République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOHAMED KHIDER, BISKRA

FACULTE des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DEPARTEMENT DE MATHEMATIQUES

Mémoire présenté par

Berramdane Ikram

En vue de l'obtention du Diplôme de

MASTER en Mathématiques

Option : **Statistique**

Titre

Analyse Factorielle Discriminante

Membres du Comité d'Examen

Pr.	Abdelhakim Necir	UMKB	Président
Pr.	Djamel Meraghni	UMKB	Encadreur
Dr.	Louiza Soltane	UMKB	Examinateur

Juin 2024

Dédicace

Je dédie ce humble travail :

À la source de ma force : Mes chers parents.

À mes soeurs.

À mon seul frère.

À mes chères amies.

À tous mes amies.

À toutes les personnes qui ont contribués de près ou de loin à la réalisation de ce travail.

Ikram Berramdane

REMERCIEMENTS

Tout d'abord je tiens à remercier **"Allah"** qui m'a accordé la santé, la force et la patience pour terminer ce travail.

Mes sincères remerciements, mon respect et ma gratitude à mes parents qui ont toujours été là pour moi.

Je tiens à exprimer toute ma reconnaissance à mon encadreur

le Professeur **Merghni Djamel**, professeur de mathématiques à l'université de Biskra, pour sa patience, ses conseils.

Mes vifs remerciements vont également aux membres du jury, le Professeur **Necir Abdelhakim** et le Docteur **Soltane Louiza**

Je remercie touts les enseignants du département de Mathématiques.

Je remercie également tous ceux qui étaient à côtés de moi et qui m'ont soutenu dans ma carrière universitaire.

Ikram Berramdane

Table des matières

Table des figures

Liste des tableaux

Introduction

L'analyse factorielle discriminante (AFD) ou Discriminant factor analysis (DFA) en anglais, est une technique statistique descriptive et prédictive utilisée pour classer des observations dans des groupes prédéfinis en fonction de deux ou plusieurs variables explicatives.

L'origine de cette méthode remonte aux travaux de Fisher (1936) ou de façon moins discrète, à ceux de Mahalanobis (1936). Elle est une des techniques d'analyse multidimensionnelle les plus utilisées en pratique (sciences médicales, contrôle de qualité, prévision de risques, ...) (voir, [8], page 251).

L'AFD est appliquée dans de nombreux domaines :

- En informatique : reconnaissance optique de caractères à partir d'information simples, comme la présence ou non de symétrie, le nombre d'extrémités,...
- En médecine : classement des patients en différents groupes de diagnostic en fonction de leurs symptômes, résultats de tests ou données biomédicales.

L'objet de l'AFD est de répondre aux questions suivantes :

Quelles sont les variables et leurs groupes (ou classes) ?

Quelles sous-aspects permettent la meilleure discrimination entre les groupes (ou classes) ?

A quelle classe (ou groupe) appartiennent de nouveaux individus à la lumière de leurs valeurs pour les variables quantitatives ?

Il y a deux aspects distincts à l'AFD :

1. Aspect descriptif : trouver la combinaison linéaire des variables explicatives qui permet de séparer au mieux les différents groupes.
2. Aspect décisionnel : affecter un nouvel individu à une classe spécifique. Il s'agit donc d'un problème de classement.

Ce mémoire se compose de deux chapitres :

- Chapitre 1 : présentation des données (individus et variables) et leurs caractéristiques.
- Chapitre 2 : description de l'AFD, ses étapes et ses objectifs (méthode géométrique et méthode bayésienne). Une application sur des données réelles (victimes d'infarctus du myocarde) est faite à la fin du chapitre en utilisant les packages "cluster", "klaR", "MASS" et "mda" du logiciel de traitement statistique R.

Chapitre 1

Données multivariées

Dans ce chapitre, on présente des données individus et variables, et leurs caractéristiques.

1.1 Tableaux de données

On dispose d'un échantillon de $n \geq 1$ individus $e_1, ..., e_n$ décrits par $p \geq 1$ variables quantitatives $x_1, ..., x_p$. L'ensemble des observations est présenté sous la forme d'un tableau appelé tableau des données, (voir [7], page 5).

C'est une matrice réelle à n lignes et p colonnes, notée par X, de terme général x_{ij} qui représente la valeur prise par la variable x_j sur l'individu e_i.

$$X = . \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & x_{n-1p} \\ x_{n1} & \cdots & x_{np-1} & x_{np} \end{pmatrix} \tag{1.1}$$

1.2 Espace des individus

Un individu e_i est représenté par un vecteur de $\mathbb{R}^p$ dont les composantes sont les valeurs qu'il prend par rapport aux différentes variables $x_1, ..., x_p$. C'est la $i^{ème}$ ligne du tableau X :

$$e_i = (x_{i1}, ..., \ x_{ip})^t \in \mathbb{R}^p; \ i = 1, ..., n.$$

1.2.1 Matrice des poids

Chaque individu e_i est doté d'un poids p_i qui reflète son importance relative par rapport aux autres individus. Les $p_1, ..., p_n$ sont regroupés dans une matrice diagonale, de dimension $(n \times n)$, notée par D, (d'apré [9] page 156),

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

Remarque 1.2.1

1. *Pour $i = 1, ..., n$, on a $0 < p_i < 1$ et $\sum_{i=1}^{n} p_i = 1$.*
2. *Dans le cas usuel où tous les individus ont la même importance, les poids sont tous égaux à $1/n$. Dans ce cas, on a*

$$D = \frac{1}{n} I_n,$$

où $I_n$représente la matrice identité d'ordre n.

Preuve. Comme on a $p_1 = ... = p_n$ et $\sum_{i=1}^{n} p_i = 1$, alors

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} p_1 = p_1 \sum_{i=1}^{n} 1 = np_1 = 1.$$

Par conéquent

$$p_i = p_1 = \frac{1}{n},\ i = 1, ..., n,$$

et

$$D = \begin{pmatrix} \frac{1}{n} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{n} \end{pmatrix} = \frac{1}{n} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix} = \frac{1}{n} I_n.$$

C'est le résultat voulu. ■

1.2.2 Centre de gravité

Le centre de gravité g est un vecteur de $\mathbb{R}^p$ dont la j-*ème* coordonnée g_j correspond à la moyenne arithmétique $\bar{x_j}$ de la variable j sur l'ensemble des n individus. En d'autres termes, le centre de gravité généralise le concept de moyenne arithmétique dans le contexte d'une analyse multidimensionnelle. Il est définit par :

$$g := (\bar{x_1},\ \ \ldots, \bar{x_p})^t \in \mathbb{R}^p. \tag{1.2}$$

où

$$\bar{x_j} := \sum_{i=1}^{n} p_i x_{ij},\ \ j = 1,\ \ \ldots,\ p.$$

désigne la moyenne arithmétique de la variable x_j,(voir [7], page 6).

