République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la Recherche Scientifique
Université Mohammed Seddik Ben Yahia - Jijel

Faculté des Sciences Exactes et Informatique
Département des Mathématiques

№ d'ordre : .....

№ de séries : .....

Mémoire de fin d'études

Présenté pour l'obtention du diplôme de

Master

Filière : Mathématiques.

Spécialité : Probabilités et Statistiques.

Thème

Analyse factorielle discriminante

Présenté par :

BOUCHICHA Imen

DIF Roumaissa Nour El Yakine

Devant le jury :

Président	: **BOUDJARDA Khawla**	MAB Université de Jijel
Encadreur	: **CHERAITIA Hassen**	MCB Université de Jijel
Examinateur	: **LAOUDJ Farida**	MCA Université de Jijel
Examinateur	: **YAKOUBI Fatma**	MAA Université de Jijel

Promotion **2017/2018**

Remerciements

Tout d'abord, nous remercions Dieu, notre Créateur, le Miséricordieux, qui nous a donné l'opportunité d'étudier, la volonté, le courage et la patience afin d'accomplir et de mener à bien ce travail.

Nous remercions en particulier notre encadreur, Mr. **CHERAITIA Hassen** pour avoir dirigé ce travail. Pour son aide, ses encouragements, sa grande disponibilité, ses précieux conseils et pour la patience qu'il nous a accordé pendant la réalisation de ce travail.

Un grand merci également aux membres du jury, Mme.la présidente **BOUDJARDA Khawla** , Mms. les examinatrices **LAOUDJ Farida** et **YAKOUBI Fatma** pour l'honneur qu'elles nous ont fait en acceptant de juger notre mémoire.

Nos vifs remerciements vont à tous les enseignants qui nous ont suivis durant nos cinq années d'études à l'université, à tous les camarades de notre promotion 2018.

Nous adressons nos remerciements les plus chaleureux à nos familles, nos amis pour leur patience et leur intérêt.

Enfin, nous remercions toutes les personnes qui auraient contribué d'une manière ou d'une autre à la réalisation de ce travail.

♡.....**ROUMAISSA.IMAN**♡

Dédicace

Un grand merci au bon Dieu, le seigneur des mondes pour le courage et la force qui nous a offert pour terminer ce mémoire.

Je dédie ce modeste travail :

À celle qui m'a aidé par ses sincères prières et douaa à la plus chère personne de ma vie ***ma mère****.À celui qui a bien travaillé pour m'apprendre c'est quoi le combat et qui m'a fait ce que je suis* ***mon chère père*** *que Dieu le protegè pour nous.*

♡

♡ ***À mes frères : FoFo, Oussama*** *Je vous dédie ce travail et veuillez trouver dans ce mémoire l'expression de mon respect.*

À mes chéres Mama Nawara et DoDa *Meilleurs voeux de bonheur dans ta vie.*

Àmes sœurs : Nessrinr, Lamiss, Chaima, Hassina *Meilleurs vœux de sucées dans ses études et de bonheur dans leurs vie.*

À mes tantes : Rachida, Naima, Samira, Ibtissam *et leurs enfants* ***Jouri, Mina, Rahil, Rahma.***

À mes tentants : Kamel, Omer et Ahmed.

À ma collègue Roumaissa : *que Dieux réunisse nos chemins pour un long commun serein et que ce travail soit témoignage de ma profonde reconnaissance et de mon amour sincère et fidèle.*

À mes chères amies : *Un dédicace particulier est sincère pour mes amies* ***Meryam, Sara, Samira, Ahlem, Widad*** *Je vos souhaite une vie pleine de joie et de prospérité.*

À la fin, je prie le bon Dieu de faire ce travail très utile pour les autres candidats de cette spécialité ♡.

♡.....**IMAN**♡

Dédicace

Un grand merci au bon Dieu, le seigneur des mondes pour le courage et la force qui nous a offert pour terminer ce mémoire.

Je dédie ce modeste travail :

À celle qui m'a aidé par ses sincères prières et douaa à la plus chère personne de ma vie ***ma mère****.À celui qui a bien travaillé pour m'apprendre c'est quoi le combat et qui m'a fait ce que je suis* ***mon chère père*** *que Dieu le protegè pour nous.*

♡

♡

À mes frères : Mouadh, Houdaifa*.*

Je vous dédie ce travail et veuillez trouver dans ce mémoire l'expression de mon respect.

À ma sœur : Roufaida *.*

Meilleurs vœux de sucées dans tes études et de bonheur dans ta vie.

À ma collègue Iman*.*

Que Dieux réunisse nos chemins pour un long commun serein et que ce travail soit témoignage de ma profonde reconnaissance et de mon amour sincère et fidèle.

À mes chères amies*.*

Un dédicace particulier est sincère pour mes amies : "***Wafa***""***Hanen***"

Je vos souhaite une vie pleine de joie et de prospérité.

À la fin, je prie le bon Dieu de faire ce travail très utile pour les autres candidats de cette spécialité ♡*.*

♡.....**ROUMAISSA NOUR EL YAKINE**♡

Table des matières

Table des figures

Liste des tableaux

Introduction générale

L'analyse factorielle discriminante (AFD) est une méthode ancienne(RONALD FISHER, 1936) qui dans sa version classique a peu évolué au cours des vingt dernières années. Cette méthode est une technique d'analyse des données qui est descriptive et explicative qui vise à décrire, expliquer et prédire l'appartenance d'un individu statistique à un groupe prédéfini (classe, modalité de la variable à prédire).

L'analyse discriminante est utilisée dans nombreux domaines :
En médecine, par exemple pour détecter les groupes à haut risque cardiaque à partir de caractéristique telles que l'alimentation, le fait de fumer ou pas, les antécédents familiaux...etc
Dans le domaine financier, lorsque l'on veut évaluer la fiabilité d'un demandeur de crédit à partir de ses revenus, du nombre de personnes à charge, des encours de crédit qu'il détient...etc
En biologie, lorsque l'on veut affecter un objet à sa famille d'appartenance à partir de ses caractéristiques physiques. Les iris de Sir Ronald Fisher (qui est à l'origine de cette méthode), il s'agit de reconnaître le type d'iris (setosa, virginica, et versicolor) à partir de la longueur/largeur de ses pétales et sépales.
En informatique, pour la reconnaissance optique de caractères. L'analyse discriminante est utilisée pour reconnaître un caractère imprimé à partir d'informations simples, comme la présence ou non de symétrie, le nombre d'extrémités.

Vu la variété des domaines d'application de AFD, que nous somme motivés de comprendre cette méthode afin de répondre aux questions suivantes :
• Quelles variables, quels groupes de variables, quels sous-espaces discriminent-ils au mieux les groupes (les classes) ?
• A partir de ces variables quantitatives, peut-on décider de la classe (groupe) à laquelle appartient la nouvelle individus ?

Nous visons dans ce travail à :
- L'aspect descriptif : Déterminer les fonctions linéaires discriminantes sur l'échantillon d'apprentissage, c-à-d la combinaison linéaire des p variables explicatives dont les valeurs séparent au mieux les q classes (qui prennent des valeurs les plus proches possible pour des éléments de la même classe, et les plus éloignées possible entre éléments de classes différentes).
- L'aspect décisionnel : Déterminer la classe de nouveaux individus pour lesquels nous observons les valeurs des p variables explicatives. Cette étape est une étape d'affectation d'un nouvel individu dans une classe. Il s'agit d'un problème de classement par opposition au problème de classification qui est la construction de classes les plus homogènes possibles dans un échantillon.

Pour bien atteindre nos objectifs et afin de répondre à nos questions, nous avons structuré notre travail comme suit :

• dans **le premier chapitre** intitulé "la méthode géométrique", nous présentons brièvement quelques notations et présentations des données. Ensuite, nous présentons la fonction linéaire discriminante, la méthode disqual et la règle géométrique d'affectation afin de montrer l'intérêt de cette méthode géométrique dans l'analyse factorielle discriminante. Nous terminons ce chapitre par les limites et les extensions de cette méthode.

• dans **le deuxième chapitre** intitulé "Méthode probabiliste" nous abordons la méthode probabiliste, nous commençons par expliquera la règle bayésienne. Puis ses différents types, avec modèle normale et avec estimations non paramétrique. Nous parlons par la suite de la qualité des règles de classement en savoir la validation par Resubstitution, échantillon test et validation croisée et nous clôturons ce chapitre par la réduction de nombre des variables en donnant deux méthodes différentes (passage par l'ACP et démarche du pas à pas).

• **le troisième chapitre** intitulé "Applications" réservé à l'application de tous qui est cités dans les deux chapitres précédents sur un exemple réel(victimes d'infarctus de myocarde). nous avons fini ce chapitre par une comparaison entre les deux approches géométrique et probabiliste en terme d'efficacité et de la qualité de classement.

A la fin, une conclusion générale dresse une synthèse des principaux résultats obtenus au cours de notre travail.

Chapitre 1

Méthode géométrique

1.1 Introduction

Cette méthodes de recherche, essentiellement descriptive, ne reposent que sur des notions de distance et ne font pas intervenir d'hypothèses probabilistes, celle dont les représentations graphiques des individus discriminent "au mieux" les q classes engendrées par la variable X.
L'idée de base est très simple. Il s'agit de calculer la distance (définie par distance de Mahalanobis) entre la nouvelle observation et le centre de chacun des groupes. On classera la nouvelle observation dans le groupe pour lequel cette distance est minimale. Pour cela, on va définir et étudier dans ce chapitre la fonction linéaire discriminante et des règles de décision (ou d'affectation) géométrique et donner ensuite les extensions et limites de cette règle.

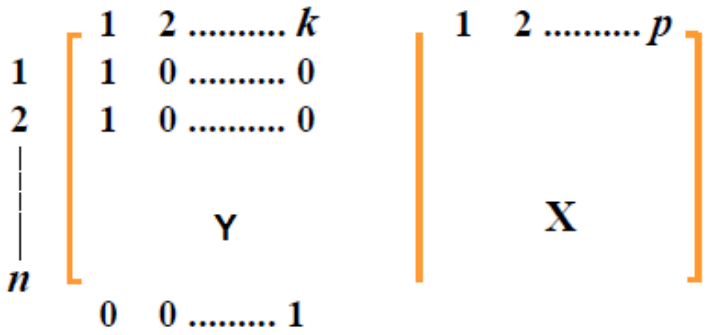
1.2 Notations et données

1.2.1 Notions

- Y : variable qualitative de q modalités jouant le rôle de la variable à expliquer comme dans le modèle linéaire
- X : tableau de p variables quantitative (X_1,...,X_p) jouant le rôle des variables explicatives
- x_{ij} : valeur de la variable X_j
- G_k : classe(groupe) des individus
- g : centre de gravité totale
- g_j : centre de gravité de chaque classe
- n : nombre total d'observations

- q : nombre de groupes
- n_k : nombre d'observations(individus) dans le groupe k
- p_i : poids de chaque individu
- q_j : poids de chaque nuage
- D : axe discriminant
- e_i : observations
- x : nouvelle observation de groupe inconnu.

