant le $i$ème individu.
— De même $x^j = (x_{1j}, \cdots, x_{nj})' \in \mathbb{R}^n$celle de la $j$ème variable.
— G_k est le groupe des individus possédant la modalité k.
— $n_k = card(G_k)$ est le nombre d'individus qui possédant la modalité k.
— $q_k = \sum_{i \in G_k} p_i$ représente le poids relatif de la classe k.
Nous remarquons ainsi que $q_1 + q_2 + \cdots + q_m = 1$

— $g_k \in \mathbb{R}^p$ le centre de gravité du groupe G_k

Définition 19 *La variance interclasse*[1] *B est estimée par la variance empirique des m centres de gravité.*

$$B = \sum_{k=1}^{m} q_k (g_k - g)(g_k - g)' \tag{4.1}$$

Définition 20 *La variance intraclasse*[2] *W est la dispersion à l'intérieur d'un groupe*

$$W = \sum_{k=1}^{m} q_k V_k \tag{4.2}$$

où $V_k = \dfrac{1}{q_k} \sum_{i \in G_k} p_i (x_i - g_k)(x_i - g_k)'$ *la variance-covariance du groupe G_k .*

Proposition 7 (Formule de décomposition de Huygens) *La variance totale V se décompose*

$$V = B + W \tag{4.3}$$

Preuve Par définition

$$V = \sum_{i=1}^{n} p_i (x_i - g)(x_i - g)'$$

Autrement dit,

$$V = \sum_{k=1}^{m} \underbrace{\sum_{i \in G_k} p_i (x_i - g)(x_i - g)'}_{SS(k)\ sum\ of\ squares}$$

On a la décomposition suivante :

$$\begin{aligned} SS(k) &= \sum_{i \in G_k} p_i (x_i - g)(x_i - g)' \\ &= \sum_{i \in G_k} p_i \left((x_i - g_k) + (g_k - g)\right)\left((x_i - g_k) + (g_k - g)\right)' \end{aligned}$$

Par suite,

$$SS(k) = \sum_{i \in G_k} p_i \left((x_i - g_k)(x_i - g_k)' + (g_k - g)(g_k - g)'\right)$$

Car

$$\sum_{i \in G_k} (x_i - g_k) = 0$$

En revanche,

$$\begin{aligned} SS(k) &= \sum_{i \in G_k} p_i (x_i - g_k)(x_i - g_k)' + \sum_{i \in G_k} p_i (g_k - g)(g_k - g)' \\ &= \sum_{i \in G_k} p_i (x_i - g_k)(x_i - g_k)' + q_k (g_k - g)(g_k - g)' \end{aligned}$$

1. En anglais : variance between
2. En anglais : variance within

Par conséquent,

$$V = \sum_{k=1}^{m} \sum_{i \in G_k} p_i (x_i - g_k)(x_i - g_k)' + \sum_{k=1}^{m} q_k (g_k - g)(g_k - g)'$$

On conclut que,

$$V = W + B$$

Cette décomposition est illustrée par les schémas ci-dessous :

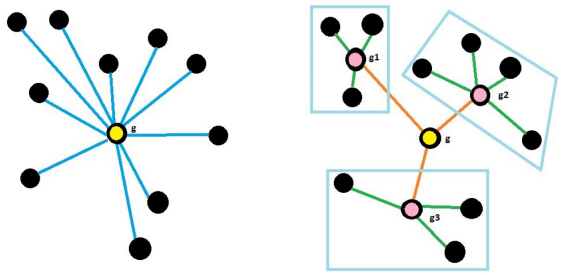


FIGURE 4.1 – Représentation graphique de la décomposition de Huygens

4.2 Axes, facteurs et variables discriminantes

L'Analyse Factorielle Discriminante (AFD) consiste à rechercher de nouvelles variables (les variables discriminantes) correspondant à des directions de $\mathbb{R}^p$ qui séparent le mieux possible en projection les K groupes d'individus.

Définition 21 *On muni $\mathbb{R}^p$ d'une métrique M et on projette les n points de $\mathbb{R}^p$ sur un axe de vecteur directeur a. On effectue des projections M-orthogonales et a est M-normé.*
La projection de x_i sur l'axe est donc $s_i = x_i' Ma$, forme la nouvelle variable s appelé la variable discriminante telle que :

$$s = (s_1, s_2, \cdots, s_n)' = XMa \tag{4.4}$$

où

- *$a \in \mathbb{R}^p$ est appelé l'axe discriminant ($a'Ma = 1$).*
- *$u = Ma \in \mathbb{R}^p$ est appelé le facteur discriminant ($u'M^{-1}u = 1$).*

La variance de la variable s est définie par :

$$Var(s) = \frac{1}{n} \sum_{i=1}^{n} (s_i - \bar{s})^2 \tag{4.5}$$

or $s_i = \sum_j^p x_{ij} u_j$
alors $\bar{s} = \frac{1}{n} \sum_{i=1}^n \sum_j^p x_{ij} u_j = \sum_j^p u_j \frac{1}{n} \sum_{i=1}^n x_{ij}$
autrement dit $\bar{s} = \sum_{j=1}^p u_j g_j$
En effet

$$s_i - \bar{s} = \sum_{j=1}^{p} u_j (x_{ij} - g_j)$$

Donc

$$Var(s) = \frac{1}{n} \sum_{i=1}^{n} \left[\sum_{j=1}^{p} u_j (x_{ij} - g_j) \right]^2$$

Il s'ensuit que

$$Var(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} u_j u_k (x_{ij} - g_j)(x_{ik} - g_k)$$

Donc

$$Var(s) = \sum_{j=1}^{p}\sum_{k=1}^{p} u_j u_k \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - g_j)(x_{ik} - g_k) = \sum_{j=1}^{p}\sum_{k=1}^{p} u_j u_k v_{jk}$$

On conclut que

$$Var(s) = u^t V u \tag{4.6}$$

D'après l'équation (4.3), nous avons :

$$u^t V u = u^t B u + u^t W u \tag{4.7}$$

La quantité $u^t V u$ étant indépendante des groupes, minimiser $u^t W u$ est équivalent à maximiser $u^t B u$. On veut maximiser le rapport entre la variance inter-groupe et la variance totale.

Le critère à maximiser,

$$\frac{u^t B u}{u^t V u} \in [0, 1] \tag{4.8}$$

Remarque 13 *La relation (4.8) est appelée le rapport de corrélation entre s et la variable à expliquer.*

Il est encore équivalent de chercher le maximum de la forme quadratique $u^t B u$ sous la contrainte quadratique $u^t V u = 1$.

On utilise la méthode des multiplicateurs de Lagrange :
Posons

$$f(u, \lambda) = u^t B u - \lambda(u^t V u - 1)$$

$$\frac{\partial f(u, \lambda)}{\partial u} = 2Bu - 2\lambda V u$$

Condition nécessaire du 1$^{\text{er}}$ ordre :

$$\frac{\partial f(u, \lambda)}{\partial u} = 2Bu - 2\lambda V u = 0$$

Par suite,

$$Bu = \lambda V u \tag{4.9}$$

Lorsque V est une matrice inversible, nous obtenons :

$$V^{-1} B u = \lambda u \tag{4.10}$$

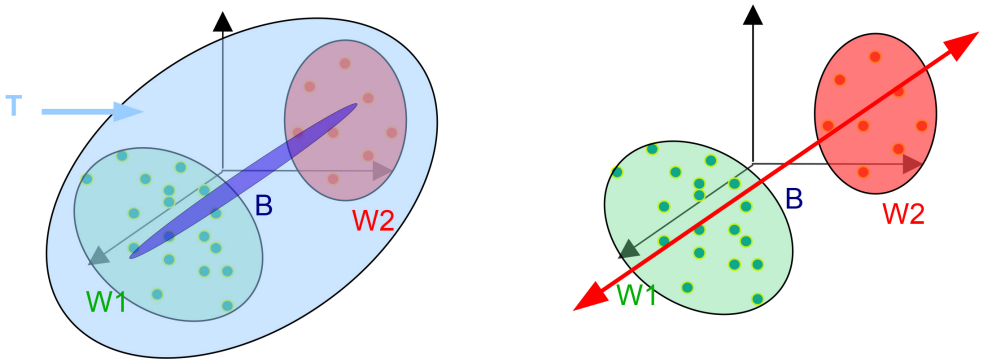
Il s'ensuit que u_1 est le vecteur propre de $V^{-1}B$ associé à la plus grande valeur propre λ_1.