Remarque 1.2.2 *Si on désigne par* 1_n *le vecteur de* $\mathbb{R}^n$ *dont toutes les composantes sont égales à* 1, *alors le centre de gravité peut être écrit sous la forme matricielle :*

$$g = X^t D 1_n.$$

1.2.3 Nuage des individus

Chaque individu e_i est un point dans l'espace vectoriel $\mathbb{R}^p$ (appelé espace des individus), où chaque dimension correspond à une variable. L'ensemble des n points représentant les individus forme un nuage dans $\mathbb{R}^p$, appelé le nuage des individus (voir [9], pages157-159).

Métrique

La métrique (notée M) est une matrice carrée symétrique définie positive de taille p. Elle permet de mesurer la distance entre deux individus e_i et $e_{i'}$, dans l'espace $\mathbb{R}^p$ selon la formule suivante :

$$d^2_M(e_i,\ e_{i'}) := (e_i - e_{i'})^t\, M\, (e_i - e_{i'})\,. \tag{1.3}$$

C'est le produit scalaire, par rapport à M, entre la ligne $(e_i - e_{i'})^t$ et la colonne $(e_i - e_{i'})$. Ce produit scalaire est défini sur $\mathbb{R}^p \times \mathbb{R}^p$ par

$$\langle a, b\rangle_M := a^t M b.$$

Les métriques les plus utilisées sont la métrique usuelle $M = I_p$ et la métrique diagonale des inverses des variances s_j^2 définies par (1.6) :

$$M = D_{1/s^2} = \begin{pmatrix} \frac{1}{s_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{s_2^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{s_p^2} \end{pmatrix}.$$

Il est à noter que le produit scalaire usuel $\langle a, b\rangle := a^t b$ correspond à la métrique canonique $M = I_p$.

Ressemblance entre individus

Pour les individus, on s'intéresse à leur ressemblance. Deux individus sont considérés comme semblables ou proches s'ils partagent des valeurs similaires pour l'ensemble des variables. Cette définition implique une conception de proximité, par rapport à une métrique M, traduite par la distance définie par (1.3) . Plus les points représentant deux individus sont proches dans la nuage, plus les deux individus sont semblables.

1.3 Espace des variables

Une variable x_j est représentée par un vecteur de $\mathbb{R}^n$ dont les composantes sont les valeurs prises par cette variable sur les différents individus $e_1, ..., e_n$. C'est la $j^{ème}$ colonne du tableau X :

$$x_j = (x_{1j}, \ldots, x_{nj})^t \in \mathbb{R}^n;\ j = 1, ..., p.$$

Il existe deux types des variables :

1. Variables quantitatives : quantités exprimées en nombres réels et peuvent être soumises à des opérations arithmétiques telles que le calcul de la moyenne. Des exemples courants incluent l'âge, le poids et la taille.
2. Variables qualitatives : quantités descriptives à valeurs non numériques telles que le sexe, la situation maritale, la couleur des cheveux,... d'une personne.

1.3.1 Matrice de covariance-matrice de corrélation

Matrice de covariance

C'est une matrice carrée symétrique, d'ordre p, notée par S, de terme général (voir [8], page 254) :

$$s_{jj'} = Cov(x_j, x_{j'}) := \sum_{i=1}^{p} p_i \left(x_{ij} - \bar{x_j}\right) \left(x_{ij'} - \bar{x_{j'}}\right);\ j, j' = 1, ..., p. \tag{1.4}$$

Matrice de corrélation

C'est une matrice carrée symétrique, de dimension p, notée par R, de terme général :

$$r_{jj'} = \frac{s_{jj'}}{s_j s_{j'}};\ j, j' = 1, ..., p. \tag{1.5}$$

Remarque 1.3.1

1. *Le terme diagonal s_{jj} de S est notée par s_j^2. Il représente la variance de la variable x_j:*

$$s_j^2 = Var(x_j) := \sum_{i=1}^{p} p_i \left(x_{ij} - \bar{x_j}\right)^2;\ j = 1, ..., p. \tag{1.6}$$

   Sa racine carrée (positive) s_j est appelée écart type de x_j.

2. *Le terme diagonal r_{jj} de R est égal à* 1.

1.3.2 Standardisation des données

Il y a deux catégories de transformations appliquées aux données initiales le centrage et la réduction, ([7] pages 6-8) :

Tableau centré

Le tableau centré correspondant au tableau initial X est une matrice Y de terme général :

$$y_{ij} = x_{ij} - \bar{x_j};\ i = 1, ..., n \text{ et } j = 1, ..., p.$$

Sa forme matricielle est :

$$Y = X - 1_n g^t = \begin{pmatrix} x_{11} - \bar{x_1} & \cdots & x_{1p} - \bar{x_p} \\ \vdots & \ddots & \vdots \\ x_{n1} - \bar{x_1} & \cdots & x_{np} - \bar{x_p} \end{pmatrix} \in \mathcal{M}(n, p). \tag{1.7}$$

Tableau centré-réduit

Le tableau centré-réduit correspondant au tableau initial X est une matrice Z de terme général :

$$z_{ij} = \frac{x_{ij} - \bar{x_j}}{s_j} = \frac{y_{ij}}{s_j};\ i = 1, ..., n \text{ et } j = 1, ..., p.$$

Son écriture matricielle est :

$$Z = Y D_{1/s} = \begin{pmatrix} \frac{y_{11}}{s_1} & \dots & \frac{y_{1p}}{s_p} \\ \vdots & \ddots & \vdots \\ \frac{y_{n1}}{s_1} & \dots & \frac{y_{np}}{s_p} \end{pmatrix} \in \mathcal{M}(n, p). \tag{1.8}$$

où $D_{1/S}$ désigne la matrice diagonale, de taille p, des inverses des écarts-types :

$$D_{1/s} = \begin{pmatrix} \frac{1}{s_1} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{s_p} \end{pmatrix}.$$

Remarque 1.3.2 *On peut exprimer les matrices de covariance et de corrélation en termes des tableaux Y et Z comme suit :*

$$S = Y^t D Y \ \textit{et}\ R = Z^t D Z. \tag{1.9}$$

En remplaçant Y et Z par leurs formes respectives (1.7) et (1.8), on obtient :

$$S = X^t D X - g g^t \ \textit{et}\ R = D_{1/s} S D_{1/s}.$$

1.3.3 Nuage des variables

Chaque variable x_j pêut être exprimée comme un vecteur dans l'espace vectoriel $\mathbb{R}^n$,où chaque dimension représente un individu. L'ensemble des p variables forme un nuage de points dans $\mathbb{R}^n$, appelé le nuage des variables, ([9] page161).