Tableau des données :



- Y : matrice des indicatrices de la variable qualitative à prédire
- X : matrice de prédicteurs (variables explicatives quantitatives).

1.2.2 Données statistiques

On dispose de n individus où observations décrit par un ensemble de p variables $(X_1,...,X_p)$ et répartit en q classes définies a priori par le rang nominale à p modalités, les n individus e_i de l'échantillon constituent un nuage E des points de $\mathbb{R}^p$ partagé en q sous nuages $(E_1,\ldots,E_k,\ldots E_q)$ de centre de gravité $(g_1,\ldots,g_k,\ldots,g_q)$ et de matrice de variance $(V_1,\ldots,V_k,\ldots,V_q)$.
Soit $\bar{g}$ le centre de gravité global et V la variance totale de E. Si les n individus e_i sont affectés des poids $p_1,...,p_n$ et les poids $q_1,...,q_q$ de chaque nuage alors :

$$g = (g^1,g^2,...,g^p) \quad \text{et} \quad g^j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

$$g_k = (g_{k1},g_{k2}...,g_{kp}) \quad \text{et} \quad g_k^j = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ij}.$$

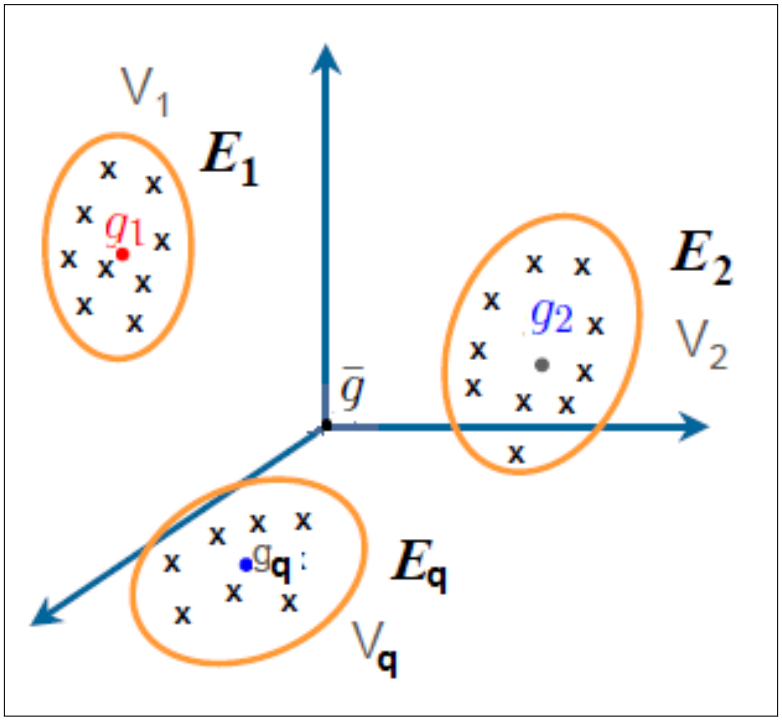


FIGURE 1.1 – Représentation des nuages des points

1.3 Fonction linéaire discriminante

1.3.1 Objectif de la fonction linéaire discriminante

Il s'agit de trouver une nouvelle variable, combinaison linéaire des variables explicatives, qui "discrimine" au mieux les groupes définis par les modalités de la variable à expliquer Y.
On pose : $S = X.a = a_1.x^1 + ... + a_p.x^p$ où $a = (a_1, ..., a_p)' \in \mathbb{R}$ est le vecteur des coefficients de cette combinaison linéaire.

1.3.2 Décomposition de la matrice variance-covariance

Nous pousuivons donc l'étude à partir des indices numériques que nous avons commencé à mettre en évidence, on débute l'analyse en étendant la formule de la variance totale au cas vectoriel, si en effet X représente le vecteur formé des variables X_j tel que $1 \leqslant j \leqslant p$, il y a deux types de variance : variance inter-classe (between) et variance intra-classe (within).
Soient :
V, W et B les matrices de taille $(p.p)$ de termes respectifs $v_{jj'}$, $w_{jj'}$ et $b_{jj'}$.
- V : est la matrice de variance-covariance du nuage E.
- B : la matrice de variance-covariance inter-classe (between).

- W : la matrice de variance-covariance intra-classe (within).

Variance inter-classe (between) :

B matrice de variance-covariance inter-classe (between) :

• C'est la matrice de variance-covariance du nuage constitué des points moyens g_{kj} des groupes E_k affectés des poids respectif $\frac{n_k}{n}$.

• C'est la matrice des q centres de gravité

$$B = \frac{1}{n}\sum_{k=1}^{q} n_k(g_{kj} - g_j)(g_{kj'} - g_{j'})'$$

• Elle rend compte alors de la dispersion des centres de gravité des classes autour du centre global g.

Variance intra-classe (within) :

W la matrice de variance-covariance intra-classe (within) :

• C'est la somme pondérée par les poids relatifs $\frac{n_k}{n}$ de classe E_k, des matrices de variance-covariance V_k dans chaque groupe E_k.

• C'est la moyenne des variances.

$$W = \sum_{k=1}^{q} \frac{n_k}{n} V_k \quad avec \quad V_k = \frac{1}{n_k}\sum_{i \in E_k}(x_{ij} - g_{kj})(x_{ij'} - g_{kj'})$$

Démonstration

L'analyse discriminante repose sur l'analyse de la variance. La décomposition de la variance sur une partition de l'ensemble des données Y joue un rôle fondamentale.

En tenant compte de g_{kj} le centre de gravité sur chacun des groupes E_k tel que ($k = 1$ à q),

le terme général $v_{jj'}$ de la matrice de variance-covariance peut être réécrit comme suit :

$$\begin{aligned}
v_{jj'} &= \frac{1}{n}\sum_{i=1}^{n}(x_{ij}-g_j)(x_{ij'}-g_{j'}) \\
&= \frac{1}{n}\sum_{k=1}^{q}\sum_{i\in E_k}(x_{ij}-g_j)(x_{ij'}-g_{j'}) \\
&= \frac{1}{n}\sum_{k=1}^{q}\sum_{i\in E_k}(x_{ij}-g_{kj}+g_{kj}-g_j)(x_{ij'}-g_{kj'}+g_{kj'}-g_{j'}) \\
&= \frac{1}{n}\sum_{k=1}^{q}\sum_{i\in E_k}\Big[(x_{ij}-g_{kj})(x_{ij'}-g_{kj'})+(x_{ij}-g_{kj})(g_{kj'}-g_{j'})+(g_{kj}-g_j)(x_{ij'}-g_{kj'}) \\
&\qquad +(g_{kj}-g_j)(g_{kj'}-g_{j'})\Big] \\
&= \frac{1}{n}\sum_{k=1}^{q}\left[\sum_{i\in E_k}(x_{ij}-g_{kj})(x_{ij'}-g_{kj'})+\sum_{i\in E_k}(g_{kj}-g_j)(g_{kj'}-g_{j'})\right]
\end{aligned}$$

puisque $$(g_{kj'}-g_{j'})\sum_{i\in E_k}(x_{ij}-g_{kj}) = (g_{kj}-g_j)\sum_{i\in E_k}(x_{ij'}-g_{kj'}) = 0$$

D'où : $$v_{jj'} = \underbrace{\frac{1}{n}\sum_{k=1}^{q} n_k\left\{\frac{1}{n_k}\sum_{i\in E_k}(x_{ij}-g_{kj})(x_{ij'}-g_{kj'})\right\}}_{w_{jj'}} + \underbrace{\frac{1}{n}\sum_{k=1}^{q} n_k(g_{kj}-g_j)(g_{kj'}-g_{j'})}_{b_{jj'}}$$

Puisqu'il y a n_k individus dans le groupe E_k

$$\text{Donc :} \qquad v_{jj'} = w_{jj'} + b_{jj'}$$

On obtient alors la formule de décomposition de la variance :

$$\underbrace{V}_{variance\ totale} = \underbrace{B}_{variance\ des\ moyennes} + \underbrace{W}_{moyenne\ des\ variance}$$

(Voir Figure A.3 pour la décomposition de la variance en annexe)

Puisqu'on cherche une combinaison linéaire qui permet de séparer au mieux les différents groupes (maximiser la variance inter-classe B) en gardant dans chaque groupe une "étendue" minimal (minimiser la variance intra-classe W), alors soit pour l'individu i la valeur de la combinaison linéaire a des p variables préalablement centrée :

$$a(i) = \sum_{j=1}^{p} a_j(x_{ij} - g_j)$$

Où a est le vecteur des coefficients (p composantes ie variable) permettant de définir le premier axe factoriel ou fonction discriminante.

La variance de a est alors égale :

$$\begin{aligned} var(a) &= \frac{1}{n}\sum_{i=1}^{n} a(i)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{j=1}^{p} a_j(x_{ij} - g_j)\right]^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{j'=1}^{p} a_j a_{j'}(x_{ij} - g_j)(x_{ij'} - g_{j'}) \\ &= \sum_{j=1}^{p}\sum_{j'=1}^{p} a_j a_{j'} cov(x_j, x_{j'}) = a'Va \end{aligned}$$

De la même manière on peut écrire :

$$a'Va = a'Ba + a'Wa \qquad (*)$$

Donc on cherche parmi toutes les combinaisons linéaires des variables, celle qui ont une variance intra-minimale et inter-maximale.

Alors il suffit de chercher le vecteur de coefficients a qui permet de définir un axe, qui en projetant sur cet axe, les q centres de gravité doivent être aussi séparés que possible, tandis que chaque sous-nuage doit se projeter de manière groupé autour de la projection de son centre de gravité(voir figure 1.2).