La décomposition spéctrale de $V^{-1}B$ donne aussi les autres axes discriminants de l'AFD. En effet, on cherche ensuite $s_2 = Xu_2$ non corrélée à s_1, telle que $\frac{u_2^t B u_2}{u_2^t V u_2}$ soit maximum et ainsi de suite. On pourra construire $m-1$ axes discriminants car le rang de $V^{-1}B$ est au plus $\min(p, m-1)$ et on suppose que $m - 1 < p < n$.

4.3 Exemple (Cas de deux groupes)

Dans le cas de deux groupes, on parle de fonction discriminante, fonction décrivant l'axe unique séparant les deux groupes et qui passe par leurs deux centroïdes car $\min(p, 2-1) = 1$.

FIGURE 4.2 – Cas de deux groupes



Le facteur discriminant vaut alors

$$u = W^{-1}(\bar{x}_1 - \bar{x}_2)$$

et la variable discriminante vaut Xu.
Pour l'individu i de l'échantillon initial, cette variable prend la valeur

$$(Xu)_i = x_i^T W^{-1}(\bar{x}_1 - \bar{x}_2) = d_{Mahalanobis}(x_i, \bar{x}_1 - \bar{x}_2)$$

La variable discriminante s'obtient donc en projetant les observations sur l'axe reliant les deux centres de gravité pour la métrique de Mahalanobis.

Appliquons sur le tableau ci-dessous :

Individu	A	B	Classe
1	-6	4	1
2	-2	1	1
3	2	-2	1
4	-2	2	2
5	2	-4	2
6	6	-1	2

Le centre de gravité global

$$g = \sum_{i=1}^{6} p_i x_i = \frac{1}{6}\sum_{i=1}^{6} x_i = \frac{1}{6}\begin{pmatrix} -6-2+2-2+2+6 \\ 4+1-2+2-4-1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

les données sont centrées. On en déduit que, la matrice 2×2 de variance-covariance globale est

$$V = \frac{1}{6}X'X = \begin{pmatrix} -6 & -2 & 2 & -2 & 2 & 6 \\ 4 & 1 & -2 & 2 & -4 & -1 \end{pmatrix} \begin{pmatrix} -6 & 4 \\ -2 & 1 \\ 2 & -2 \\ -2 & 2 \\ 2 & -4 \\ 6 & -1 \end{pmatrix} = \begin{pmatrix} \frac{44}{3} & -8 \\ -8 & 7 \end{pmatrix}$$

Ainsi,la matrice 2×2 de variance-covariance inter-groupe est

$$B = \sum_{k=1}^{2} \frac{n_k}{n} g_k g_k'$$

$$g_1 = \frac{1}{3}\sum_{i=1}^{3} x_i = \frac{1}{3}\begin{pmatrix} -6-2+2 \\ 4+1-2 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$g_2 = \frac{1}{3}\sum_{i=3}^{6} x_i = \frac{1}{3}\begin{pmatrix} -2+2+6 \\ 2-4-1 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$$

Par suite,

$$B = \frac{1}{2}\begin{pmatrix} -2 \\ 1 \end{pmatrix}\begin{pmatrix} -2 & 1 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 2 \\ -1 \end{pmatrix}\begin{pmatrix} 2 & -1 \end{pmatrix} = \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}$$

Par conséquent, la matrice 2×2 de variance-covariance intra-groupe est

$$W = V - B = \begin{pmatrix} \frac{32}{3} & -6 \\ -6 & 6 \end{pmatrix}$$

Calculons le détérminant de W, $det(W) = 64 - 36 = 28$

En effet, W est inversible tel que $W^{-1} = \frac{1}{28}\begin{pmatrix} 6 & 6 \\ 6 & \frac{32}{3} \end{pmatrix} = \begin{pmatrix} \frac{3}{14} & \frac{3}{14} \\ \frac{3}{14} & \frac{8}{21} \end{pmatrix}$

Donc, le facteur discriminant est

$$u = W^{-1}(g_1 - g_2) = \begin{pmatrix} \frac{3}{14} & \frac{3}{14} \\ \frac{3}{14} & \frac{8}{21} \end{pmatrix}\left[\begin{pmatrix} -2 \\ 1 \end{pmatrix} - \begin{pmatrix} 2 \\ -1 \end{pmatrix}\right] = \begin{pmatrix} \frac{-3}{7} \\ \frac{-2}{21} \end{pmatrix}$$

D'où,la variable discriminante $s = Xu = \begin{pmatrix} -6 & 4 \\ -2 & 1 \\ 2 & -2 \\ -2 & 2 \\ 2 & -4 \\ 6 & -1 \end{pmatrix}\begin{pmatrix} \frac{-3}{7} \\ \frac{-2}{21} \end{pmatrix} = \frac{1}{21}\begin{pmatrix} 46 \\ 16 \\ -14 \\ 14 \\ -10 \\ -52 \end{pmatrix}$

Remarque 14 — *L'AFD peut être vue comme une ACP de centre de gravité g_k avec la métrique V^{-1}.*

— *L'AFD est aussi un cas particulier de l'analyse canonique. Elle correspond en effet à l'analyse canonique des tableaux A et X où A (de dimension $n \times m$) représente l'ensemble des variables indicatrices associées à la variable qualitative à expliquer.*

Chapitre 5

Classification, segmentation

5.1 Introduction

La classification des individus est le domaine de la classification automatique et de l'analyse discriminante. Parmi les méthodes de statistique exploratoire multidimensionnelle, dont l'objectif est d'extraire d'une masse de données des " informations utiles ", on distingue les méthodes d'analyse factorielle des méthodes de classification automatique. Classifier consiste à définir des classes, classer est l'opération permettant de mettre un objet dans une classe définie au préalable.

5.2 Les objectifs

L'objectif de la classification ou segmentation est de former des groupes d'individus ou de variables afin de structurer un ensemble de données. C'est la recherche d'une typologie, ou segmentation, c'est-à-dire, d'une partition ou répartition des individus en classes ou catégories. Autrement dit, l'objectif est de regrouper ou classer les individus qui se ressemblent le plus ou qui ont des caractéristiques semblables.

5.3 Tableau de données

On considère un ensemble $\Omega = \{1, ..., i, ..., n\}$ de n individus décrits par p variables $X^1, ..., X^p$ dans une matrice X de n lignes et p colonnes.

$$X = \left(x_i^j\right)_{n\times p} = \begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^p \\ x_2^1 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \ddots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^p \end{pmatrix}$$

TABLE 5.1 – Matrice des données à classifier

Un individu $i \in \Omega$ est donc décrit par un vecteur $x_i \in \mathbb{R}^p$ (ligne i de X). Un poids p_i est associé à chaque individu i.