Liaison entre deux variables

Pour les variables, on s'intéresse à leurs liaisons. On mesure la liaison entre deux variables x_j et $x_{j'}$, à l'aide de leur covariance $s_{jj'}$ ou de leur coefficient de corrélation $r_{jj'}$ respectivement définis par (1.4) et (1.5) .

Métrique des variables

Pour étudier la proximité entre les variables, il est nécessaire d'équiper l'espace des variables d'une métrique. En d'autres termes, il s'agit de trouver une matrice symétrique et définie positive d'ordre n. Le choix de cette matrice diagonale est imposé par la définition de la covariance où la matrice des poids D joue le rôle de métrique. Ainsi, on définit le produit scalaire entre deux variables x_j et $x_{j'}$, comme suit :

$$\langle x_j, x_{j'} \rangle_D = x_j^t D x_{j'} = \sum_{i=1}^{n} p_i x_{ij} x_{ij'};\ j, j' = 1, \ldots, p.$$

Remarque 1.3.3

1. *Si les deux variables x_j et $x_{j'}$ sont centrées alors :*

$$\langle x_j, x_{j'} \rangle_D = Cov(x_j, x_{j'}) = s_{jj'};\ j, j' = 1, \ldots, p.$$

2. *La norme d'une variable centrée x_j est égale à son écart-type :*

$$\| x_j \|_D = s_j;\ j = 1, \ldots, p.$$

3. *Géométriquement, la corrélation entre deux variables centrées x_j et $x_{j'}$ est égale au cosinus de l'angle $\theta_{jj'}$ formé par ces deux variables :*

$$r_{jj'} = \frac{s_{jj'}}{s_j s_{j'}} = \frac{\langle x_j, x_{j'} \rangle_D}{\| x_j \|_D \| x_{j'} \|_D} = \cos \theta_{jj'}.$$

Exemple 1.3.1 *Les évaluations de cinq élèves dans trois matières (mathématiques, anglais et sciences) sont notées sur une échelle de* 10 *et résumées dans le Tableau 1.1 qu'on note par* $X \in \mathcal{M}(5,3)$*.([7], page 8)*

mathématiques	anglais	sciences
4	6	8
5	2	7
6	3	6
10	2	5
0	7	4

TAB. 1.1 – Notes des élèves

La matrice des poids est égale à $D = I_5/5$*. En appliquant la formule* (1.2) *, on otient le centre de gravité de* X *:*

$$g = (5, 4, 6)^t.$$

On déduit, d'après (1.7) *, le tableau centré* Y *correspondant :*

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

En utilisant la première formule dans (1.9) *, on obtient la matrice de covariance :*

$$S = \begin{pmatrix} 10.4 & -5.6 & 0.6 \\ -5.6 & 4.4 & -0.4 \\ 0.6 & -0.4 & 2 \end{pmatrix}.$$

Les écarts-type des trois variables sont donc :

$$s_1 = \sqrt{10.4} = 3.22,\ \ s_2 = \sqrt{4.4} = 2.10 \ \textit{et}\ s_3 = \sqrt{2} = 1.41.$$

et ainsi on a la matrice :

$$D_{1/s} = \begin{pmatrix} 0.31 & 0 & 0 \\ 0 & 0.48 & 0 \\ 0 & 0 & 0.71 \end{pmatrix}.$$

Le tableau standard Z *correspondant à* X *est obtenu par la formule* (1.8) :

$$Z = \begin{pmatrix} -0.31 & 0.94 & 1.42 \\ 0 & -0.94 & 0.71 \\ 0.31 & -0.47 & 0 \\ 1.55 & -0.94 & -0.71 \\ -1.55 & 1.41 & -1.42 \end{pmatrix}.$$

Enfin, on utilise la deuxième formule de la relation (1.9) *pour calculer la matrice de corrélation :*

$$R = \begin{pmatrix} 1 & -0.82 & 0.13 \\ -0.82 & 1 & -0.13 \\ 0.13 & -0.13 & 1 \end{pmatrix}.$$

Chapitre 2

Discrimination

L'analyse factorielle discriminante (AFD) est une technique destinée à classer des individus à un groupe prédéfini (classe, modalité de la variable à prédire). On distingue deux aspects dans cette méthode : aspect descriptif et aspect décisionnel. Ce dernier est réalisé à travers des méthodes géométriques et/ou probabilistes.

2.1 Les données

Soient n individus $e_1, ..., e_n$ $(n \geq 1)$, de poids respectifs $p_1, ..., p_n$, décrits par p variables $x_1, ..., x_p$ $(p \geq 1)$ et répartis en q classes (groupes) $C_1, ..., C_q$ de poids $P_1, ..., P_q$, de centres de gravité $g_1, ..., g_q$ et de matrices de covariance $S_1, ..., S_q$ respectivement définis par :

$$P_k := \sum_{i \in C_k} p_i, \ \ k = 1, ..., q. \tag{2.1}$$

$$g_k = (g_{k1}, ..., g_{kp})' := \frac{1}{P_k} \sum_{i \in C_k} p_i e_i, \ \ k = 1, ..., q.$$

et

$$S_k := \frac{1}{P_k} \sum_{i \in C_k} p_i \left(e_i - g_k\right) \left(e_i - g_k\right)', \ \ k = 1, ..., q.$$

Les différentes observations sont résumées dans un tableau X défini, au premier chapitre, par (1.1) . On rappelle que le centre gravité global g et la matrice de covariance totale S sont définies par (1.2) et (1.6). Les classes $C_1, ..., C_q$ correspondent à q modalités d'une variable qualitative associée aux données, (voir [9], pages 440-441).

2.2 Décomposition de la variance

La matrice de covariance totale S se décompose en deux composantes, à savoir la variance inter-classe (between) et la variance intra-classe (within), (voir [8], pages 254-255).

2.2.1 Variance inter-classe (Between)

La matrice de covariance inter-classe B est définie comme la matrice de covariance des q centres de gravité $g_1, ..., g_q$:

$$B := \sum_{k=1}^{q} P_k (g_k - g)(g_k - g)' \in \mathcal{M}(p) .$$

Le terme général de la matrice B est donc :

$$b_{jj'} = \sum_{k=1}^{q} P_k (g_{kj} - \bar{x_j})(g_{kj'} - \bar{x_{j'}}). \tag{2.2}$$

2.2.2 Variance intra-classe (Within)

La matrice de covariance intra-classe W est définie comme la matrice moyenne des q matrices $S_1, ..., S_q$:

$$W := \sum_{k=1}^{q} P_k S_k \in \mathcal{M}(p) .$$

Le terme général de la matrice W est donc :

$$w_{jj'} = \sum_{k=1}^{q} \sum_{i \in C_k} p_i (x_{ij} - g_{kj})(x_{ij'} - g_{kj'}). \tag{2.3}$$

Proposition 2.2.1 *La covariance totale du nuage est égale à la somme des covariances inter-classe et intra-classe :*

$$S = W + B.$$

En d'autres termes, la somme des inerties inter-groupe et intra-groupe est équivalente à l'inertie totale du nuage des points individuels, (voir [5], pages 70-71).