En d'autre terme, on cherche a tel que le quotient :

$$\frac{a'Ba}{a'Wa} \quad \text{soit maximal} \quad (\text{ où } \quad \frac{a'Wa}{a'Ba} \quad \text{soit minimal})$$

D'après la relation $(*)$: il est équivalent de minimiser $\frac{a'Va}{a'Ba}$ ou de rendre maximale $\frac{a'Ba}{a'Va}$ et donc :

$$a'\mathrm{V}a = a'\mathrm{B}a + a'\mathrm{W}a \Longrightarrow 1 = \frac{a'Ba}{a'Va} + \frac{a'Wa}{a'Va} \Longrightarrow 0 < \frac{a'Ba}{a'Va} < 1$$

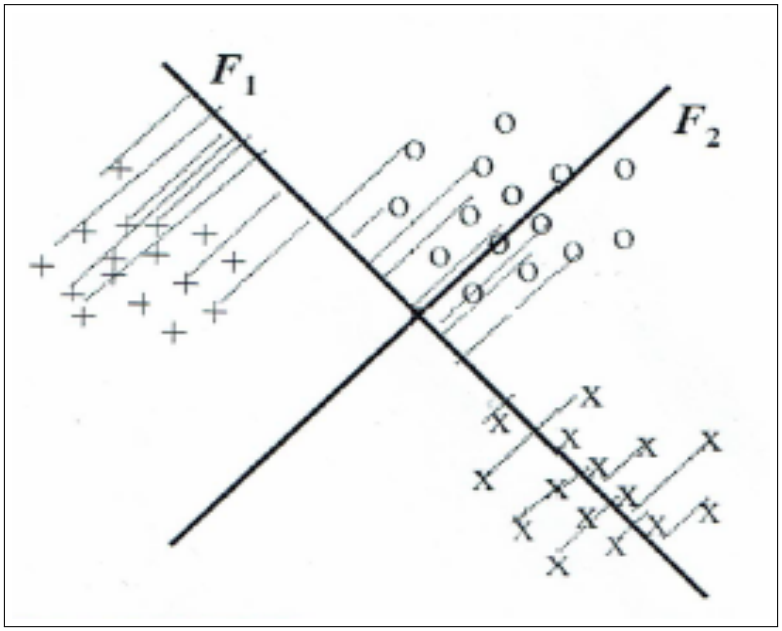


FIGURE 1.2 – Séparation de trois groupes

L'axe 1 de la figure 1.2 possède un bon pouvoir discriminant tandis que l'axe 2 ne permet pas de séparer en projection les trois groupes.
En effet en projection sur l'axe, les centres de gravité doivent être aussi séparés que possible, tandis que chaque sous-nuage doit se projeter de manière groupée autour de la projection de son centre de gravité.

1.3.3 Calcul de la fonction linéaire discriminante

En réservant la notation a pour un vecteur colonne et a' pour le vecteur ligne transposé du précédent, on cherche une représentation géométrique du nuage qui sépare le mieux possible les groupes. Pour cela il faut se donner un critère de séparation optimale.
Le critère s'écrit :

$$\max_{a} \frac{a'Ba}{a'Va} \tag{1.1}$$

correspondant à la plus grande valeur propre $\lambda = \frac{a'Ba}{a'Va}$

Démonstration

La solution du problème (1.1) est invariante par multiplication de a par une constante scalaire quelconque. Il est donc équivalent de résoudre le problème (1.2) :

$$\max_a a'Ba \tag{1.2}$$

Sous la contrainte $a'Va = 1$

En utilisant la méthode du multiplicateur de Lagrange :

$$\frac{\delta}{\delta a}(a'Ba - \lambda a'Va) = 0$$

En dérivant par rapport à a, on obtient :

$$2Ba = 2\lambda Va \tag{1.3}$$

Si V est inversible, on a :

$$V^{-1}Ba = \lambda a$$

Et donc le vecteur a, assurant le maximum de $a'Ba$ est le vecteur propre de $V^{-1}B$ correspondant à la valeur propre λ.
En multipliant les deux membres de (1.3) par le vecteur ligne a', on obtient :

$$a'Ba = \lambda a'Va \qquad \text{d'où} \qquad \lambda = \frac{a'Ba}{a'Va} \qquad \square$$

La valeur propre λ mesure le pouvoir discriminant de l'axe défini par a. en d'autre terme, elle mesure la part de la variance inter-classe dans la variance totale.
C'est-à-dire :

$$\lambda = \frac{a'Ba}{a'Va} = \frac{a'Ba}{a'Ba+a'Wa}$$

Ces formes quadratiques sont toutes définies positives et $V = B + W$, il résulte que $0 \leqslant \lambda \leqslant 1$.
Soit a_1 le vecteur propre correspondant à la plus grande valeur propre λ_1 de $V^{-1}B$, a_1 définis le premier axe factoriel discriminant.
Les valeurs propres étant ordonnées en ordre décroissant, le deuxième axe factoriel discriminant de vecteur directeur a_2 est le vecteur propre correspondant à la deuxième valeur propre λ_2. Il est le meilleur facteur discriminant après a_1 et indépendamment de lui, en d'autre termes a_1 et a_2 sont orthogonaux pour la métrique V^{-1}.
Et ainsi de suite, on prend les valeurs propres successives et les vecteurs propres correspondants

de $V^{-1}B$.

Le nombre de vecteurs propres est égal à $q-1$. C'est la dimension de l'espace affine B engendré par les points moyen des q groupes.

Géométriquement : le premier facteur détermine un axe dans le nuage de points (passant par l'origine) tel que les projections des points sur cet axe aient une variance inter-classe maximale. Le deuxième facteur est non corrélé (perpendiculaire) au premier est de variance inter classe maximale.

• **Si la valeur propre** $\lambda = 1$ ceci correspond à une dispersion intra-classes nulles. Les q sous nuages sont donc chacun dans un hyperplan orthogonal à a (l'axe factoriel discriminant). Il y a évidemment discrimination parfaite si les q centres de gravité se projettent en des points différents.

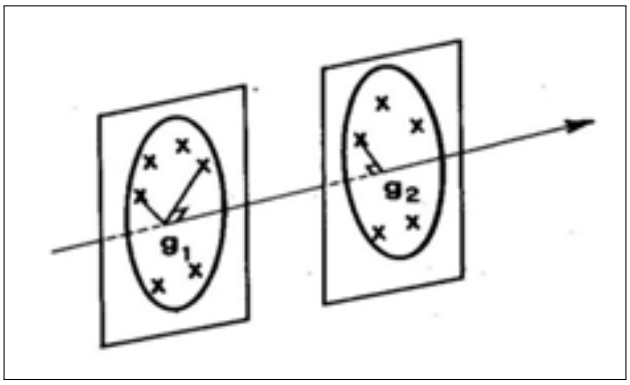


FIGURE 1.3 – Séparation des centres de gravité

• **Et si** $\lambda = 0$ cela correspond au cas où le meilleur axe ne permet pas de séparer les centres de gravité. C'est le cas où ils sont confondus. Les nuages sont donc concentriques et aucune séparation linéaire n'est possible. Il se peut alors qu'il y est de discrimination non linéaire : la distance au centre permet ici de séparer les groupes, mais il s'agit d'une fonction quadratique des variables.

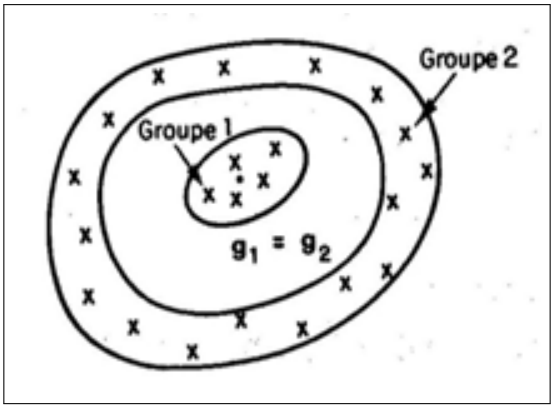


FIGURE 1.4 – Nuages concentriques

• La valeurs propre λ est une valeur pessimiste du pouvoir discriminant d'un axe, car on peut avoir un cas où les classes sont parfaitement séparées, et pourtant on a $\lambda < 1$, comme on le vois dans la figure suivante 1.5.

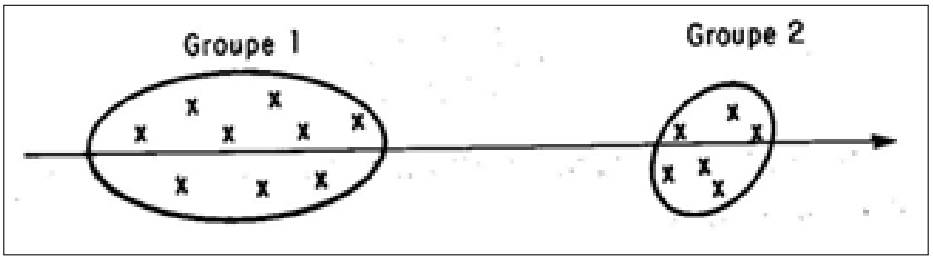


FIGURE 1.5 – Classes séparées

- Un autre critère équivalent est souvent utilisé :

$$\max_a \frac{a'Ba}{a'Wa}$$

En effet :

$$\max_a \frac{a'Ba}{a'Va} = \max_a \frac{a'Ba}{a'(B+W)a} \Longleftrightarrow \min_a (1 + \frac{a'Wa}{a'Ba}) \Longleftrightarrow \max_a \frac{a'Ba}{a'Wa}$$

Exprimé sous cette forme, le critère signifie explicitement qu'on cherche un axe de vecteur directeur a le long du quel le rapport de la variance inter-classe sur la variance intra-classe est maximal, c'est-à-dire, que les groupes apparaissent les plus ramassés possible autour de leurs centres respectifs en même temps que les groupes, représentés par leurs points moyens, apparaissent les plus écartés possible les uns des autres.

La solution est alors l'ensemble des vecteurs propre a de $W^{-1}B$ et les valeurs propres correspondantes sont : $\mu = \frac{a'Ba}{a'Wa}$

Démonstration :

$$V^{-1}Ba = \lambda a \Longrightarrow Ba = \lambda Va = \lambda(B+W)a \Longrightarrow (1-\lambda)Ba = \lambda Wa$$
$$\Longrightarrow W^{-1}Ba = \tfrac{\lambda}{1-\lambda}a$$

Ainsi si a est le vecteur propre de $V^{-1}B$ il l'est aussi de $W^{-1}B$.

Mais par contre les valeurs propres respectives correspondantes diffèrent et sont liées par la relation :

$$\mu = \tfrac{\lambda}{1-\lambda}$$

1.3.4 Détermination des vecteurs propres de $V^{-1}B$

Dans la pratique la matrice $V^{-1}B$ est rarement symétrique. Elle ne peut donc être diagonalisée. On utilise alors une matrice C telle que $CC' = B$.