5.4 Mesures d'éloignement

La ressemblance (similarité/dissimilarité) des individus est mesurée par un indice de similarité, un indice de dissimilarité ou une distance

Définition 22 *Un indice de similarité est une mesure* $s : \Omega \times \Omega \to \mathbb{R}^+$ *telle que pour tout* $i, k \in \Omega$ *on a :*

$$\begin{aligned} s(i,k) &\geq 0 \\ s(i,k) &= s(k,i) \\ s(i,i) = smax &\geq s(i,k) \end{aligned} \tag{5.1}$$

Si $smax = 1$*, alors* s *est une similarité normalisée.*

Définition 23 *Une dissimilarité est une application* $d : \Omega \times \Omega \to \mathbb{R}^+$ *est telle que pour tout* $i, k \in \Omega$ *on a :*

$$\begin{aligned} d(i,i) &= 0 \\ d(i,k) &= d(k,i) \end{aligned} \tag{5.2}$$

Définition 24 *Une distance est une dissimilarité qui vérifie en plus l'inégalité triangulaire : pour tout* $i, i', k \in \Omega$

$$d(i,i') \leq d(i,k) + d(k,i') \tag{5.3}$$

Remarque 15 *Les notions de similarité* s *et dissimilarité* d *se correspondent de façon élémentaire. Pour tout* $i, k \in \Omega$

$$d(i,k) = smax - s(i,k)$$

La mesure de ressemblance utilisée varie en fonction du type des données c'est à dire du type des variables : tableau quantitatif, qualitatif, ou mixte. Quand les données sont quantitatives, la distances classique est définie par :

$$d_M(x_1, x_2) = \sqrt{(x_1 - x_2)' M (x_1 - x_2)} = \|x_1 - x_2\|_M \tag{5.4}$$

Mesures de proximité dans le cas des données numériques

Une mesure générale de la distance dans le cas des données numériques est donnée par l'indice de Minkowski défini de la manière suivante :

$$\text{Pour tout } x, y \in \Omega^p, d(x,y) = \left[\sum_{i=1}^{p} \mid x_i - y_i \mid^q \right]^{\frac{1}{q}} \tag{5.5}$$

Définition 25 *Une partition* P *en* K *classes de* Ω *est un ensemble* $(P_1, P_2, \cdots, P_K)$ *de classes non vides, d'intersections disjointes et dont leur réunion forme* Ω *:*

- *Pour tout* $i \in \{1, 2, \cdots, K\}, P_i \neq \emptyset$
- *Pour tout* $i \in \{1, 2, \cdots, K\}, P_i \cap P_j = \emptyset$
- $\bigcup_{j=1}^{K} P_j = \Omega$

Définition 26 *Une hiérarchie* H *de* Ω *est un ensemble de classes non vides (appelés paliers) qui vérifient :*

- $\Omega \in H$

— *Pour tout* $i \in \Omega, \{i\} \in H$ *(la hiérarchie contient tous les singletons)*
— *Pour tout* $A, B \in H, A \cap B \in \{A, B, \emptyset\}$ *(deux classes de la hiérarchie sont soit disjointes soit contenues l'une dans l'autre)*

Définition 27 *La classification automatique est un ensemble des techniques qui fournissent directement une ou plusieurs partitions d'un ensemble ; certaines d'entre elles, dites de classification hiérarchique, permettent d'obtenir des partitions "non contradictoires" qui sont présentées sous forme d'un arbre de classification.*

5.5 La classification hiérarchique

Il existe deux stratégies de construction d'une hiérarchie indicée :
— On construit la hiérarchie en partant du bas de l'arbre (des singletons) et on agrège, deux par deux les classes les plus proches, et ce jusqu'à l'obtention d'une seule classe. On parle de classification ascendante hiérarchique (C.A.H.).
— On construit la hiérarchie à partir du haut de l'arbre en procédant par divisions successives de l'ensemble Ω jusqu'à obtenir des classes réduites à un élément, ou des classes ne contenant que des individus identiques. On parle de classification divisive ou classification descendante hiérarchique (C.D.H.).

Ces deux approches conduisent à une hiérarchie des partitions des éléments. La seconde approche est beaucoup moins employée que la première, nous présentons donc ici la première approche.

Définition 28 *Une hiérarchie binaire est une hiérarchie dont chaque palier est la réunion de deux paliers. Le nombre de paliers non singletons d'une hiérarchie binaire vaut* $n - 1$.

Cette définition d'une hiérarchie est ensembliste. Maintenant, pour pourvoir représenter une hiérarchie par un graphique, il faut pouvoir valuer ses paliers, c'est à dire, leur attribuer une hauteur.

Définition 29 *Une hiérarchie indicée est un couple* (H, h) *où* H *est une hiérarchie et* h *une application de* H *dans* $\mathbb{R}^+$ *telle que :*

$$\begin{aligned} &\textit{Pour tout } A \in H, h(A) = 0 \Leftrightarrow A \textit{ est un singleton} \\ &\textit{Pour tout } A, B \in H, A \neq B, A \subset B \Longrightarrow h(A) \leq h(B) \end{aligned} \tag{5.6}$$

Le graphique représentant une hiérarchie indicée est appelé un dendogramme ou arbre hiérarchique.

5.5.1 Classification ascendante hiérarchique

La classification hiérarchique ascendante est une méthode itérative qui consiste, à chaque étape, à regrouper les classes les plus proches. A la première étape, chaque individu constitue une classe. L'algorithme démarre donc de la partition triviale des n singletons. Et l'algorithme s'arrête avec l'obtention d'une seule classe. Les regroupements successifs sont représentés sous la forme d'un arbre ou dendogramme.

Algorithme

L'algorithme général est le suivant :

Initialisation Les classes initiales sont les singletons $P = (P_1, P_2, \cdots, P_n)$ avec $P_k = \{k\}$. Calculer la matrice de leurs distances deux à deux.

Itérer

1. On part de la partition $P = (P_1, P_2, \cdots, P_k)$ en k classes obtenue à l'étape précédente et on agrège les deux classes P_i et P_j qui minimisent une mesure d'agrégation $D(P_i, P_j)$. On construit ainsi une nouvelle partition en $k-1$ classes. En cas d'égalité, on choisit la première solution rencontrée. La hiérarchie obtenue est donc toujours binaire.
2. On recommence l'étape (1) jusqu'à obtenir la partition la plus grossière, c'est-à-dire, la partition en une seule classe Ω

— **Avantage :** facile à implémenter.
— **Défaut :** méthode très coûteuse, complexité temporelle en $\mathcal{O}(n^2)$.

Mesure d'agrégation

La mesure d'agrégation $D : \mathcal{P}(\Omega) \times \mathcal{P}(\Omega) \longrightarrow \mathbb{R}^+$ qui mesure la ressemblance entre deux classes.

$$D(A, B) = \min_{i \in A, j \in B} d(i, j) \text{ (saut minimum, single linkage)} \tag{5.7}$$

$$D(A, B) = \max_{i \in A, j \in B} d(i, j) \text{ (saut maximum ou diamètre, complete linkage)} \tag{5.8}$$

Agrégation selon l'inertie

On note :
— m_k le poids de P_k, $m_k = \sum_{i \in P_k} p_i$.
— g_k le centre de gravité de P_k, $g_k = \frac{1}{m_k} \sum_{i \in P_k} p_i x_i$.