Preuve. Le terme général $s_{jj'}$ de la matrice de covariance peut être réécrit, pour $j, j' = 1, ..., p$, comme suit :

$$\begin{aligned}
s_{jj'} &= \sum_{i=1}^{n} p_i(x_{ij} - \bar{x_j})(x_{ij'} - \bar{x_{j'}}) \\
&= \sum_{k=1}^{q} \sum_{i \in C_k} p_i(x_{ij} - \bar{x_j})(x_{ij'} - \bar{x_{j'}}) \\
&= \sum_{k=1}^{q} \sum_{i \in C_k} p_i \left[(x_{ij} - g_{kj}) + \left(g_{kj} - \bar{x_j}\right)\right)\left((x_{ij'} - g_{kj'}) + \left(g_{kj'} - \bar{x_{j'}}\right)\right] \\
&= \sum_{k=1}^{q} \sum_{i \in C_k} p_i \left[(x_{ij} - g_{kj})(x_{ij'} - g_{kj'}) + (x_{ij} - g_{kj})(g_{kj'} - \bar{x_{j'}})\right] \\
&+ \sum_{k=1}^{q} \sum_{i \in C_k} p_i \left[(g_{kj} - \bar{x_j})(x_{ij'} - g_{kj'}) + (g_{kj} - \bar{x_j})(g_{kj'} - \bar{x_{j'}})\right].
\end{aligned}$$

Or, on a :

$$(g_{kj} - \bar{x_j}) \sum_{i \in C_k} p_i(x_{ij'} - g_{kj'}) = (g_{kj'} - \bar{x_{j'}}) \sum_{i \in C_k} p_i(x_{ij} - g_{kj}) = 0.$$

D'où

$$\begin{aligned}
s_{jj'} &= \sum_{k=1}^{q} \left[\sum_{i \in C_k} p_i(x_{ij} - g_{kj})(x_{ij'} - g_{kj'}) + \sum_{i \in C_k} p_i(g_{kj} - \bar{x_j})(g_{kj'} - \bar{x_{j'}}) \right] \\
&= \sum_{k=1}^{q} \left[\sum_{i \in C_k} p_i(x_{ij} - g_{kj})(x_{ij'} - g_{kj'}) + (g_{kj} - \bar{x_j})(g_{kj'} - \bar{x_{j'}}) \sum_{i \in C_k} p_i \right].
\end{aligned}$$

En utilisant le fait que la somme des poids à l'intérieur de la classe C_k est égal (par définition) à son poids P_k, on obtient :

$$s_{jj'} = \sum_{k=1}^{q} \sum_{i \in C_k} p_i (x_{ij} - g_{kj})(x_{ij'} - g_{kj'}) + \sum_{k=1}^{q} P_k (g_{kj} - \bar{x_j})(g_{kj'} - \bar{x_{j'}}).$$

ce qui en vertu des relations (2.2) et (2.3) donne :

$$s_{jj'} = w_{jj'} + b_{jj'}, \ j, j' = 1, ..., p.$$

D'où la formule matricielle de la décomposition de la covariance totale. ■

2.3 Fonction discriminante

L'AFD vise à trouver de nouvelles variables discriminantes qui correspondent à des directions dans l'espace $\mathbb{R}^p$. Ces variables permettent de séparer au mieux les q groupes d'observations lorsqu'elles sont projetées. En d'autres termes, l'AFD cherche à identifier les combinaisons linéaires des variables quantitatives initiales qui optimisent la séparation entre les différentes classes définies par la variable qualitative explicative. ([9], pages 442-444).

2.3.1 Axes et variables discriminants

On munit $\mathbb{R}^p$ d'une métrique M et on projette (de façon M-orthogonales) les n points $(e_1, ..., e_n)$ de $\mathbb{R}^p$ sur un axe $\triangle_j$ de vecteur directeur a_j M-normé à 1. La liste des coordonnées $c_{ij} = \langle e_i, a_j \rangle_M$, $i = 1, ..., n$, des projections des individus sur $\triangle_j$ forme une nouvelle variable (artificielle) c_j :

$$c_j = (c_1, ..., c_n)' = XMa, \ j = 1, ..., p.$$

C'est une combinaison linéaire des colonnes de X, c-à-d des variables initiales.

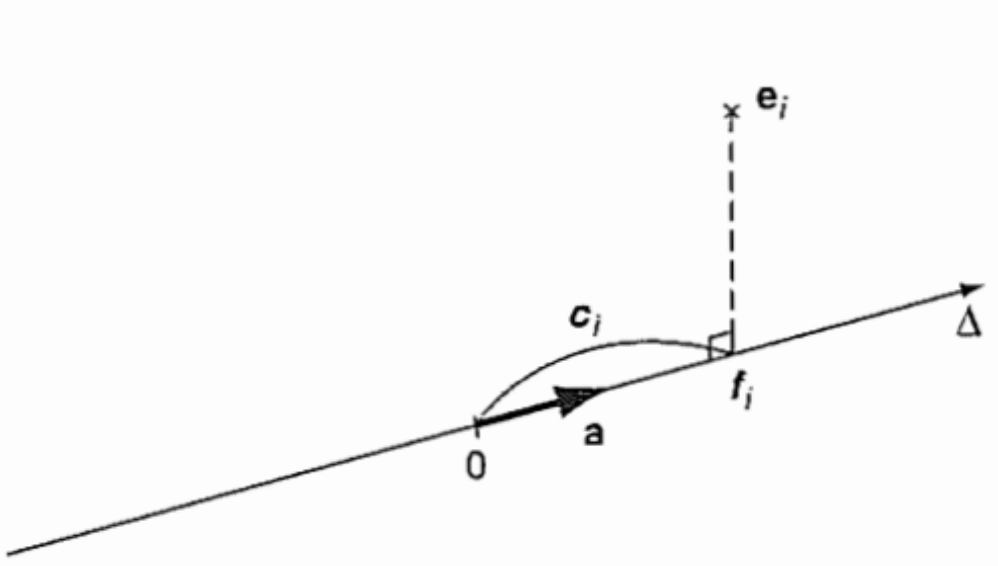

FIG. 2.1 – Projection d'un individu sur un axe ([9],page 161)