En effet on sait que

$$B = \tfrac{1}{n}\sum_{k=1}^{q} n_k(g_{kj} - g_j)(g_{kj'} - g_{j'})$$

Cette matrice est le produit d'une matrice C à p lignes et q colonnes par sa transposée, cette matrice a pour terme général :

$$C_{jk} = \tfrac{n_k}{n}(g_{kj} - g_j)$$

Alors on écrit : $V^{-1}CC'a = \lambda a$

Ou si on pose : $a = V^{-1}Cw$

Cette relation s'écrit alors :

$$CC'V^{-1}CW = \lambda VV^{-1}CW \Rightarrow C'V^{-1}CW = \lambda W$$

Les vecteurs propres w sont ceux de la matrice $C'V^{-1}C$ d'ordre $(q.q)$.

Il suffit en pratique d'effectuer la diagonalisation de cette matrice symétrique, puis d'en déduire a par la transformation : $a = V^{-1}Cw$.

1.4 Règle géométrique d'affectation

Ayant trouvé la meilleure représentation de la séparation en q groupes des n individus, on va chercher ici à affecter une nouvelle observation à l'un de ces groupes. Notons x le vecteur des valeurs de p variables quantitatives mesurées sur ce nouvel individu.
La règle consiste à calculer les distances de cette observation à chacun des q centres de gravité et à affecter naturellement cette observation au groupe le plus proche. Pour cela, il faut préciser la métrique à utiliser dans le calcul des distances. La règle la plus utilisée est celle de Mahalanobis-Fisher.

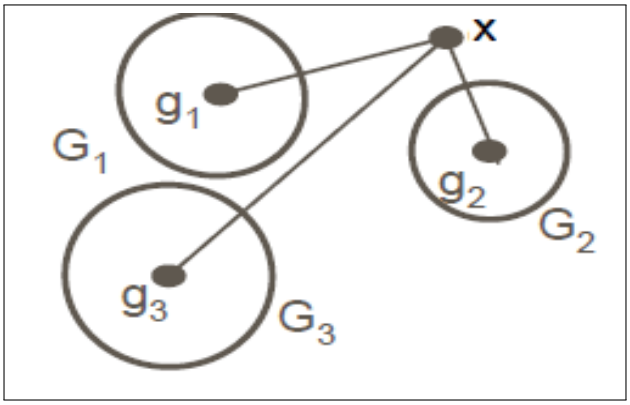


FIGURE 1.6 – Affectation des groupes et ces distances

1.4.1 Métrique de Mahalanobis

Définition de la métrique Mahalanobis

La distance de Mahalanobis est introduite par Prasanta Chandra Mahalanobis en 1936, elle est basée sur la corrélation entre des variables par lesquelles différents modèles peuvent être identifiés et analysés, est une métrique (c-à-d une définition de ce que, on appelle distance entre deux points), qui est mieux adapté que la métrique euclidienne habituelle pour décrire des situations dans lesquelles les distribuions considérées ne sont pas à symétrie sphérique. Bien que sa définition ne l'exige pas, elle est plus particulièrement adaptée aux distributions multi normales.
La métrique Mahalanobis joue un rôle important dans
• Distance d'un point à la moyenne d'une distribution
• Et distance entre les moyennes de deux distributions.

Si on prend par exemple deux points A et B qui sont à égale distance de la moyenne μ (une classe sphérique) alors dans ce cas la distance euclidienne habituelle :

$$d^2(A,\mu) = \sum_i (a_i - \mu)^2$$

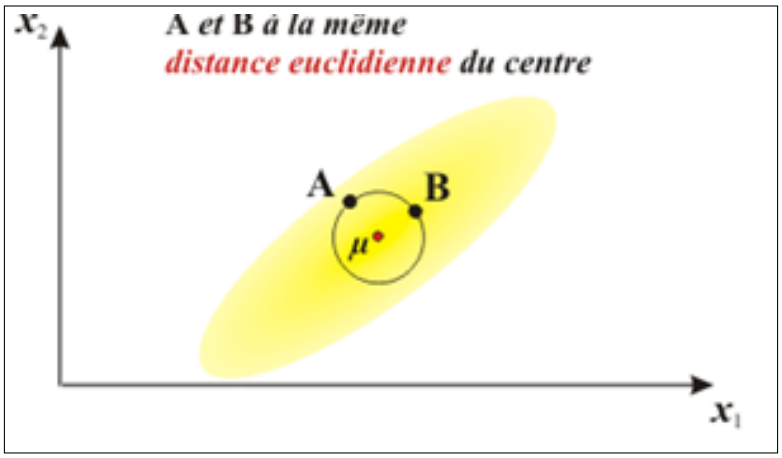


FIGURE 1.7 – Distance euclidienne du centre

Mais si la classe n'est plus sphérique, et en raison de la forme analytique de la distribution normale multi variée, ces deux points conduiraient à la même valeur de la quantité :

$$D^2 = (x-\mu)'\Sigma^{-1}(x-\mu)$$

Où Σ est la matrice de covariance de la distribution.

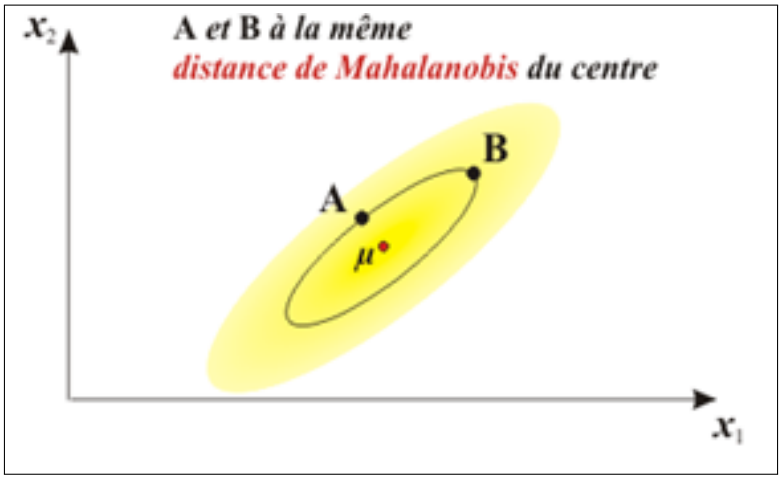


FIGURE 1.8 – Distance de Mahalanobis du centre

D s'appelle la distance de Mahalanobis du point x à la moyenne μ.
Grâce à cette métrique, la distance est interprétée en termes de « vraisemblance d'appartenance ».
Donc on voit l'intérêt de faire intervenir cette métrique dans l'analyse des centres.

Cas de deux groupes

Quand la population est divisée en deux classes, l'analyse discriminante linéaire est ramenée au cas de l'analyse de régression multiple $Y = a * X + \epsilon$ où Y ne prend que deux valeurs.
Lorsque la variable Y ne prend que deux modalités, il n'y a qu'une seule variable discriminante car $q - 1 = 2 - 1 = 1$.
Le facteur discriminant est le facteur principal unique de l'ACP sur le nuage de deux points g_1 et g_2 pondérés par n_1 et n_2, avec la métrique V^{-1} ou W^{-1}.
L'axe discriminant a est la droite reliant les deux centres de gravité g_1 et g_2 :

$$a = (g_1 - g_2)$$

Règle de Mahalanobis-Fisher ou critère métrique

Un nouvel individu x sera affecté au groupe k si la distance qui le sépare du centre de gravité du groupe k est inférieure à toutes les distances qui le séparent des autres centres de gravité. On définit donc la distance qui sépare le nouvel individu x du centre de gravité d'un groupe k par :

$$d^2(x, g_k) = \delta_k(x) = (x - g_k)' M (x - g_k)$$

Avec M la métrique utilisée qui peut être W^{-1} ou V^{-1}
On compare cette distance avec les distances qui séparent l'individu des centres de gravité des autres groupes k' : $\delta_{k'}(x) - \delta_k(x)$, c'est-à-dire :

$$\delta_{k/k'}(x) = \delta_{k'}(x) - \delta_k(x) = (x - g_{k'})' M (x - g_{k'}) - (x - g_k)' M (x - g_k)$$

La décision est alors :

$$\delta_{k'}(x) - \delta_k(x) \geq 0 \Longleftrightarrow \delta_{k/k'}(x) \geq 0 \Longleftrightarrow x \in E_k$$

La différence de distances conduites à la formule de score :

$$\begin{aligned}
&\delta_{k/k'}(x) = (x-g_{k'})'M(x-g_{k'}) - (x-g_k)'M(x-g_k)\\
&\Longrightarrow \delta_{k/k'}(x) = x'Mx - g'_{k'}Mx - x'Mg_{k'} + g'_{k'}Mg_{k'} - [x'Mx - x'Mg_k - g'_kMx + g'_kMg_k]\\
&\Longrightarrow \delta_{k/k'}(x) = x'M(g_k-g_{k'}) + (g'_k-g'_{k'})Mx + g'_{k'}Mg_{k'} + g'_kMg_k\\
&\Longrightarrow \delta_{k/k'}(x) = 2(g_k-g_{k'})'Mx + g'_{k'}Mg_{k'} - g'_kMg_{k'} + g'_kMg_{k'} - g'_kMg_k\\
&\Longrightarrow \delta_{k/k'}(x) = 2(g_k-g_{k'})'Mx + (g'_{k'}-g'_k)Mg_{k'} + g'_kM(g_{k'}-g_k)\\
&\Longrightarrow \delta_{k/k'}(x) = 2(g_k-g_{k'})'Mx + (g_{k'}-g_k)'Mg_{k'} + (g_{k'}-g_k)'Mg_k\\
&\Longrightarrow \delta_{k/k'}(x) = 2(g_k-g_{k'})'Mx + (g_{k'}-g_k)'M(g_{k'}+g_k)\\
&\Longrightarrow \delta_{k/k'}(x) = 2(g_k-g_{k'})'Mx - (g_k-g_{k'})'M(g_k+g_{k'})\\
&\Longrightarrow \delta_{k/k'}(x) = 2\left[(g_k-g_{k'})'M(x-\frac{g_k+g_{k'}}{2})\right]
\end{aligned}$$

On a l'expression du premier degré, encore appelée fonction score :

$$f_{k,k'}(x) = \tfrac{1}{2}\delta_{k/k'}(x) = (g_k-g_{k'})'M(x-\tfrac{g_k+g_{k'}}{2})$$

telle que l'équation de l'hyperplan frontière entre les groupes E_k et $E_{k'}$, optimal au sens du critère métrique :

$$(g_k-g_{k'})'M(x-\tfrac{g_k+g_{k'}}{2}) = 0$$

Si $M = W^{-1}$, on considère l'expression :

$$D^2_W(x;g_k) = (x-g_k)'W^{-1}(x-g_k)$$

- $D^2_W(x;g_k)$ s'appelle la distance de Mahalanobis entre x et g_k.
- $D^2_W(g_1;g_2)$ s'appelle le D^2 de Mahalanobis.