Définition 30 *L'inertie totale T du nuage des n individus est*

$$T = \sum_{i=1}^{n} p_i d_M^2(x_i, g) \tag{5.9}$$

Définition 31 *L'inertie I_a du nuage des n individus par rapport à un point $a \in \mathbb{R}^p$ est*

$$I_a = \sum_{i=1}^{n} p_i d_M^2(x_i, a) \tag{5.10}$$

Définition 32 *L'inertie inter-classe A de la partition P est*

$$A = \sum_{i=1}^{K} m_i d_M^2(g_i, g) \tag{5.11}$$

A est donc l'inertie du nuage des centres de gravité des K classes, munis des poids m_i

Définition 33 *L'inertie intra-classe W de la partition P est*

$$W = \sum_{i=1}^{K} I(P_i) \tag{5.12}$$

$$\text{où } I(P_i) = \sum_{k \in P_i} p_k d_M^2(x_k, g_i) \tag{5.13}$$

Définition 34 *Le pourcentage d'inertie expliquée d'une partition P est :*

$$\left(1 - \frac{W}{T}\right) \times 100 \tag{5.14}$$

Remarque 16 *La qualité d'une partition est mesurée par :*

$$0 \leq \frac{\textit{Inertie inter-classe}}{\textit{Inertie totale}} \leq 1 \tag{5.15}$$

Décomposition de Huygens : Pour toute partition $\mathcal{P}$ de $\mathcal{N}$, on a

$$T = \sum_{i=1}^{n} p_i d_M^2(x_i, g) = \sum_{k=1}^{K} \sum_{i \in P_k} p_i d_M^2(x_i, g)$$

Par suite,

$$T = \sum_{k=1}^{K} \sum_{i \in P_k} p_i \|x_i - g_{P_k} + g_{P_k} - g\|_M^2$$

Autrement dit,

$$T = \sum_{k=1}^{K} \sum_{i \in P_k} p_i (\|x_i - g_{P_k}\|_M^2 + \|g_{P_k} - g\|_M^2 + 2 < x_i - g_{P_k}, g_{P_k} - g >_M)$$

En effet,

$$T = \sum_{k=1}^{K} \sum_{i \in P_k} p_i \|x_i - g_{P_k}\|_M^2 + \sum_{k=1}^{K} \sum_{i \in P_k} p_i \|g_{P_k} - g\|_M^2$$

Il s'ensuit que

$$T = \sum_{k=1}^{K} \sum_{i \in P_k} p_i \|x_i - g_{P_k}\|_M^2 + \sum_{k=1}^{K} m_k \|g_{P_k} - g\|_M^2$$

On conclut que

$$T = A + W$$

Remarquant qu' à chaque pas, on cherche à obtenir un minimum local de l'inertie intra-classe ou un maximum de l'inertie inter-classe.
Par conséquent l'indice de dissimilarité entre deux classes (ou niveau d'agrégation de ces deux classes) est alors égal à la perte d'inertie inter-classe résultant de leur regroupement.
D'une part, l'inertie inter-classe étant la moyenne des carrés des distances des centres de gravité des classes au centre de gravité total, la variation d'inertie inter-classe, lors du regroupement de A et B est égale à :

$$m_A d^2(g_A, g) + m_B d^2(g_B, g) - (m_A + m_B) d^2(g_{A \cup B}, g) \tag{5.16}$$

où

$$m_A = \sum_{i \in A} p_i$$

$$g_A = \frac{1}{m_A} \sum_{i \in A} p_i x_i$$

$$g_{A \cup B} = \frac{m_A g_A + m_B g_B}{m_A + m_B}$$

De plus, la distance de Ward entre deux classes (A,B) de barycentres respectifs g_A etg_B est définie par

$$D(A,B) = m_A d^2(g_A,g) + m_B d^2(g_B,g) - (m_A+m_B)d^2(g_{A\cup B},g)$$
$$= m_A d^2(g_A,g) + m_B d^2(g_B,g) - (m_A+m_B)d^2(\frac{m_A g_A + m_B g_B}{m_A+m_B},g)$$

Autrement dit,

$$D(A,B) = m_A\|g_A-g\|^2 + m_B\|g_B-g\|^2 - (m_A+m_B)\parallel \frac{m_A g_A+m_B g_B}{m_A+m_B} - g \parallel^2$$

$$= m_A\|g_A-g\|^2 + m_B\|g_B-g\|^2 - \frac{1}{m_A+m_B}\|m_A(g_A-g)+m_B(g_B-g)\|^2$$

$$= m_A\|g_A-g\|^2+m_B\|g_B-g\|^2-\frac{m_A^2}{m_A+m_B}\|g_A-g\|^2-\frac{m_B^2}{m_A+m_B}\|g_B-g\|^2-2\frac{m_A m_B}{m_A+m_B} < g_A-g, g_B-g >$$

Par suite,

$$D(A,B) = \frac{m_A m_B}{m_A+m_B}(\|g_A-g\|^2 + \|g_B-g\|^2 - 2 < g_A-g, g_B-g >)$$

Nécessairement,

$$D(A,B) = \frac{m_A m_B}{m_A+m_B}\|(g_A-g)-(g_B-g)\|^2$$

Par conséquent,

$$D(A,B) = \frac{m_A m_B}{m_A+m_B}\|(g_A-g_B)\|^2 \tag{5.17}$$

La hiérarchie obtenue par C.A.H. avec cette mesure est la hiérarchie de Ward. D'autre part, la relation généralement utilisée pour définir l'indice h, une hiérarchie H construite par C.A.H. est définie comme suit

$$\text{Soit } A,B \in H, h(A\cup B) = \max(D(A,B),h(A),h(B)) \tag{5.18}$$

5.5.2 Classification Hiérarchique Descendante (CHD)

Les méthodes de classification descendante hiérarchique partent d'un ensemble d'individus Ω et construisent, de manière itérative, une partition de l'ensemble d'individus.
A l'inverse de la classification ascendante hiérarchique, à chaque étape de l'algorithme il y a deux processus à faire :

— Chercher une classe à scinder.
— Choisir un mode d'affectation des objets aux sous-classes.

Dans la première étape, les données sont divisées en deux classes au moyen des dissimilarités. Dans chacune des étape suivantes, la classe avec le diamètre le plus grand se divise de la même façon. Après $n-1$ divisions,tous les individus sont bien séparés. La dissimilarité moyenne entre l'individu x qui appartient à la classe C qui contient n individus et tous les autres individus de la classe C est définie par :

$$d_x = \frac{1}{n-1}\sum_{x\in C, y\neq x} d(x,y)$$

	Crev. Entiers	Poissons
SANTIG-DU	16.6308	17.9489
POSEIDON II	14.0892	7.1728
AGIOS SPYRIDON	22.5171	20.04
AFRODITI	21.78164	27.76
MELAKY 7	11.8066	18.4195
MELAKY 8	11.2254	19.4965
MELAKY 2	8.5105	13.8635
MELAKY 3	9.245	10.97

FIGURE 5.1 – Production de Crevettes en tonnes

5.6 Exemple

Si on calcul la distance entre les variables, on obtient :

Distance	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
w_1	0	11.072	6.246	11.081	4.847	5.623	9.09	10.161
w_2	11.072	0	15.382	21.977	11.476	12.652	6.691	6.155
w_3	6.246	15.382	0	7.755	10.832	11.305	15.308	16.075
w_4	11.081	21.977	7.755	0	13.666	13.406	19.216	20.954
w_5	4.847	11.476	10.832	13.666	0	1.224	5.623	7.878
w_6	5.623	12.652	11.305	13.406	1.224	0	6.253	8.753
w_7	9.09	6.691	15.308	19.216	5.623	6.253	0	2.985
w_8	10.161	6.155	16.075	20.954	7.878	8.753	2.985	0