Remarque 2.3.1

1. *Le vecteur $a_j \in \mathbb{R}^p$ est appelé axe discriminant* $\left(\|a_j\|_M^2 = 1\right)$.
2. *On définit le vecteur $u_j := Ma_j \in \mathbb{R}^p$. On l'appelle facteur discriminant* $\left(\|u_j\|_{M^{-1}}^2 = 1\right)$.
3. *La variable c_j est appelée variable discriminante ou fonction linéaire discriminante.*

Remarque 2.3.2 (ACP particulière) *L'AFD peut être considérée comme l'ACP du nuage des q centres de gravité avec la métrique $M = S^{-1}$. Les axes, facteurs et variables discriminants correspondent aux axes, facteurs et composantes principaux de cette ACP. On conclut que les variables discriminantes sont non corrélées 2 à 2.*

2.3.2 Critère de discrimination

L'inertie en projection, sur l'axe $\triangle_j$, du nuage des centres de gravité $g_1, ..., g_q$ est égale à $u'Bu$. On cherche à la rendre maximale tout en minimisant l'inertie (en projection)

intra-classe $u'Wu$. Ceci revient à maximiser le rapport d'inertie

$$\frac{u'Bu}{u'Wu}.$$

L'égalité de la décomposition des covariances $S = B + W$ entraîne que :

$$\arg\max \frac{u'Bu}{u'Wu} = \arg\max \frac{u'Bu}{u'Su}.$$

Donc, la fonction à maximiser est le rapport de la variance inter-classe B à la variance totale S.

Proposition 2.3.1 *Le maximum est atteint pour u égal au vecteur propre de la matrice $S^{-1}B$ associé à la plus grande valeur propre.*

Preuve. La fonction

$$f(u) = \frac{u'Bu}{u'Su},$$

à maximiser est sous la forme d'un quotient de deux formes quadratiques, à savoir $u'Bu$ et $u'Su$. La solution à ce problème peut être trouvée dans plusieurs références dont ([9], page 607).

On annulle la dérivée de f, qui est égale à :

$$f'(u) = \frac{(u'Su)(2Bu) - (u'Bu)(2Su)}{(u'Su)^2}.$$

Ceci revient à résoudre l'équation :

$$(u'Su)Bu - (u'Bu)Su = 0 \Longleftrightarrow (u'Su)Bu = (u'Bu)Su.$$

En divisant les deux membres par le nombre réel $u'Su$ et en les multipliant par la matrice inverse de S, on obtient :

$$S^{-1}Bu = \lambda u.$$

où $\lambda := u'Bu/u'Su \in \mathbb{R}$. Donc, la solution u est un vecteur propre de $S^{-1}B$ associé à la valeur propre λ. En d'autres termes, le maximum de la fonction f est atteint en u égal à ce vecteur propre (c-à-d pour $Bu = \lambda Su$) :

$$\max_u f(u) = \frac{u'(\lambda Su)}{u'Su} = \lambda \frac{u'Su}{u'Su} = \lambda.$$

On conclut que la valeur propre correspondant à la solution est la plus grande parmi les valeurs propres de $S^{-1}B$. ■

Interprétation des valeurs propres

On range, par ordre décroissant, les p valeurs propres de la matrice $S^{-1}B$:

$$\lambda_1 > \lambda_2 > ... > \lambda_p.$$

Le premier axe (facteur) discriminant correspond à λ_1, le deuxième correspondra à λ_2, et ainsi de suite. La valeur propre λ_j permet de mesurer le pouvoir de discrimination de l'axe a_j, $j = 1, ..., p$.

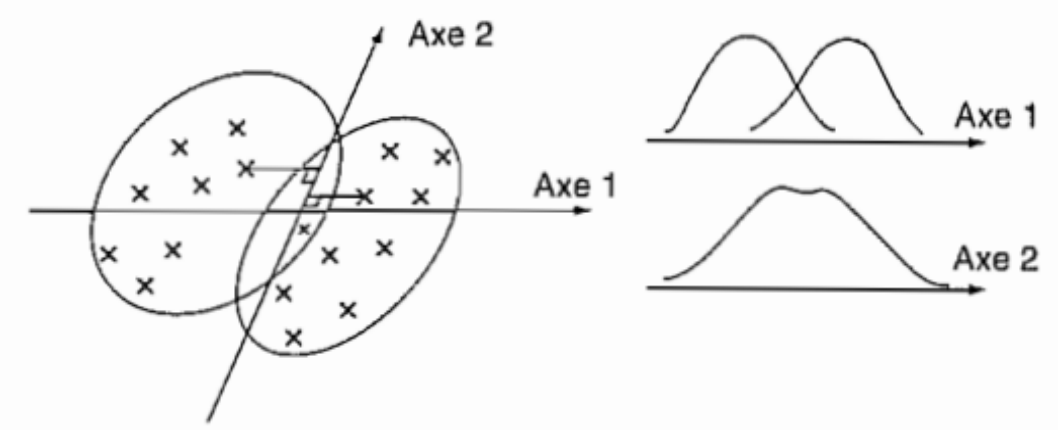

FIG. 2.2 – Axe 1 et 2 ([9], page 442)

Sur la figure 2.2, on voit que l'axe 1 possède un bon pouvoir de discrimination alors que l'axe 2 ne permet pas de séparer en projection les 2 groupes.

- On a $0 \leq \lambda \leq 1$
- $\lambda = 1$: en projection sur Δ, les dispersions intra-classes sont nulles, les q nuages sont donc chacun dans un hyperplan orthogonal à a. Il y a discrimination parfaite si les centres de gravité se projettent en des points différents.

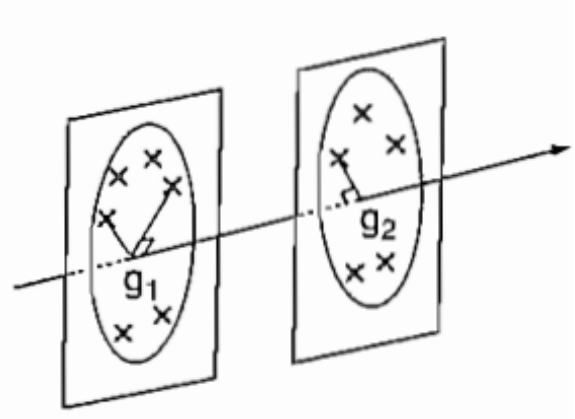

FIG. 2.3 – Séparation des centres de gravité ([9])

- $\lambda = 0$: le meilleur axe ne permet pas de séparer les centres de gravité g_j. C'est le cas où ils sont confondus. Les nuages sont donc concentriques et aucune séparation linéaire n'est possible.