Pour la discrimination de q groupes on dispose de q distances $\delta_1, \delta_2, ..., \delta_q$. Pour affecter x à un des groupes, les $\delta_k(x)$ sont comparés entre eux et on affecte x au groupe correspondant à la plus petite distance δ_k.

Pour les $\delta_{k/k'}(x)$ on a $\frac{q(q-1)}{2}$ expressions distinctes utiles. En effet $\delta_{k/k'}(x) = -\delta_{k'/k}(x)$

1.4.2 Cas de deux groupes

Plutôt que considérer δ_1 et δ_2, on considère une seule formule :

$$Score = f(x) = \tfrac{1}{2}\delta_{1/2}(x) = (g_1-g_2)'M(x-\tfrac{g_1+g_2}{2}) = \alpha'x + \beta$$

Où $\alpha' = (g_1 - g_2)'M$ et $\beta = -(g_1 - g_2)'M\frac{g_1+g_2}{2}$

C'est le signe de cette expression qui nous intéresse. Le nombre 0 joue le rôle de seuil de décision.

1. Si $f(x) > 0$, on affecte x au groupe 1

2. Si $f(x) < 0$, on affecte x au groupe 2

3. Si $f(x) = 0$, il n'y a pas d'affectation

$f(x) = 0$ est l'équation de l'hyperplan médiateur du segment $[g_1; g_2]$. Il sépare le nuage en deux demi-espaces qui sont les régions de décisions.

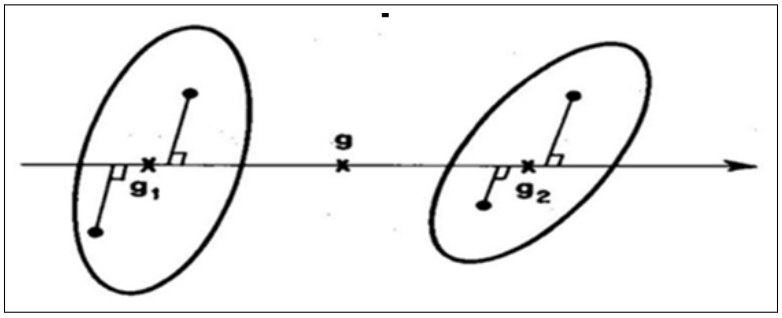


FIGURE 1.9 – Mahalanbis-Fisher cas de deux groupes

(Voir Figure A.2 pour le cas de trois groupes)

Le facteur discriminant f' vaut donc :

$$f' = (g_1 - g_2)V^{-1}$$

On peut retrouver l'unique valeur propre de $V^{-1}B$ en remarquant que pour deux groupes

$$B = \frac{1}{n}\sum_{k=1}^{2} n_k(g_{kj} - g_j)(g_{kj'} - g_{j'})$$
$$\Longrightarrow B = \frac{1}{n}(n_1(g_1 - g)(g_1 - g)' + n_2(g_2 - g)(g_2 - g)') \quad \text{avec} \quad g = \frac{1}{n}(n_1g_1 + n_2g_2)$$

En remplaçant g_j par sa valeur et en tenant compte de fait que $n_1 + n_2 = n$, on trouve :

$$\Longrightarrow B = \frac{n_1n_2}{n}(g_1 - g_2)(g_1 - g_2)'$$

La matrice des covariances entre les classes B d'ordre $(p.p)$ et de rang 1 peut être considérée comme le produit d'une matrice colonne C par sa transposée $B = CC'$, avec :

$$c_j = \frac{\sqrt{n_1 n_2}}{n}(g_1 - g_2)$$

Ainsi la relation $V^{-1}Ba = \lambda a$ devient : $V^{-1}CC'a = \lambda a \Longrightarrow (C'V^{-1}C)C'a = \lambda C'a$
Et finalement : $\lambda = C'V^{-1}C$. $C'V^{-1}C$ est un scalaire, égale par conséquent à.
En effet :

$$\lambda = \frac{n_1 n_2}{n}(g_1 - g_2)V^{-1}(g_1 - g_2)'$$

Puisque B est de rang 1, la valeur propre λ est unique (λest la distance Mahalanobis entre les deux classes) et son vecteur propre $a = V^{-1}C$ est l'unique fonction discriminante.
Considérons maintenant le problème comme s'il s'agissait de régression multiple. Le modèle est $w = X\beta$ où X est la matrice ayant les p variables explicatives centrées en colonnes. Le vecteur w à n composantes est définis par :

$$w_i = \begin{cases} \sqrt{\frac{n_2}{n_1}} & \text{Si l'individus } i \text{ à la classe 1} \\ -\sqrt{\frac{n_2}{n_1}} & \text{Si l'individus } i \text{ à la classe 2} \end{cases}$$

Alors la régression multiple expliquant w par les colonnes de X conduit au vecteur de coefficients noté ici b estimateur de β :

$$b = (X'X)^{-1}X'w \quad \text{avec} \quad \frac{1}{n}X'X = V$$

On vérifie que : $\frac{1}{n}X'w = Cd'$ où $b = V^{-1}C$
Le vecteur des coefficients de régression b coïncide par conséquent avec le vecteur des composantes de la fonction discriminante a calculé précédemment.

1.5 Analyse discriminante sur une variable qualitative

1.5.1 Méthode Disqual

L'FAD dans sa version usuelle n'utilise que des variables quantitatives. SAPORTA 1975 a proposé de généraliser cette méthode dans le cas où les variables explicatives sont p variables qualitatives $x_1, ..., x_p$ respectivement à $m_1, ..., m_p$ modalités.

La méthode Disqual (discrimination sur variables qualitatives) construit une fonction score en deux étapes. Elle est issue d'une analyse des correspondances multiples, menée sur le tableau disjonctif complet des modalités des variables, suivie d'une analyse discriminante de Fisher sur les axes factoriels les plus discriminants issus de l'analyse des correspondances multiples.

L'analyse des correspondances multiples répond à deux objectifs : d'une part, elle met en évidence, sur un nuage de points, les tendances dominantes et les proximités entre les différents individus statistiques et d'autre part, elle permet de résumer au maximum l'information en substituant au nuage de points initial un nuage de dimension plus réduite caractérisé selon ses axes factoriels. Ces axes factoriels se définissent comme une combinaison linéaire des variables indicatrices des modalités, parmi lesquels seuls les plus discriminants seront conservés.
Une analyse discriminante linéaire de Fisher est alors menée sur les axes factoriels les plus discriminants. Donc les p variables qualitatives sont remplacées par les q coordonnées sur les axes factoriels, et une AD est effectuée sur ces q variables numériques $z_1, ..., z_q$.
Une fonction discriminante d est une combinaison linéaire des z_j qui sont elles-mêmes des combinaisons linéaires des indicatrices.
On exprime ensuite directement d comme une combinaison linéaire des indicatrices ce qui revient à attribuer à chaque catégories de chaque variables une valeur numérique ou score.
d est alors simplement égale à l'addition des scores obtenues dans les catégories des p variables.
Ceci revient donc à transformer chaque variable qualitative en une variable discrète à m valeurs (associées à chaque modalité).

1.6 Extension et limite de la règle géométrique

1.6.1 Extension

La règle de fisher se généralisé pour k et p quelconque un point e dans $\mathbb{R}^p$ est classé dans le groupe l si g_l est le centre de gravité le plus proche au sens de Mahalanobis $(x-g_l)'W^{-1}(x-g_l) = (x-g_k)'W^{-1}(x-g_k)$.

1.6.2 Limite

La métrique W^{-1} est estimée sur l'ensemble des données, ce loi engendre plusieurs problèmes :
• pas adapté au cas où les classes ont de dispersions différentes (W_k dépend de k).
• De façon implicite, chaque classe est supposé avoir le même poids : l'appartenance à une classe n'est pas forcément uniforme (pensez au cas où n_k dépend de k).

1.6.3 Qualité de règle de classification

• Une règle de classification est bonne si les individus sont rarement mal-classés

On cherche donc à minimiser la probabilité de mauvais classement.

• La règle géométrique est optimale dans le contexte suivant :

- Les observations d'une classe q suivant une loi normale (multivariée) de matrice de variance-covarince W.

- W ne dépend pas de q : elle est identique pour toutes les classes.

- Les classes sont équidistribuées : les probabilité apriori d'appartenance à chaque classe sont égales.

Chapitre 2

Méthode probabiliste

2.1 Introduction

Dans le chapitre précédent on disposait d'un échantillon de taille n sur lequel étaient observées à la fois les variables explicatives X_j, $j = 1, ..., p$ et la variable d'intéret Y. En général, cet échantillon est appelé échantillon d'apprentissage.

L'idée de base est de classer une observation(un nouvel individu ou de plusieurs sur lequel on a observé les X_j mais pas Y) dans le groupe pour lequel la probabilité conditionnelle d'appartenir à ce groupe étant données les valeurs observées est maximale. En pratique on ne peut calculer ces probabilités que si les observations proviennent d'une loi multi normale. Si tel n'est pas le cas on devra au préalable transformer les données pour s'en rapprocher le plus possible. (La pratique a toutefois prouvée que l'analyse des données était très robuste face à l'hypothèse de multi normalité). On parle aussi de problème d'affectation. Pour cela, on va définir et étudier dans ce chapitre des règles de décision (ou d'affectation) et donner ensuite les moyennes de les évaluer.

2.2 Règle décision bayésienne

2.2.1 Introduction

La classification bayésienne est une autre approche probabiliste qui suppose connues les probabilités a priori et les distributions des probabilités d'appartenance à chaque classe.
Dans ce cas c'est une méthode optimale c'est-à-dire celle qui maximiser la probabilité $\mathbb{P}(G_k \backslash y)$: probabilité conditionnelle a posteriori. En pratique, ces probabilités sont estimées à partir de

données d'apprentissage. On présente brièvement cette méthode très utilisée en classification, comme méthode de classement de l'analyse factorielle discriminante.