Agrégation selon le lien minimum On les rassemble pour former le groupe $A = \{w_5, w_6\}$. On a une nouvelle partition de Γ :

$$\mathcal{P}_1 = \{\{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}, A, \{w_7\}, \{w_8\}\}$$

Tableau des dissimilarités associé à $\mathcal{P}_1$ est

$$D(w_1, A) = \min(D(w_1, w_5); D(w_1, w_6)) = 4.847$$

$$D(w_2, A) = \min(D(w_2, w_5); D(w_2, w_6)) = 11.476$$

$$D(w_3, A) = \min(D(w_3, w_5); D(w_3, w_6)) = 10.832$$

$$D(w_4, A) = \min(D(w_4, w_5); D(w_4, w_6)) = 13.406$$

$$D(w_7, A) = \min(D(w_7, w_5); D(w_7, w_6)) = 5.623$$

$$D(w_8, A) = \min(D(w_8, w_5); D(w_8, w_6)) = 7.878$$

	w_1	w_2	w_3	w_4	A	w_7	w_8
w_1	0	11.072	6.246	11.081	4.847	9.09	10.161
w_2	11.072	0	15.382	21.977	11.476	6.691	6.155
w_3	6.246	15.382	0	7.755	10.832	15.308	16.075
w_4	11.081	21.977	7.755	0	13.406	19.216	20.954
A	4.847	11.476	10.832	13.406	0	5.623	7.878
w_7	9.09	6.691	15.308	19.216	5.623	0	2.985
w_8	10.161	6.155	16.075	20.954	7.878	2.985	0

Les éléments(individus) w_7 et w_8 sont les plus proches. On les rassemble pour former le groupe $B = \{w_7, w_8\}$. On a une nouvelle partition de Γ :

$$\mathcal{P}_2 = \{\{w_1\}, \{w_2\}, \{w_3\}, \{w_4\}, A, B\}$$

Tableau des dissimilarités associé à $\mathcal{P}_2$ est

$$D(w_1, B) = \min(D(w_1, w_7); D(w_1, w_8)) = 9.09$$

$$D(w_2, B) = \min(D(w_2, w_7); D(w_2, w_8)) = 6.155$$

$$D(w_3, B) = \min(D(w_3, w_7); D(w_3, w_8)) = 15.308$$

$$D(w_4, B) = \min(D(w_4, w_7); D(w_4, w_8)) = 19.216$$

$$D(A, B) = \min(D(w_7, A); D(w_8, A)) = 5.623$$

	w_1	w_2	w_3	w_4	A	B
w_1	0	11.072	6.246	11.081	4.847	9.09
w_2	11.072	0	15.382	21.977	11.476	6.155
w_3	6.246	15.382	0	7.755	10.832	15.308
w_4	11.081	21.977	7.755	0	13.406	19.216
A	4.847	11.476	10.832	13.406	0	5.623
B	9.09	6.155	15.308	19.216	5.623	0

Les éléments w_1 et A sont les plus proches. On les rassemble pour former le groupe $C = \{w_1, A\}$. On a une nouvelle partition de Γ :

$$\mathcal{P}_3 = \{\{w_2\}, \{w_3\}, \{w_4\}, C, B\}$$

Tableau des dissimilarités associé à $\mathcal{P}_3$ est

$$D(w_2, C) = \min(D(w_2, w_1); D(w_2, A)) = 11.072$$

$$D(w_3, C) = \min(D(w_3, w_1); D(w_3, A)) = 6.246$$

$$D(w_4, C) = \min(D(w_4, w_1); D(w_4, A)) = 11.081$$

$$D(C, B) = \min(D(B, A); D(w_1, B)) = 5.623$$

	w_2	w_3	w_4	C	B
w_2	0	15.382	21.977	11.072	6.155
w_3	15.382	0	7.755	6.246	15.308
w_4	21.977	7.755	0	11.081	19.216
C	11.476	10.832	13.406	0	5.623
B	6.155	15.308	19.216	5.623	0

Les éléments B et C sont les plus proches. On les rassemble pour former le groupe $E = \{B, C\}$. On a une nouvelle partition de Γ :

$$\mathcal{P}_4 = \{\{w_2\}, \{w_3\}, \{w_4\}, E\}$$

Tableau des dissimilarités associé à $\mathcal{P}_4$ est

$$D(w_2, E) = \min(D(w_2, B); D(w_2, C)) = 6.155$$

$$D(w_3, E) = \min(D(w_3, B); D(w_3, C)) = 6.246$$

$$D(w_4, E) = \min(D(w_4, B); D(w_4, C)) = 11.081$$

	w_2	w_3	w_4	E
w_2	0	15.382	21.977	6.155
w_3	15.382	0	7.755	6.246
w_4	21.977	7.755	0	11.081
E	6.155	6.246	11.081	0

Les éléments w_2 et E sont les plus proches. On les rassemble pour former le groupe $F = \{w_2, E\}$. On a une nouvelle partition de Γ :

$$\mathcal{P}_5 = \{\{w_3\}, \{w_4\}, F\}$$

Tableau des dissimilarités associé à $\mathcal{P}_5$ est

	w_3	w_4	F
w_3	0	7.755	6.246
w_4	7.755	0	11.081
F	6.246	11.081	0

Les éléments w_3 et F sont les plus proches. On les rassemble pour former le groupe $G = \{w_3, F\}$. On a une nouvelle partition de Γ :

$$\mathcal{P}_6 = \{\{w_4\}, G\}$$

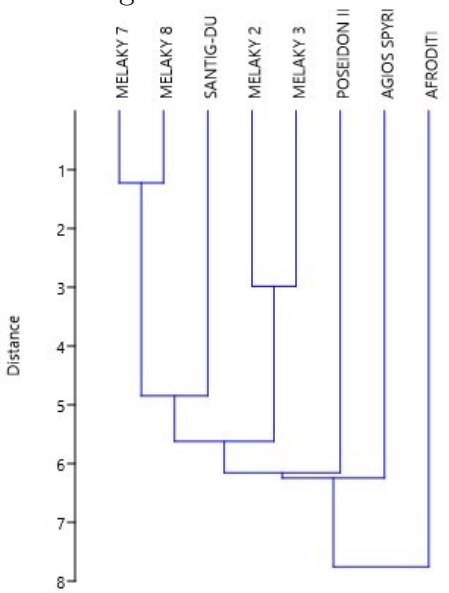
Tableau des dissimilarités associé à $\mathcal{P}_6$ est

$$D(w_4, F) = \min(D(w_3, w_4); D(w_4, F)) = 7.755$$

	w_4	G
w_4	0	7.755
G	7.755	0

Il ne reste plus que 2 éléments, w_4 et G ; on les regroupe.
On obtient la partition $\mathcal{P}_7 = \{w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8\}$. Cela termine l'algorithme de CAH.