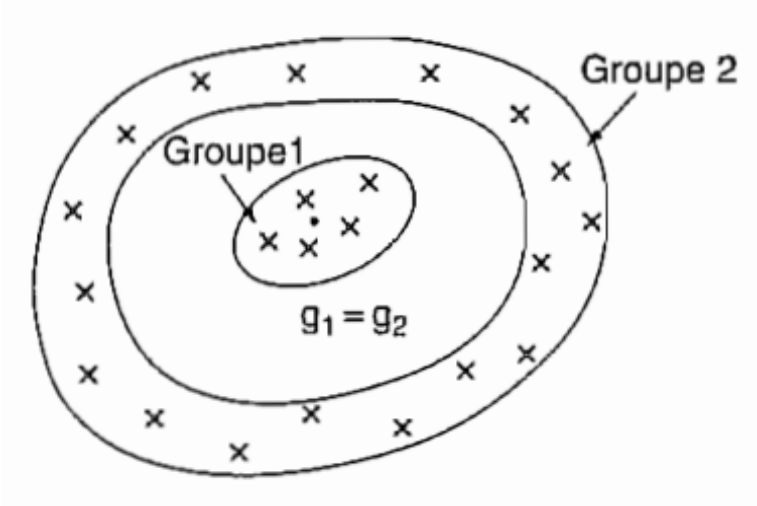

FIG. 2.4 – Nuages concentriques ([9])

2.4 Méthodes de classement

Soit $e = (w_1, ..., w_p)' \in \mathbb{R}^p$ un nouvel individu. Le but est d'assigner (affecter) cet individu à une des q classes $C_1, ..., C_q$ représentées par leurs centres de gravité respectifs $g_1, ..., g_q$.

2.4.1 Méthode géométrique

C'est la méthode de classement la plus courante. Elle consiste à calculer les distances entre e et chacun des centres de gravités $g_1, ..., g_q$ et à affecter e au groupe dont la distance est la plus petite. Cependant, il est essentiel de définir la métrique à utiliser pour mesurer ces distances.(voir [9], page 447).

La métrique la plus couramment utilisée est la matrice W^{-1}(ou de façon équivalente S^{-1}). On l'appelle métrique de Mahalanobis et la distance correspondante est appelée distance de Mahalanobis.

$$d^2(e, C_k) := d^2(e, g_k) = (e - g_k)' W^{-1}(e - g_k), \ k = 1, ..., q. \tag{2.4}$$

Ainsi, la règle de décision est la suivante : on attribue l'individu e à la classe C_m pour laquelle on a :

$$d^2(e, g_m) = \min_{1 \leq k \leq q} d^2(e, g_k). \tag{2.5}$$

En développant les quantités définies dans l'équation (2.4) , on obtient :

$$d^2(e, g_k) = e' W^{-1} e + g_k' W^{-1} g_k - 2 g_k' W^{-1} e, \ k = 1, ..., q.$$

On note que le terme $e' W^{-1} e$ ne dépend pas k. On définit la fonction linéaire discriminante (en e) par :

$$\ell_{C_k}(e) := g_k' W^{-1} g_k - 2 g_k' W^{-1} e. \tag{2.6}$$

La règle de décision (2.5) se reformule comme suit : on choisit d'attribuer e à la classe C_m tellque :

$$\ell_{C_m}(e) = \min_{1 \leq k \leq q} \ell_{C_k}(e).$$

2.4.2 Méthode bayésienne

Les q classes $C_1, ..., C_q$ se trouvent dans l'échantillon en proportions $P_1, ..., P_q$ qui représentent leurs poids respectifs (2.1) . On suppose que la probabilité d'un vecteur aléatoire $U = (U_1, ..., U_p)'$ est donnée pour chaque classe k par une densité (ou loi discrète) $f_k(u)$. En observant un point u, la probabilité qu'il appartienne à la classe C_k est exprimée par la formule de Bayes :

$$P\left(C_k/u\right) = \frac{P_k f_k(u)}{\sum_{k=1}^{q} P_k f_k(u)}, \ k = 1, ..., q.$$

P_k est appelée probabilité à priori de la classe C_k, alors que $P\left(C_k/u\right)$ représente la probabilité a postériori de la même classe. (voir [9], pages 467-468).

Ainsi, la règle bayésienne de décision est la suivante : on affecte l'individu u à la classe qui a la plus forte probabilité à postériori. En d'autres termes, on attribue u à la classe C_m pour laquelle on a :

$$P\left(C_m/u\right) = \max_{1 \leq k \leq q} P\left(C_k/u\right), \tag{2.7}$$

ce qui revient à maximiser $P_k f_k(u)$: $\max_{1 \leq k \leq q} P_k f_k(u)$.

Règle bayésienne avec modèle gaussien

Dans ce cas, on suppose que le vecteur U est de distribution multinormale dans chaque classe C_k :

$$U \sim \mathcal{N}_p(\mu_k, \Sigma_k), \ k = 1, ..., q.$$

Les densités de probabilité f_k sont donc de la forme :

$$f_k(u) = \frac{1}{(2\pi)^{P/2}\left(\det(\Sigma_k)\right)^{1/2}} \exp(-\frac{1}{2}(u-\mu_k)'\Sigma_k^{-1}(u-\mu_k)).$$

Puisque la fonction logarithme est croissante, on maximise $\ln\left(P_k f_k(u)\right)$. On a :

$$\ln\left(P_k f_k(u)\right) = -\frac{1}{2}(u-\mu_k)'\Sigma_k^{-1}(u-\mu_k) + \ln P_k - \frac{1}{2}\ln\left(\det(\Sigma_k)\right) - \frac{p}{2}\ln(2\pi)$$

De façon équivalente, la règle bayésienne revient donc à minimiser (par rapport à k) la quantité :

$$(u-\mu_k)'\Sigma_k^{-1}(u-\mu_k) - 2\ln P_k + \ln(\det\Sigma_k),$$

appelée (carré) de la distance de Mahalanobis théorique généralisée.

2.5 Cas de deux classes

2.5.1 Analyse linéaire discriminante

L'équation des valeurs propres et des vecteurs propres en analyse linéaire discriminante est :

$$Bu = \lambda Su. \iff S^{-1}Bu = \lambda u;$$

La matrice de covariance inter-classe B dans le cas de deux groupes est une matrice d'ordre p et de rang 1 dont le terme rénéral est :

$$b_{jj'} = P_1(g_{1j} - \bar{x_j})(g_{1j'} - \bar{x_{j'}}) + P_2(g_{2j} - \bar{x_j})(g_{2j} - \bar{x_{j'}}) = P_1P_2(g_{1j} - g_{2j})(g_{1j'} - g_{2j'}).$$

Donc peut B se mettre sour la forme $B = mm'$ ou m est une matrice colonne $(p, 1)$ de terme général

$$m_j = \sqrt{P_1P_2}(g_{1j} - g_{2j}).$$

Ainsi l'unique valeur propre de $S^{-1}B$ est $\lambda = m'S^{-1}m$ et le vecteur propre associé est $u = S^{-1}m$. ([6], page 14).