2.2.2 Explication de la règle

Au moment de l'apprentissage, on sait que l'individu i appartient au groupe G_k (appartenance codée par la valeur : $Y_i = k$) et on calcule une estimation de la probabilité $\mathbb{P}(x_i \backslash G_k)$, c'est-à-dire la probabilité de x_i sachant que G_k est réalisé.
Au moment de l'affectation d'un individu nouveau noté x, on peut calculer les différents $\mathbb{P}(x \backslash G_k)$ pour $k = 1, 2, ..., q$. Il paraît raisonnable d'affecter x à la classe G_k pour laquelle $\mathbb{P}(x \backslash G_k)$ est maximale.
Cependant, ce ne sont pas les probabilités $\mathbb{P}(x \backslash G_k)$ qu'il faudrait connaître mais les probabilités $\mathbb{P}(G_k \backslash x)$, c'est-à-dire la probabilité du groupe G_k sachant que x est réalisé.
Le théorème de bays permet de procéder à cette inversion des probabilités.
Il exprime $\mathbb{P}(G_k \backslash x)$ en fonction de $\mathbb{P}(x \backslash G_k)$, $\mathbb{P}(G_k)$ et $\mathbb{P}(x)$:

$$\mathbb{P}(G_k \backslash x) = \frac{\mathbb{P}(x \backslash G_k)\mathbb{P}(G_k)}{\mathbb{P}(x)}$$

est la probabilité a posteriori du groupe k.$\mathbb{P}(x)$ en fonction de $\mathbb{P}(x \backslash G_k)$ et de $\mathbb{P}(G_k)$; d'où la formulation classique du théorème de Bayes :

$$\mathbb{P}(G_k \backslash x) = \frac{\mathbb{P}(x \backslash G_k)\mathbb{P}(G_k)}{\sum_{k=1}^{q} \mathbb{P}(x \backslash G_k)(G_k)}$$

$\mathbb{P}(x \backslash G_k)$ est la probabilité d'observer x au sein de la classe G_k.
$\mathbb{P}(x \backslash G_k) = \mathbb{P}(x = x_q^k \backslash G_k)$ dans le cas discret.
$\mathbb{P}(x \backslash G_k) = f(x \backslash G_k)$ dans le cas continu.
dans tous les deux cas, on utilise $f_k(x)$.
La règle de décision s'écrit finalement :

$$\max_{k=1,...,q} \mathbb{P}(G_k) f_k(x)$$

La règle bayésienne d'affectation consiste alors à affecter l'observation x au groupe de la probabilité à postériori maximale.

2.2.3 Détermination des probabilités à priori

Les probabilités a priori $\mathbb{P}(G_k)$ peuvent effectivement être connues a priori : proportions de divers groupes dans une population, de diverses maladies...etc, sinon elles sont estimées sur l'échantillon d'apprentissage :

$$\hat{\mathbb{P}}(G_k) = \frac{n_k}{n} \text{ (si tous les individus ont le même poids)}$$

Les différentes méthodes d'estimation des densités conditionnelles $f_k(x)$ conduisent aux méthodes classiques de discrimination bayésienne.

Cas particuliers

• Dans le cas où les probabilités a priori sont égales, c'est par exemple le cas du choix de probabilités non informatives, la règle de décision bayésienne revient alors à maximiser $f_k(x)$ qui est vraisemblance, au sein de G_k, de l'observation x. La règle consiste alors à choisir la classe pour laquelle cette vraisemblance est maximum.

• Dans le cas où $q = 2$, on affecte x à G_1 si :

$$\frac{f_1(x)}{f_2(x)} > \frac{\mathbb{P}(G_2)}{\mathbb{P}(G_1)}$$

faisant ainsi apparaitre un rapport de vraisemblance. D'autre part, l'introduction de coûts de mauvais classement différent selon les classes amène à modifier la valeur limite $\frac{\mathbb{P}(G_2)}{\mathbb{P}(G_1)}$.

2.3 Règle bayésienne avec modèle gaussien

2.3.1 Introduction

On suppose dans cette section que, conditionnellement à G_k, $x = (x_1, ..., x_P)$ est l'observation d'un vecteur aléatoire gaussien $\mathcal{N}(\mu_k, \Sigma_k)$; μ_k est un vecteur de $\mathbb{R}^p$ et Σ_k une matrice $(p.p)$ symétrique et définie-positive. La densité de la loi, au sein de la classe k, s'écrit donc :

$$f_k(x) = \frac{1}{\sqrt{(2\pi det(\Sigma_k))}} \exp\left[-\frac{1}{2}(y-\mu_k)'\Sigma_k^{-1}(y-\mu_k)\right].$$

L'affectation de x à une classe se fait en maximisant $\mathbb{P}(G_k)$. $f_k(x)$ par rapport à k soit encore la quantité :

$$\ln\left(\mathbb{P}(G_k) - \frac{1}{2}\ln(det(\Sigma_k)\right) - \frac{1}{2}(y-\mu_k)'\Sigma_k^{-1}(y-\mu_k).$$

Dans les applications, on distingue en fait entre deux modèles d'analyse discriminante selon que l'on suppose que les Σ_k sont différentes d'un groupe à un autre (modèle hétéroscédastique) ou que ces matrices sont identiques (modèle homoscédastique).

2.3.2 Hétéroscédasticité

Dans le cas générale, il n'y a pas d'hypothèse supplémentaire sur la loi de x et donc les matrices Σ_k sont fonction de k. Le critère d'affectation est alors quadratique en x. Les probabilités $\mathbb{P}(G_k)$ sont supposées connues mais il est nécessaire d'estimer les moyennes μ_k ainsi que les covariances Σ_k en maximisant, compte tenu de l'hypothèse de normalité, la vraisemblance. Ceci conduit à estimer la moyenne.

$$\hat{\mu}_k = \bar{x}_k$$

par la moyenne empirique de x dans la classe k pour l'échantillon d'apprentissage et Σ_k par la matrice de covariance empirique S^*_{Rk} :

$$S^*_{Rk} = \frac{1}{n_k - 1} \sum_{i \in \Omega} (x_i - \bar{x}_k)(x_i - \bar{x}_k)'$$

pour ce même échantillon.

2.3.3 Homoscédasticité

On suppose dans ce cas que les lois de chaque classe partagent la même structure de covariance $\Sigma_k = \Sigma$. Supprimant les termes indépendant de k, le critère à maximiser devient :

$$\ln\left(\mathbb{P}(G_k)\right) - \frac{1}{2}\mu_k' \sum_k^{-1} \mu + \mu_k' \sum_k^{-1} x$$

qui est cette fois linéaire en x. Les moyennes μ_k sont estimées comme précédemment tandis que Σ est estimé par la matrice de covariance intra empirique :

$$S^*_R = \frac{1}{n - q} \sum_{k=1}^{q} \sum_{i \in \Omega_k} (x_i - \bar{x}_k)(x_i - \bar{x}_k)'$$

Si, de plus les probabilités $\mathbb{P}(G_k)$ sont égales, après estimation le critère s'écrit :

$$\bar{x}_k' S_R^{*-1} x - \frac{1}{2} \bar{x}_k' S_R^{*-1} \bar{x}_k'$$

2.3.4 Commentaire

Les hypothèses : normalité, éventuellement l'homoscédasiticité, doivent êtres vérifiées par la connaissance a priori du phénomène ou par une étude préalable de l'échantillon d'apprentissage. L'hypothèse d'homoscédasticité, lorsqu'elle est vérifiée permet de réduire très sensiblement le nombre de paramètres à estimer d'aboutir à des estimateurs plus fiable car de variance moins élevée. dans le cas contraire, l'échantillon d'apprentissage doit être de taille importante.

2.4 Régression logistique

Lorsqu'il ya deux groupes, sous l'hypothèse de normalité et d'égalité des matrices de variances intra-classes, on a vu que la probabilité à posteriori était une fonction logistique de score, lui-même fonction linéaire des variables explicatives.

$$On\ a\ donc : \qquad \ln\left(\frac{f_1(x)}{f_2(x)}\right) = B_0 + B'x \quad tel\ que \quad B = (B_1, ..., B_p)' \qquad (2.1)$$

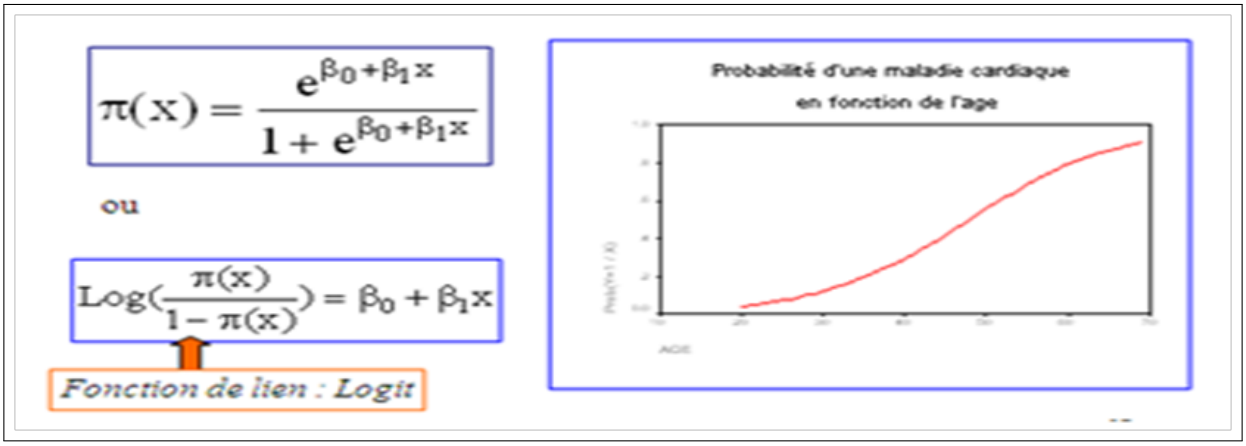


FIGURE 2.1 – Fonction logistique

Ce modèle consiste à partir de la relation (2.1) à estimer $(P+1)$ paramètres selon la méthode de maximum de vraisemblance par rapport à la discrimination linéaire usuelle, le modèle logistique impose (implique) moins de paramètres et couvre une gamme étendu de loi de probabilité (les variables explicatives peuvent être binaire).
On va estimer le B par le maximum de vraisemblance, la règle bayésienne peut être appliqué pour les classements comme :

$$\ln \frac{\mathbb{P}(G_1\backslash x)}{\mathbb{P}(G_2\backslash x)} = B_0 + \ln \frac{p_1}{p_2} + B'x$$

On a affecté x au groupe 1 si : $\qquad B_0 + \ln(\frac{p_1}{p_2}) + B'x > 0$

2.5 Règle bayésienne avec estimation non paramétrique

2.5.1 Introduction

En statistique, on parle estimation non paramétrique ou fonctionnelle lorsque le nombre de paramètres à estimer est infini. L'objet statistique à estimer est lors une fonction par exemple

de régression $y = f(x)$ ou encore une densité de probabilité. Dans ce cas, au lieu de supposer a affaire à une densité de type connu (normale) dont on estime les paramètres, on cherche une estimation $\hat{f}$ de la fonction de densité f. Pour tout x de $\mathbb{R}$, $f(x)$ est donc estimée par $\hat{f}(x)$.
Cette approche très souple a l'avantage de ne pas nécessiter d'hypothèse particulière sur la loi (seulement la régularité de f pour de bonnes propriétés de convergence), en revanche elle n'est applicable qu'avec des échantillons de grande taille d'autant plus que le nombre de dimensions p est grand (curse of dimensionality).
Dans le carde de l'analyse discriminante, ces méthodes permettent d'estimer directement les densités $f_k(y)$. On considère ici deux approches : la méthode du noyau et celle des k plus proches voisins.