FIGURE 5.2 – Dendogramme selon la méthode du lien minimum

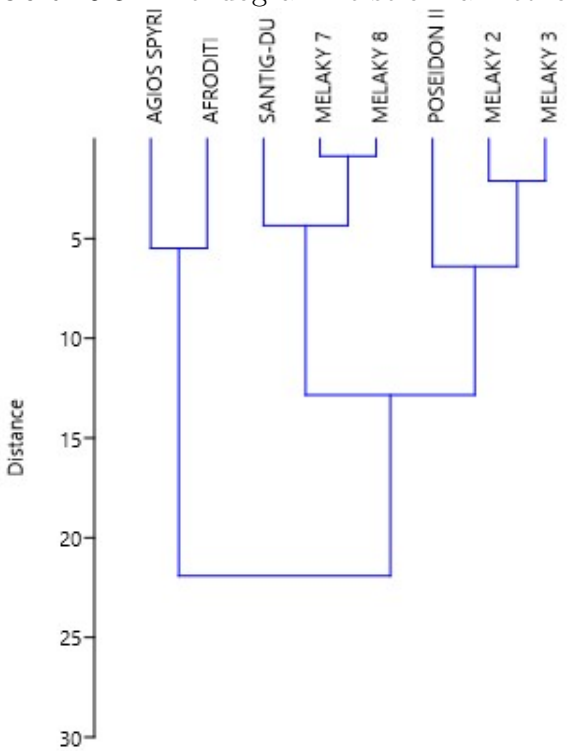


Méthode de Ward Même raisonnement mais la dissimilarité change :

$$D(A, B) = \frac{m_A m_B}{m_A + m_B} \|(g_A - g_B)\|^2 = \frac{n_A n_B}{n(n_A + n_B)} \left((x_A - x_B)^2 + (y_A - y_B)^2\right)$$

On obtient le dendrogramme associé :

FIGURE 5.3 – Dendogramme selon la méthode de Ward



5.7 Agrégation autour de centres mobiles

Cette méthode consiste à construire une partition en k classes en sélectionnant k individus comme centres des classes tirés au hasard de l'ensemble d'individus. Après cette sélection, on affecte chaque individu au centre le plus proche en créant k classes, les centres des classes seront remplacé par les centres de gravité et les nouveaux classes seront crée par le même principe. Il s'agit de la méthode (kmeans) proposée dans Forgy (1965).

Algorithme des kmeans

Initialisation Choisir k individus au hasard (comme centre des classes initiales)
Itérer jusqu'à ce que le critère de variance interclasse ne croisse plus de manière significative.
Pour $i = 1, \cdots, n$,

— Allouer l'individu i à la classe k telle que $d(x_i, g_k) \leq d(x_i, g_l)$ pour tout $l \neq k$
— Calculer les centres de gravités g_k des K classes.

Méthode de Lloyd

- On choisit q points au hasard dans $\mathbb{R}^p$ (Ces points sont appelés centres).
- On calcule le tableau de distances entre tous les individus et les q centres.
- On forme alors q groupes de la manière suivante : chaque groupe est constitué d'un centre et des individus les plus proches de ce centre que d'un autre. On obtient une partition $\mathcal{P}_1$ de Ω.
- On calcule le centre de gravité de chacun des q sous-nuages de points formés par les q groupes. Ces q centres de gravité sont nos nouveaux q centres.
- On calcule le tableau de distances entre tous les individus et les nouveaux q centres.
- On forme alors q groupes, chaque groupe étant constitué d'un centre et des individus les plus proches de ce centre que d'un autre. On a une nouvelle partition $\mathcal{P}_2$ de Ω.
- On itère la procédure précédente jusqu'à ce que deux itérations conduisent à la même partition.

— **Avantage :**
  — simple
  — Compréhensibles
  — Applicables à des données de grande taille
  — Compléxité des calculs en $\mathcal{O}(k.n)$
— **Défaut :**
  — Le nombre des classes doit être fixé au départ.
  — Le résultat dépend du tirage initial des centres des classes.

5.8 EXEMPLE

En utilisant l'exemple de donnée précédente, choisissant SANTIG-DU et POSEIDON II pour centres initiaux.
Distances entre les individus et ces centres est

	w_3	w_4	w_5	w_6	w_7	w_8
$c_1^0 = w_1$	6.25	11.08	4.85	5.62	9.09	10.16
$c_2^0 = w_2$	15.38	21.98	11.47	12.65	8.71	6.16

$\left(\frac{16.6308+22.5171+21.78164+11.8066+11.2254}{5}; \frac{17.9489+20.04+27.46+18.4195+19.4965}{5}\right)$ et $\left(\frac{14.0892+8.5105+9.245}{3}; \frac{7.1728+13.8635+10.97}{3}\right)$

- Classe1 $= \{w_1, w_3, w_4, w_5, w_6\}$ avec comme centre de classe $c_1^1 = (16.79; 20.73)$
- Classe2 $= \{w_2, w_7, w_8\}$ avec comme centre de classe $c_2^1 = (10.61; 10.67)$

Distances entre les individus et ces nouveaux centres est

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8
c_1^1	2.79	13.82	5.77	8.62	5.49	5.70	10.76	12.34
c_2^1	9.45	4.93	15.15	20.41	7.84	8.85	3.82	1.40

D'où les deux Classes :

- Classe1 $= \{w_1, w_3, w_4, w_5, w_6\}$
- Classe2 $= \{w_2, w_7, w_8\}$

On retrouve la même classification que l'étape précédente. Alors, l'algorithme s'arrête.

Item	Cluster
SANTIG-DU	1
POSEIDON II	2
AGIOS SPYRIDON	1
AFRODITI	1
MELAKY 7	1
MELAKY 8	1
MELAKY 2	2
MELAKY 3	2

5.8.1 Variantes

Méthodes des nuées dynamiques

La variante proposée par Diday (1971) consiste à remplacer chaque centre de classe par un noyau constitué d'éléments représentatifs de cette classe. Cela permet de corriger l'influence d'éventuelles valeurs extrêmes sur le calcul du barycentre.

5.9 Combinaison

Il est courant de combiner les méthodes de classification introduites précédemment. En effet, la méthode de classification hiérarchique n'est raisonnablement applicable que si le nombre d'observations est relativement faible. Son résultat constitue néanmoins souvent une initialisation intéressante pour une méthode des k-moyennes. Il fournit en effet à la fois des critères pour sélectionner le nombre de classes et une initialisation des centres de classes . La stratégie suivante, adaptée aux grands ensembles de données, permet de contourner ces difficultés.

1. Réaliser une classification par nuées dynamique sur un sous échantillon tiré au hasard et de taille environ 10% de n. On choisit un nombre de classes grand.
2. Sur les barycentres des classes précédentes, exécuter une classification hiérarchique puis déterminer un nombre de classes "optimal" K.
3. Exécuter une classification par k-moyennes pour K classes et en choisissant comme valeurs initiales des centres de classe les barycentres des classes de l'étape précédente. On pourra pondérer ces centres par le nombre d'individus dans les classes.

5.10 La Classification des Données Qualitatives

Les n individus à classer décrits par des variables qualitatives

On peut présenter les données sous la forme d'un tableau disjonctif complet (TDC) de dimension $n \times r$, où r est le nombre total de modalités des p caractères considérés. On utilise un des indices de dissimilarité déduit des indices de similarité proposés qui combinent de diverses manières les quatre nombres suivants associés à un couple d'individus.

— a est le nombre de présence (cas 1) pour le couple d'individus.
— b est le nombre de présence pour l'individu i et absence pour l'individu j.
— c est le nombre d'absence pour l'individu i et présence pour l'individu j.
— d est le nombre d'absence (cas 0) pour le couple d'individus .