2.5.2 Règle géométrique d'affectation

L'individu e est affecté au groupe le plus proche. La comparaison des distances de e à C_1 et C_2 revient à la comparaison des distances de e aux points moyens g_1 et g_2, puisque les deux groupes sont supposés également dispersés autour de leurs deux points moyens. ([6], page 21). On affecte e à C_1 si $d^2(e, C_1) < d^2(e, C_2)$ ou $\ell_{C_1}(e) < \ell_{C_2}(e)$, c-à-d si :

$$\ell_{C_2}(e) - \ell_{C_1}(e) > 0, \tag{2.8}$$

où $d^2(e, C_k)$ et $\ell_{C_k}(e)$ sont definies par (2.4) et (2.6) respectivement.
En remplaçant $\ell_{C_1}(e)$ et $\ell_{C_2}(e)$ par leurs valeurs spécifiées en (2.6) , on a :

$$\begin{aligned}
\ell_{C_2}(e) - \ell_{C_1}(e) &= g_2'W^{-1}g_2 - g_1'W^{-1}g_1 - 2g_2'W^{-1}e + 2g_1'W^{-1}e \\
&= g_2'W^{-1}g_2 - g_1'W^{-1}g_1 + 2(g_1 - g_2)'W^{-1}e \\
&= 2(g_1 - g_2)'W^{-1}e - (g_1 - g_2)'W^{-1}(g_1 + g_2).
\end{aligned}$$

En conclusion, la règle d 'affectation (2.8) devient, après la division par 2, comme suit :

$$h(e) : \begin{cases} > 0 : & \text{affectation à } C_1, \\ < 0 : & \text{affectation à } C_2, \end{cases}$$

où,

$$h(e) = (g_1 - g_2)'W^{-1}e - \frac{1}{2}(g_1 - g_2)'W^{-1}(g_1 + g_2),$$

est appelée fonction lineaire discriminante de Fisher.

2.5.3 Règle bayésienne avec modèle gaussien

([9], page 469) On affectera l'individu u au groupe C_1 si :

$$u'\Sigma^{-1}(\mu_1 - \mu_2) > \frac{1}{2}(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) + \ln\frac{P_2}{P_1},$$

Score d'Anderson

Le score (ou statistique) d'Anderson est une fonction définie sur $\mathbb{R}^p$ par :

$$S(u) := u'\Sigma^{-1}(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1 + \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2) - \ln\frac{P_2}{P_1}.$$

Elle est simplement liée à la probabilité a postériori d'appartenir au groupe C_1. En termes du score la règle de décision devient :

$$S(u) : \begin{cases} > 0 : & \text{affectation à } C_1, \\ < 0 : & \text{affectation à } C_2, \end{cases}$$

Fonction logistique du score

La fonction logistique du score est la probabilité à postériori d'appartenir au groupe $P(C_1/u)$:

$$P := P(C_1/u) = \frac{P_1 f_1(u)}{P_1 f_1(u) + P_2 f_2(u)},$$

On a :

$$\frac{1}{P} = 1 + \frac{P_2 f_2(u)}{P_1 f_1(u)} = 1 + \frac{P_2}{P_1}\exp(-\frac{1}{2}(u + \mu_2)'\Sigma^{-1}(u - \mu_2) + \frac{1}{2}(u + \mu_1)'\Sigma^{-1}(u - \mu_1)),$$

En appliquant le logarithme, on obtient :

$$\ln(\frac{1}{P} - 1) = -S(u),$$

ce qui implque la forme suivante de la fonction logistique du score :

$$P = \frac{1}{\exp(-S(u)) + 1} = \frac{\exp(S(u))}{1 + \exp(S(u))}.$$

On exprime la règle de décision en termes de la fonction logistique du score comme suit :

$$P : \begin{cases} > \frac{1}{2} : & \text{affectation à } C_1, \\ < \frac{1}{2} : & \text{affectation à } C_2. \end{cases}$$

2.6 Etude de cas : Infarctus du myocarde

Les données dans cet exemple concernent 101 victimes d'infarctus du myocarde où 50 survivront et 51 décèderont (les données sont prises [9], pages $453 - 454$) sur lesquels ont été mesurées 7 variables quantitatives à leur admission dans un service de cardiologie. Ces variables sont : fréquence cardiaque (FRCAR), index cardiaque (INCAR), index systolique (INSYS), pression diastolique (PRDIA), pression artérielle pulmonaire (PAPUL), pression ventriculaire (PVENT), résistance pulmonaire (REPUL).

2.6.1 Réprésentation des groupes sur le plan principal d'une ACP

La figure (2.5) représente la séparation de l'échantillon sur le premier plan principal :

2.6.2 Statistiques descriptives univariées

Comme pour toute analyse des données, on débute toujours par des statistiques univariées (global et par classes) qui sont étudiées sur les variables d'analyse. L'objectif est de déterminer le rôle d'une variable sur la variable explicative.

Le tableau (2.1) répresente les statistiques univariées globales des sept variables. Dans les tableaux (2.2) et (2.3), on résume les moyennes et écart-types par classe de chacune des

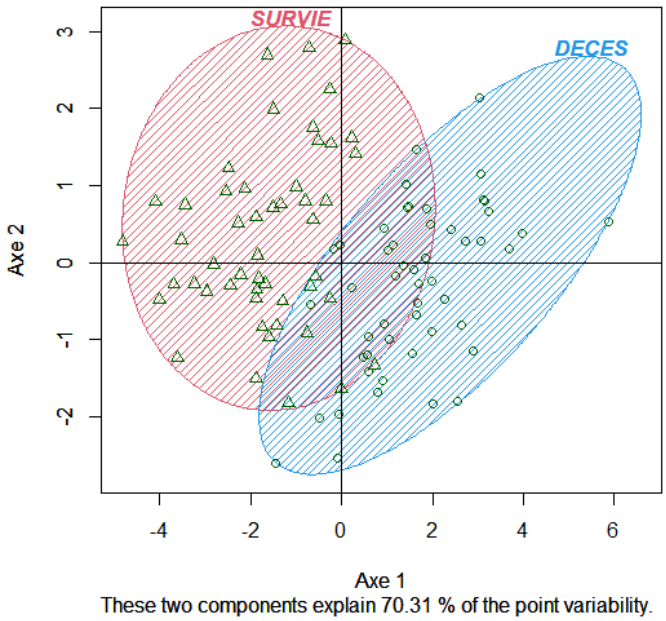

FIG. 2.5 – Premier plan principal de discrimination

sept variables explicatives.