2.5.2 Méthode de noyau

Estimation de la densité

Soit $x_1, ..., x_n$ n observation équipondérées d'une v.a.r. continue x de densité f inconnue. Soit $K(x)$ (le noyau) une densité de probabilité unidimensionnelle (sans rapport avec f) et h un réel strictement positif. On appelle estimation de f par la méthode du noyau la fonction

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x-x_i}{h})$$

Il est immédiat de vérifier que :

$$\forall x \in \mathbb{R}, \hat{f}(x) \geqslant 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} \hat{f}(x)dx = 1$$

h est appelé largeur de fenêtre ou paramètre de lissage ; plus h est grand, l'estimation $\hat{f}$ est régulière. L e noyau K est choisi centré en 0, uni modal et symétrique. Les cas les plus usuels sont la densité gaussienne, celle uniforme sur $[-1, 1]$ ou triangulaire : $K(x) = [1 - |x|]\mathbb{1}_{[-1,1]}(x)$. La forme de noyau n'est pas très déterminante sur la qualité de l'estimation contrairement à la valeur de h.

Application de la méthode de noyau à l'analyse discriminante

Cette méthode est utilisée pour calculer une estimation non paramétrique de chaque densité $f_j(x)$, le noyau k^* donc être choisi multidimensionnelle.

$$f_j(x) = \frac{1}{n_j h^p} \sum_{i \in E_l} k^*(\frac{x-x_p}{h}) \quad \textit{tel que} \quad k^* = \prod_{i=1}^{n_j} k(x_j)$$

Cette méthode consiste à diviser l'espace multidimensionnel de l'échantillon d'apprentissage en cellules de volumes comparable V_r puis de compter à l'intérieure de chaque classes $j \quad tq \quad (j < q)$ les n_{rl} observations obtenues dans chaque cellule r.
La fréquence $\frac{n_{rl}}{n_j}$ est une estimation de la probabilité qu'une observation de la catégorie j appartient à la cellule V_r, la règle de Bays permet alors d'effectuer une observation supplémentaire x à une catégorie j après avoir déterminer la cellule V_r qui la contient.
Comme cas particulier de cette méthode, on trouve **la méthode de boules :**en trace autour de x le nouvel individu une boule de rayon R donnée dans $\mathbb{R}^p$ et on compte le nombre d'observation n_{rj} de groupe j dans cette boule, on estime alors directement $\mathbb{P}(G_j \backslash x)$ par $\frac{n_{rj}}{n_j}$.

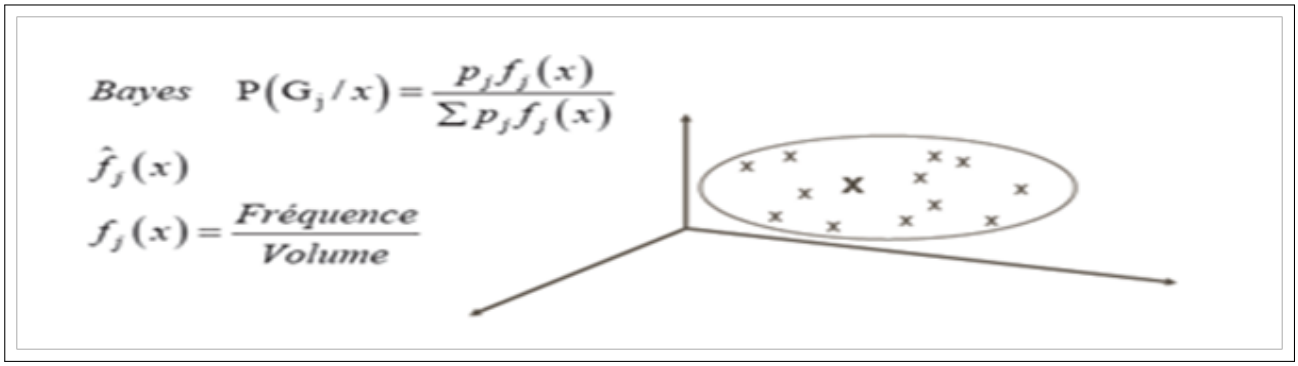


FIGURE 2.2 – Méthode de boules

Remarque : la boule peut être vide si R est trop petit.

2.5.3 K plus proches voisins

On cherche les k points plus proche de nouvel individu x au sens d'une métrique à préciser, et on classe x dans le groupe (classe) plus représenté : la probabilité à posteriori s'obtient comme par la discrimination par boule mais n'a pas un grand sens si k est faible.

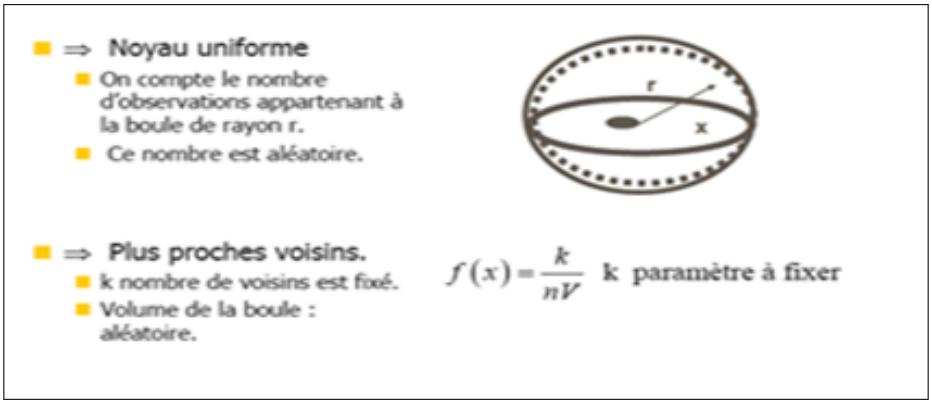


FIGURE 2.3 – Méthode de noyau

Cette méthode consiste à enchainer les étapes suivantes :

i. choix d'un entier $k : 1 \geq k \geq n$

ii. calculer les distances $d_M(x, x_i)$, $i = 1, ..., n$ où M est la métrique de Mahalanobis c'est-à-dire la matrice inverse de la matrice de covariance (ou de variance-intra)

iii. retenir les k observations $(x_1, ..., x_p)$ pour lesquelles ces distances sont les plus petites

iv. compter les nombres de fois $(k_1, ..., k_m)$ que ces k observations apparaissent dans des classes

v. estimer les densités par

$$\hat{f}_j(x) = \frac{k_j}{kV_k(x)}$$

où $V_k(x)$ est le volume de l'ellipsoïde $\{z; (z-y)'M(z-y) = d_M(x, x_{(k)})\}$.

Remarque : Pour $k = 1$, x est affecté à la classe du plus proche élément.

2.6 Qualité des règles de classement -évaluation des règles de la densité-

On décrit les trois méthodes d'évaluation les plus courantes d'une règle de décision en analyse discriminante. Cette étape est indispensable afin de s'assurer de la fiabilité des résultats.

2.6.1 Méthode de la Resubstitution

Cette méthode consiste à appliquer la règle de décision choisie sur l'échantillon d'apprentissage, on calcule ensuite le taux de mal classées. Il s'appelle le taux apparent d'erreur est fourni en standard par les logiciels. Il est en effet biaisé car est estimé sur les données qui

ont servi à définir la règle de décision. Il est d'autant plus faible que le modèle est complexe (sur-paramétrisation) et que la taille de l'échantillon est faible. Il est peu recommandé.

2.6.2 Méthode de l'échantillon test

Elle consiste à partager l'échantillon en deux parties : une partie de l'ordre de 80 sert d'échantillon d'apprentissage de la règle de décision, l'autre partie sert à tester. Cette estimation est plus faible (non biaisée) mais nécessite un échantillon plus important.
En effectuant plusieurs tirages aléatoires d'échantillons d'apprentissage, on améliore encore l'estimation du taux d'erreur en calculant la moyenne des valeurs obtenus à chaque tirage.

2.6.3 Méthode de la validation croisée

Pour tout $i = 1, ..., n$ on considère les n échantillons d'apprentissage constitués en éliminant la i ème observation. La règle de décision qui en découle est utilisée pour affecter cette i ème observation. Le taux dé erreur est estimé en divisant le nombre de mal classés par n.
Les moyens de calcul actuels permettent d'estimer les taux d'erreur en des temps raisonnables, ces techniques itératives doivent être systématiquement mises en œuvre en pratique.

2.7 Réduction de nombre des variables

L'analyse discriminante est d'autant plus couteuse que l'on considère un plus grand nombre de variables, puisque l'opérateur $D^{-1}E$ qu'il faut diagonaliser est de dimension.
Il est donc utile de limiter le nombre de variables.
On peut utiliser deux procédés : ou bien on réalise dans une première étape une analyse en composante principales, ou bien on utilise la méthode dite « pas à pas ».

2.7.1 Passage par l'analyse en composantes principales -l'ACP-

Si deux variables observées sont fortement corrélées, elles feront double emploi du point de vue de l'analyse discriminante, et on ne perdra aucune information en ne considérant que l'une des deux.
D'un point de vue plus général, si les variables observées sont reliées par des corrélations fortes, on pourra les remplacer par un petit nombre de combinaisons linéaires, celles qui sont associées

aux premiers axes d'une analyse en composantes principales, tout en conservant la plus grande partie de l'information qu'elles apportent.

2.7.2 Démarche du pas à pas

Si on observe k individus, le rapport $\frac{inertie\ entre\ les\ classes}{inertie\ dans\ les\ classes}$ est celui que l'on obtient en considérant tous les vecteurs propres de $V^{-1}B$, soit :

$$\frac{1}{k}\sum_{\alpha}^{k}\lambda_\alpha = \frac{1}{k}trace(D^{-1}E)$$

Ce rapport sera donc d'autant plus élevé que $trace(V^{-1}B)$ sera plus grand.
Ceci nous fournit un critère de sélection des variables : nous prendrons d'abord celle qui, si on la considère seule, rend maximum $traceV^{-1}B$: c'est-à-dire le rapport $\frac{inertie\ entre\ les\ classes}{inertie\ dans\ les\ classes}$.
Puis on cherchera, parmi les autres variables, celles qui lorsqu'on l'adjoint à la première rend maximal $trace(V^{-1}B)$, etc, on n'est pas certain bien sûr que cette démarche au pas à pas permettre d'obtenir à chaque stade celle des variables qui ajoute la meilleure discrimination sur les premiers axes de l'analyse discriminantes, au sens de $trace(V^{-1}B)$, pour un effectif donné de cet ensemble.
Cependant dans les applications pratiques, la démarche au pas à pas permet d'arriver rapidement à un résultat acceptable pour un faible coût.
D'autre critères peuvent être utilisés au lieu de $trace(V^{-1}B)$ pour opérer la sélection des variables retenues dans la démarche « pas à pas » : on peut par exemple chercher à maximiser le « pourcentage de biens classés », ou encore à maximiser la première valeur propre de $V^{-1}B$.