Pour tout $i \in \{1, \cdots, n\}$, la i-ème ligne du tableau est constituée du vecteur $(n_{i,1}, \cdots, n_{i,k}, \cdots, n_{i,r})$, avec

$$n_{i,k} = \begin{cases} 1 & \text{si } i \text{ possède la modalité } k \\ 0 & \text{Sinon} \end{cases}$$

Les similarités suivantes ont été proposées par différents auteurs :

- Indice de Russel et Rao :
$$s(i,j) = \frac{a}{a+b+c+d}$$
- Indice de Dice :
$$s(i,j) = \frac{2a}{2a+b+c}$$
- Indice de Jaccard :
$$s(i,j) = \frac{a}{a+b+c}$$
- Indice de d'Anderberg :
$$s(i,j) = \frac{a}{a+2(b+c)}$$
- Indice de Rogers et Tanimoto :
$$s(i,j) = \frac{a+d}{a+d+2(b+c)}$$
- Indice de Yule :
$$s(i,j) = \frac{ad-bc}{ad+bc}$$
- Indice de Pearson :
$$s(i,j) = \frac{ad-bc}{\sqrt{(a+b)(a+c)(d+b)(d+c)}}$$
- Indice de Ochiaï : :
$$s(i,j) = \frac{a}{\sqrt{(a+b)(a+c)}}$$
- Indice de Sokal et Michener :
$$s(i,j) = \frac{a+d}{a+b+c+d}$$

À partir du tableau disjonctif complet, on appelle "distance" du Chi-deux (χ^2) entre i et j la distance :

$$d^2_{\chi^2} = \sum_{k=1}^{r} \frac{1}{\rho_k}(f_{i,k} - f_{j,k})^2$$

où

$$f_{i,k} = \frac{n_{i,k}}{n_{i,\bullet}} \qquad n_{i,\bullet} = \sum_{k=1}^{r} n_{i,k} \qquad \rho_k = \frac{n_{\bullet,k}}{n} \qquad n_{\bullet,k} = \sum_{i=1}^{n} n_{i,k}$$

On peut aussi utiliser cette "distance" pour mettre en œuvre l'algorithme de CAH.
Dans le cas où les variables qualitatives à $m_1, m_2, \cdots, m_p$ modalités il est très difficile de définir une distance.

5.10.1 Caractères de natures différentes :

Si le tableau est composé de données mixtes, certains caractères sont qualitatifs et d'autres quantitatifs. Différentes stratégies sont envisageables dépendant de l'importance relative des nombres de variables qualitatives et quantitatives.

- Il suffit de rendre les caractères qualitatifs en quantitatifs en introduisant des classes de valeurs et en les considérant comme des modalités.
- **Metrique de Gower** permet de mélanger les types de variables ainsi reste très peu utilisée.
- Inversement du première point. Les variables quantitatives sont rendues qualitatives par découpage en classes.

5.11 Classification spectrale

La classification spectrale vise à obtenir un partitionnement des données (ou observations) en groupes à partir d'une représentation de ces données sous la forme d'un graphe de similarité s. Dans un tel graphe, chaque sommet correspond à une donnée (ou observation) et chaque arête qui relie deux observations est pondérée par la similarité entre ces observations. Les méthodes de partitionnement spectrales, quant à elles, fonctionnent également très bien lorsque les classes ne sont pas forcément convexes.
Avant d'aller plus loin, il est nécessaire d'introduire (ou rappeler) quelques notions concernant les graphes.

Un **graphe pondéré et non orienté** est un couple $G = (V, E)$, où $V = \{x_1, x_2, \cdots, x_n\}$ est l'ensemble des n sommets et E l'ensemble des arêtes qui sont pondérées. On appelle G le graphe de similarité.

Soit $w_{ij} \geq 0$ le poids de l'arête qui lie les sommets x_i et x_j, alors $w_{ij} = 0$ si et seulement si il n'y a pas d'arête entre les sommets x_i et x_j.On note alors $W = (w_{ij})_{i=1,\cdots,n;j=1,\cdots,n}$ la matrice d'adjacence du graphe.

Le **degré** du sommet x_i est simplement le nombre d'arêtes dont l'une des extrémités est égale à x_i. Dans le cas d'un graphe pondéré, on définit comme la somme des poids des arêtes qui relient ce sommet à d'autres $d_i = \sum_{j=1}^{n} w_{ij}$. On peut considérer une matrice diagonale des degrés $D = diag(d_1, d_2, \cdots, d_n)$.

Une **chaîne** ou un **chemin** entre les sommets x_i et x_j est une suite d'arêtes de E permettant de relier les deux sommets.
Un sous-ensemble $A \subset V$ forme une composante connexe du graphe $G = (V, E)$ si

- Quels que soient deux sommets de A, il existe une chaîne de G les reliant.

- Pour tout sommet x_i de A et toute chaîne avec des arêtes de E et partant de x_i, le sommet terminal de la chaîne est également dans A .

Le **vecteur indicateur** d'un sous-ensemble A de V est le vecteur c à n composantes binaires tel que $c_i = 1$ si le sommet $x_i \in A$ et $c_i = 0$ sinon. De même lorsque A est une composante connexe.

5.11.1 Construction de la matrice de similarités S

Mesure de Similarité

Dans la littérature, plusieurs techniques permettent d'obtenir une mesure de similarité entre deux objets. Le choix de la fonction de similarité dépend essentiellement du domaine de provenance des données. Certains ont été évoqués précédemment.
D'autres sont plus spécifiques aux algorithmes de partitionnement spectral ; cependant, les deux formules les plus répandues et les plus utilisées sont :

— Distance cosinus

$$s_{ij} = \frac{< x_i, x_j >}{\| x_i \| \| x_j \|}$$

— Noyau Gaussien

$$s_{ij} = \exp\left(-\frac{d^2(x_i, x_j)}{2\sigma^2}\right)$$

Les différents types de graphes pondérés

L'objectif de la construction d'un graphe pondéré est de rendre compte (ou de modéliser) les relations de voisinage entre les objets. Il existe alors plusieurs façons de procéder :

Graphe entièrement connecté : connexion de tous les nœuds entre eux,et on pondère l'arête reliant x_i et x_j par la mesure de similarité s_{ij} entre ces points.

Graphe du ε-voisinage : connexion des nœuds pour lesquels la similarité s_{ij} est supérieure à ε.

Graphe des k-plus proches voisins : On ne relie un sommet x_i qu'avec ses k plus proches voisins. Autrement dit, connexion du i^{e} nœud avec le j^{e} nœud si et seulement si l'objet x_j fait partie des k plus proches voisins de l'objet x_i.

5.11.2 Partitionnement du graphe

Les méthodes de partitionnement spectrales s'appuyent sur le spectre de la matrice de similarité du graphe,et plus précisément sur les valeurs et vecteurs propres de la matrice Laplacienne du graphe G.

Aucune normalisation : $L = D - W$

La normalisation symétrique : $L_{sym} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$

La normalisation « marche aléatoire » (random walk) : $L_{rw} = D^{-1} L$

La normalisation additive : $L_{ad} = \frac{L + d_{max} I - D}{d_{max}}$ avec $d_{max} = \sup_{1 \leq i \leq n} d_i$

5.11.3 Algorithme de classification spectrale

1. Soit un ensemble de points $\Omega = \{x_1, x_2, \cdots, x_n\}$ et une matrice de similarité associée S.
2. A l'aide du graphe, une représentation vectorielle « convenable » (du point de vue de la classification automatique) des données est obtenue par la méthode suivante :

(a) Calculer la matrice Laplacienne

(b) Calculer les vecteurs propres $u_1, u_2, \cdots, u_k$ associés aux k plus petites valeurs propres de la matrice Laplacienne.

(c) Soit U la matrice dont les colonnes sont les vecteurs $u_1, u_2, \cdots, u_k$.

(d) Soit $y_i \in \mathbb{R}^k, i = 1, 2, \cdots, n$, les n lignes de la matrice U, chaque donnée (observation) est représentée par le vecteur y_i.