2.6.3 Taux d'erreur empirique

Une règle de décision est évaluée au moyen de ce que l'on appelle taux d'erreur. Plus ce dernier est faible, plus la décision est bonne. Empiriquement, il est obtenu à partir de la matrice de confusion, qu'on construit en réaffectant un échantillon des observations initiales. Le taux d'erreur empirique est égal à la somme du nombre d'individus mal classés divisé par la taille de l'échantillon. Le complément à 1 est appelé taux d'efficacité empirique de la règle.

Le reclassement de classement de 21 individus donne la matrice de confusion (2.4) , à partir de laquelle on déduit que le nombre d'individus mal classés est $0 + 2 = 2$. Par conséquent, le taux d'erreur empirique est égal à 0.095, ce qui représente un pourcentage d'environ 10% de mal classés, correspondant à une efficacité de 90%.

	Min	Quart 1	Mediane	Moyenne	Quart 3	Max	Écart-type
FRCAR	51.00	80.00	90.00	92.16	102.00	135.00	16.43
INCAR	0.60	1.34	1.76	1.85	2.32	3.37	0.66
INSYS	5.20	14.80	19.60	20.82	26.30	54.00	8.81
PRDIA	8.00	15.00	19.00	19.26	24.00	34.00	5.81
PAPUL	10.00	20.50	26.00	26.00	31.00	46.00	7.32
PVENT	1.00	6.50	10.00	9.50	11.50	20.00	4.34
REPUL	345.00	807.00	1131.00	1324.00	1610.00	5067.00	741.34

TAB. 2.1 – Staisques uni-variées global

	FRACAR	INCAR	INSYS	PRDIA	PAPUL	PVENT	REPUL
SURVIE	88.34	2.31	26.75	16.50	22.84	8.33	841.38
DECES	95.90	1.39	15.00	21.96	29.10	10.65	1797.27

TAB. 2.2 – Moyenne par classe

	FRACAR	INCAR	INSYS	PRDIA	PAPUL	PVENT	REPUL
SURVIE	13.84	0.56	8.08	5.15	6.47	4.05	303.68
DECES	17.98	0.38	4.64	5.14	6.82	4.34	739.87

TAB. 2.3 – écart-type par classe

	Réel	
Prévu	DECES	SURVIE
DECES	8	2
SURVIE	0	11

TAB. 2.4 – Matrice de confusion

Conclusion

Dans ce mémoire, on a présenté l'une des méthodes de statistique exploratoire les plus couramment utilisées de nos jours : l'analyse factorielle discriminante. La facilité de sa mise en œuvre la rend répandue dans de nombreux domaines de la vie socio-économique. Elle convient parfaitement à la représentation des données dans des espaces qui permettent de mieux discriminer les individus selon des classes prédéfinies. Cette représentation offre la possibilité d'extraire des informations à partir d'un grand nombre de données qui sont complexes à interpréter. De plus, elle permet de classer de nouveaux individus dans des groupes existants.

Bibliographie

[1] Bouchier, A. (2010). L'analyse des données multivariées à l'aide du logiciel. L'analyse discriminante linéaire (LDA). Montpellier.

[2] Chessel, D., Dufour, A. B., & Lobry, J. (2007). Analyse discriminante linéaire. Lyon.

[3] Lecoutre, J.P. (2012). Statistique et probabilités. Dunod, Paris.

[4] Lejeune, M. (2010). Statistique, la théoie et ses applications. Springer.

[5] Monbet, V. (2013). Analyse des données Master Statistique et économétrie Notes de cours, cours de $1^{ère}$ année Master.

[6] Nakache, J. P., & Confais, J. (2003). Statistique explicative appliquée : analyse discriminante, modèle logistique, segmentation par arbre. Editions Technip.

[7] Necir, A. (2022). Analyse en composantes principales(Modèle lineaire), cours de $1^{ère}$ année Master, Université de Mohamed Khider Biskra.

[8] Piron, M., Lebart, L., & Morineau, A. (1995). Statistique exploratoire multidimensionnelle. Paris.

[9] Saporta, G. (2011). Probabilités, analyse des données et statistique. Technip, Paris.

Annexe : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

Notation	**Signification**
ACP	Analyse en composantes principales.
AFD	Analyse factorielle discriminante.
c-à-d	C'est-à-dire.
1_n	Vecture unitaire d'ordre n.
I_n	Matrice identité d'ordre n.
$cov(.,.)$	Covariance.
D	Matrice des poids.
$D_{1/s}$	Matrice diagonal des inverse des écart-types.
$d(e_i, e_{i'})$	Distance entre e_i et $e_{i'}$.
e_i	$i^{ème}$ individu.
g	Centre de gravité.
M	Métrique.
R	Matrice de corrélation.
$r_{jj'}$	coefficient de corrélation.
S	Matrice de covariance.

Notation	**Signification**
$s_{jj'}$	Covariance.
s_j^2	variance.
s_j	Ecart-type.
$var(.)$	Variance.
X	Tableau des données.
x_j	$j^{ème}$ variable.
Y	Tableau centré.
Z	Tableau standard.
$\| \, . \, \|_M$	Norme par rapport à la métrique M.
$\langle ., . \rangle_M$	Produit scalaire par rapport à la métrique M.

ملخص

تحليل العوامل التمييزية هي تقنية إحصائية وصفية و تنبؤية تستخدم لتصنيف الملاحظات إلى مجموعات محددة مسبقا على أساس عدة متغيرات.

في هذه المذكرة، قدمنا التعاريف والخصائص الأساسية المتعلقة بتحليل العوامل التمييزية مع تطبيقها على بيانات حقيقية، والتي تتكون من التدابير الطبية على ضحايا احتشاء عضلة القلب.

الكلمات المفتاحية: تصنيف؛ التمييز؛ طريقة هندسية؛ طريقة بايز؛ نموذج غوس؛ التباين بين الفئات؛ التباين داخل الفئات.

Abstract

Discriminant Factor Analysis (DFA) is a descriptive and predictive multivariate statistical technique used to classify observations into predefined groups based on multiple variables. In this dissertation, definitions and fundamental properties related to DFA were presented with application on real data, consisting of medical measurements on victims of myocardial infarction.

Key words : Bayesian method; Classification; Discrimination; Gaussian model; Geometric method; Interclass variance; Intraclass variance; Mahalanobis distance.

Résumé

L'analyse factorielle discriminante (AFD) est une technique statistique descriptive et prédictive multivariée, utilisée pour classer des observations dans des groupes prédéfinis en fonction de plusieurs variables.

Dans ce mémoire, on a présenté les définitions et propriétés fondamentales relatives à l'AFD avec application sur des données réelles, qui consistent en des mesures médicales sur des victimes d'infarctus du myocarde.

Mots clés : Classement ; Discrimination ; Distance de Mahalanobis ; Méthode géométrique ; Méthode bayésienne ; Modèle gaussien ; Variance interclasse ; Variance intraclasse.