3. Réaliser une classification de base (comme k-means) pour obtenir les k groupes disjoints $G_1, G_2, \cdots, G_k$

4. Définir les classes $P_1, P_2, \cdots, P_k$ par :

$$P_i = \{x_j \mid y_j \in G_i\}$$

Remarque 17 *Dans le cas partitionnement spectral normalisé.La matrice Laplacienne L se change par L_{sym} ainsi normaliser la matrice U de sorte que la norme par lignes soit égale à 1 et créer la matrice V telle que l'élèment $v_{ij} = \frac{u_{ij}}{\|u_i\|}$. L'algorithme se déroule sur $v_1, v_2, \cdots, v_k$ au lieu de $u_1, u_2, \cdots, u_k$.*

Annexe A

DERIVEES MATRICIELLES

Soit le vecteur $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ et une fonction à valeurs dans $\mathbb{R}$ tel que $f(X) = f(x_1, x_2, \cdots, x_n)$

Alors,

$$\frac{\partial f(X)}{\partial X} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

est appelé dérivée matricielle de f(X) par rapport à X.
Soit $v \in \mathbb{R}^n$ et $a \in \mathbb{R}^n$

$$\frac{\partial(v^T a)}{\partial v} = \frac{\partial(\sum_{i=1}^n v_i a_i)}{\partial v} = \frac{\partial(\sum_{i=1}^n a_i v_i)}{\partial v} = \frac{\partial(a^T v)}{\partial v} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = a$$

Soit une matrice $M \in \mathbb{R}^{n \times n}$

$$\frac{\partial(v^T M v)}{\partial v} = \frac{\partial(v^T)}{\partial v} M v + \frac{v^T \partial(Mv)}{\partial v} = (M + M^T)v$$

En effet,

$$\frac{\partial((Mv)^T(Mv))}{\partial v} = \frac{\partial(v^T M^T M v)}{\partial v}$$

Par conséquent,

$$\frac{\partial((Mv)^T(Mv))}{\partial v} = \left(M^T M + (M^T M)^T\right) v$$

On conclut que,

$$\frac{\partial((Mv)^T(Mv))}{\partial v} = 2M^T M v$$

Annexe B

Décomposition de l'inertie Totale

Si on décompose $\mathbb{R}^p$ comme la somme de sous-espaces de dimension 1 et orthogonaux entre eux :

$$\Delta_1 \perp \Delta_2 \perp \cdots \perp \Delta_p$$

En effet, en appliquant le théorème de Pythagore,on a :

$$d^2(G, u_i) = d^2(G, h_{\Delta_2 \oplus \Delta_3 \oplus \cdots \oplus \Delta_p i}) + d^2(u_i, h_{\Delta_2 \oplus \Delta_3 \oplus \cdots \oplus \Delta_p i})$$

où $h_{\Delta_2 \oplus \Delta_3 \oplus \cdots \oplus \Delta_p i}$ est la projection orthogonale de u_i sur le sous-espace

$$\Delta_1^* = \Delta_2 \oplus \cdots \oplus \Delta_p$$

complémentaire orthogonal de Δ_1 sur $\mathbb{R}^p$. Remarquant que tous les axes sont orthogonaux :

$$d^2(u_i, h_{\Delta_2 \oplus \Delta_3 \oplus \cdots \oplus \Delta_p i}) = d^2(G, h_{\Delta_1 i})$$

De la même façon ,

$$d^2(G, h_{\Delta_2 \oplus \Delta_3 \oplus \cdots \oplus \Delta_p i}) = d^2(G, h_{\Delta_3 \oplus \Delta_4 \oplus \cdots \oplus \Delta_p i}) + d^2(G, h_{\Delta_2 i})$$

Ainsi de suite,

$$d^2(G, u_i) = d^2(G, h_{\Delta_1 i}) + d^2(G, h_{\Delta_2 i}) + \cdots + d^2(G, h_{\Delta_p i})$$

Il s'ensuit que

$$I = I_{\Delta_1^*} + I_{\Delta_2^*} + \cdots + I_{\Delta_p^*}$$

Annexe C

Méthode des multiplicateurs de Lagrange

Pour chercher les optimums d'une fonction $f(x_1, x_2, \cdots, x_n)$ sous la contrainte $h(x_1, x_2, \cdots, x_n) = cte$
On calcul les dérivées partielles de la fonction

$$g(x_1, x_2, \cdots, x_n, \lambda) = f(x_1, x_2, \cdots, x_n) - \lambda(h(x_1, x_2, \cdots, x_n) - cte)$$

par rapport à chacune des variables (λ est appelé le multiplicateurs de Lagrange). En annulant ces dérivées partielles, on obtient un système de $n+1$ équations à $n+1$ inconnues. L'existence de solutions à ce système est une condition nécessaire mais pas suffisante pour avoir un optimum .

Dans le cas où les n variables sont soumises à p contraintes, le méthode reste valable mais en ajoutant une combinaison linéaire des p contraintes dont les coefficients $\lambda_1, \lambda_2, \cdots, \lambda_p$ sont les multiplicateurs de Lagrange.
On doit alors résoudre un système de $n + p$ équations à $n + p$ inconnues.

Bibliographie

[A 74] A.I.D.E.P :*Statistiques et Probabilités (Dunod 1974) 2 tomes.*

[An 67] Anderson T. :*A bibiography of multivariate analysis (Olivier & Boyd 1967)*

[Ar 04] Arnaud MARTIN :*Polycopié de cours ENSIETA - Réf. : 1463* Septembre 2004

[Ben80 a] J.P. Benzecri :*L'analyse de données (Tome 1) La taxinomie. Dunod, 1980.*

[Ben80 b] J.P. Benzecri :*L'analyse de données (Tome 2) L'analyse des correspondances. Dunod, 1980.*

[Bro 03] G. Brossier : *Analyse des données, chapitre Les éléments fondamentaux de la classification. Hermes Sciences publications,2003*

[Char 07] Charlotte Baey :*Analyse de données M2 Ingénierie Statistique et Numérique 2017-2018*

[Fen 81] Fenelon J. :*Qu'est-ce que l'analyse des Données (Lefonen 1981) .*

[Gu 77] Guignou J.L. :*Méthodes multidimensionnelles(Dunod 1977.*

[Jam 78] Jambu M.,Lebeaux M.O : *Classification automatique pour l'analyse des données (Dunod 1978)* 2 tomes.

[Mar] Marie-Christine Roubaud :*Le polycopié du cours de M1MASS d'Analyse exploratoire des données*

[Mon] Monbet V. :*Analyse des données Master Statistique et économétrie*

Résumé

Le travail de mémoire a porté sur une réflexion relative aux méthodes de classification automatique des données pour lesquelles il est bien connu qu'un effet "méthode" existe. La première partie qui présente la problématique générale de l'analyse des données sur différente méthodes d'Analyses factorielles suivi une aperçu des méthodes de classification. Chaque qui consistent à réduire le problème d'optimisation pour faciliter une représentation graphique des données.

Dans l'analyse factorielle, ils existent quatres méthodes selon le cas du type des données. De même sur la classification .

Ainsi, chaque méthodes met en évidence la qualité et la dispersion de l'observation sur la représentation.

Abstract

This memory focused on a reflexion on the automatic data classification methods for which it is well known that the "method" effect existe. A first part which presents the general problem of data analys on different methods of factor analysis and proposes a survey of classification methods. Each method consists in reducing the optimization problem to facilitate a graphical representation of the data.

In the factorial analysis, there are four methods depending on the type of data. Likewise on the classification.

So, each method demonstrater the quality and dispersion of the observation on the representation.

Titre :METHODE D'ANALYSE DES DONNEES ET APPLICATION

NOM : NIAVOMALALA Ravoniaina Harison

E-mail : niavomalalaravoniainaharison@yahoo.com

RAPPORTEUR : Monsieur RAKOTONDRALAMBO Joseph

Maître de conférences