Chapitre 3

Application

3.1 Objectif de l'étude

Les données utilisées dans cet exemple sont tirées de SAPORTA(1990) concernant 91 victimes d'infarctus du myocarde (crise cardiaque) où 45 décèderont et 46 survivront, sur lequel ont été mesurées à leurs admission 7 variables quantitatives qui sont :

1. Fréquence cardiaque notée : FRCAR : c'est le nombre de battement cardiaque(pulsation) par unité de temps.
2. Index cardiaque notée : INCAR : c'est la quantité de sang expulsé par chacun des ventricules du cœur, par minute et par mètre carré de surface corporelle.
3. Index systolique notée : INSYS : c'est la quantité de sang injectée par le ventricule gauche de cœur rapportée à la surface corporelle.
4. Pression diastolique notée : PRDIA : c'est la tension artérielle mesurée lors de la phase de relâchement du cœur.
5. Pression artérielle pulmonaire notée : PAPUL : elle mesure la force exercée par le sang sur les parois des artère. Deux chiffres sont annoncés : le premier est le plus grand(autour de 12) correspond à la pression systolique, le deuxième correspond à la pression diastolique (autour de 8). Elle se mesure de millimètre de mercure.
6. Pression ventriculaire notée : PVENT.
7. Résistance pulmonaire notée : REPUL.

3.2 Analyse des résultats

L'analyse des liaisons entre le tableau quantitatif X des descripteurs et la variable qualitative Y à q modalités, codant la partition a priori en q groupes de l'ensemble des individus, peut être menée selon deux points de vue : le premier à orientation descriptive est centré sur la décomposition de la variance en s'appuyant sur des notions géométriques, le second à orientation décisionnelle se focalise sur le risque d'erreur en faisant intervenir une modélisation probabiliste. L'analyse discriminante (AD) se déroule en trois grandes étapes :

1. Statistiques descriptives
2. Approche géométrique
3. Approche probabiliste

Par tirage aléatoire, Nous choisissons un échantillon pour appliquer l'AD, elle représente 90% de l'échantillon total (46 survivront et 45 décèderont) sur lequel, on estimé la fonction linéaire discriminante. Le reste 10% des observations est concéderé comme échantillon test (4 survivront et 6 décèderont) réserver pour la validation.
Tableau 3.1 représente l'échantillon d'apprentissage

Individus	FRCAR	INCAR	INSYS	PRDIA	PAPUL	PVENT	REPUL	PRONO
57	110	0.96	8.8	15.0	19.0	16.0	1583.00	SURVIE
3	120	1.40	11.7	23.0	29.0	8.0	1657	DECES
72	120	1.18	9.9	25.0	36.0	8.0	2441.00	DECES
88	51	1.34	26.3	11.0	17.0	6.0	1015.00	DECES
19	125	3.37	26.9	18.0	28.0	6.0	665.00	SURVIE
91	132	1.31	9.9	23.0	28.0	12.0	1710.00	DECES
46	92	3.06	33.3	10.0	15.0	6.0	392.00	SURVIE
87	94	1.21	12.9	17.0	22.0	3.0	1455.00	DECES
74	101	2.55	25.2	23.2	30.5	9.0	957.00	SURVIE
96	112	1.54	13.8	25.0	31.0	8.0	1610.00	DECES
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
31	118	2.31	19.6	22.0	27.0	10.0	935.00	SURVIE
47	94	1.31	13.9	26.0	40.0	15.0	2443.00	DECES
39	90	0.95	10.6	20.0	24.0	6.0	2021.00	DECES
14	61	2.84	47.3	11.0	17.0	12.0	479.00	SURVIE
2	90	1.68	18.7	24.0	31.0	14.0	1476.00	DECES
60	80	2.65	33.1	13.0	19.0	9.0	574.00	SURVIE
69	80	2.85	35.6	25.0	32.0	7.0	898.00	SURVIE
100	83	2.37	27.2	15.0	22.0	10.0	743.00	SURVIE
48	79	1.29	16.3	24.0	31.0	10.0	1922.00	DECES
37	90	2.08	23.1	20.0	28.0	6.0	1077.00	SURVIE

TABLE 3.1 – Échantillon d'apprentissage

Tableau 3.2 représente l'échantillon test

individu	FRCAR	INCAR	INSYS	PRDIA	PAPUL	PVENT	REPUL	PRONO
16	92	2.47	26.8	12	19.0	11.0	615	SURVIE
24	118	1.03	8.7	19	27.0	10.0	2097	DECES
25	95	1.89	19.9	25	27.0	20.0	1143	DECES
34	84	2.15	25.6	27	37	10.0	1377	SURVIE
35	103	0.91	8.8	30	33.5	10.0	2945	DECES
44	82	2.02	24.6	16	22.0	14.0	871	SURVIE
56	92	1.97	21.4	18	27.0	3.0	1096	DECES
61	102	1.60	15.7	24	31.0	16.0	1550	DECES
66	108	2.96	27.4	24	35.0	6.5	946	SURVIE
73	115	1.83	15.9	25	30.0	8.0	1311	DECES

TABLE 3.2 – Échantillon de test

3.3 Statistiques descriptives

Comme dans toutes analyses de données, on commence toujours par des statistiques uni variées pour chacun des groupes étudiés sur les variables de l'analyse. L'objectif est de déterminer le rôle d'une variable sur la variable explicative. Le nuage de points des individus pour les deux variables INCAR et PAPUL.(voir l'annexe pour les autres variables)

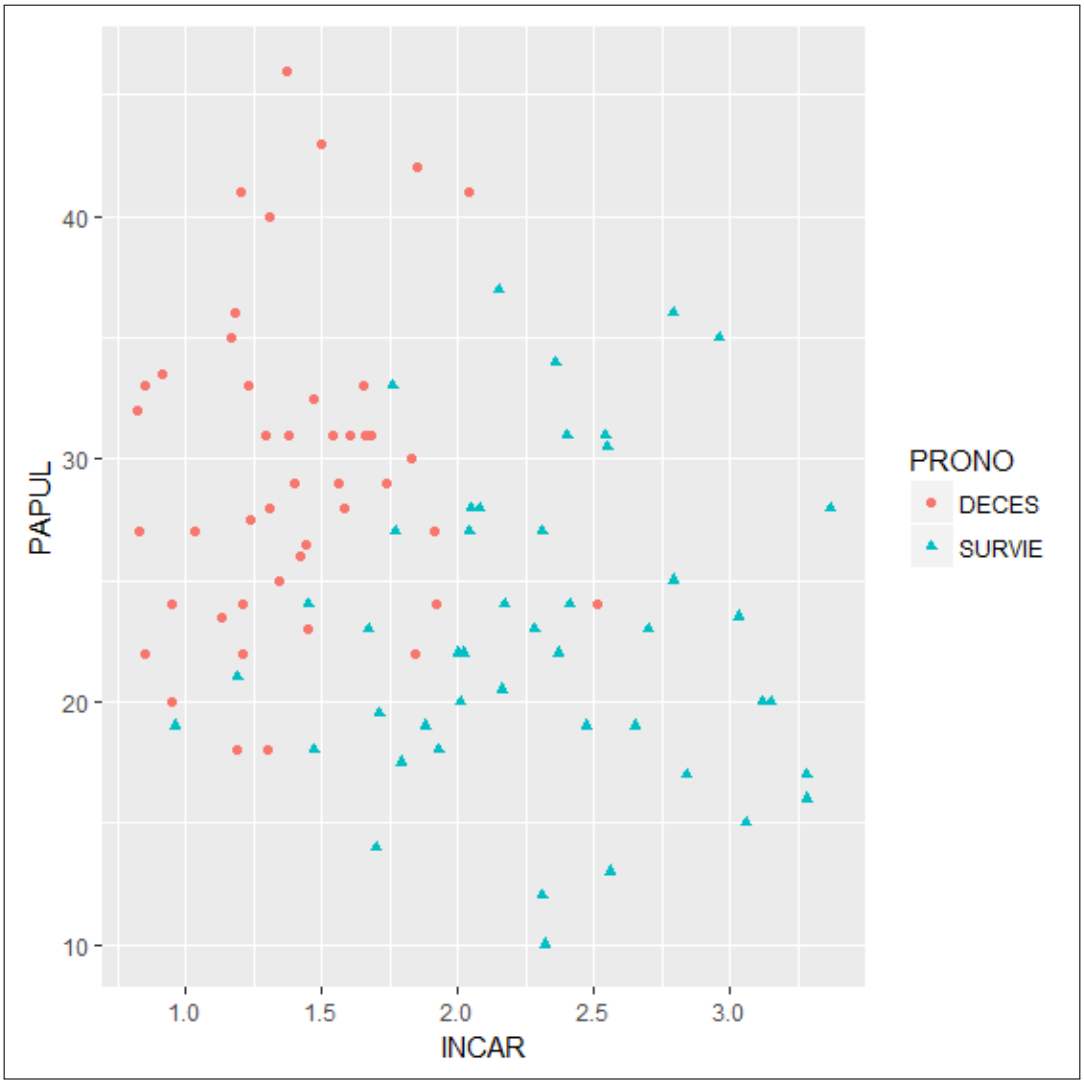


FIGURE 3.1 – Nuage de points des individus pour les deux variables INCAR et PAPUL

Tableau 3.3 représente les statistiques uni variées pour chacun des groupes à savoir les moyennes et écarts types locaux (par groupe) pour chacun des variables explicatives.

Groupe	PRONO	Moyenne	Écart-type
SURVIE	FRCAR	87.97826	14.10199
	INCAR	2.297609	0.5741203
	INSYS	26.80870	8.426198
	PRDIA	16.22174	4.967848
	PAPUL	22.36957	6.099576
	PVENT	8.152174	4.105375
	REPUL	803.2661	318.1262
DECES	FRCAR	94.80000	18.55410
	INCAR	1.375556	0.3656784
	INSYS	14.98667	4.601413
	PRDIA	21.75556	5.240383
	PAPUL	29.07778	7.211068
	PVENT	10.577778	4.158720
	REPUL	1811.4889	749.8438

TABLE 3.3 – Statistiques uni variées

On remarque pour chacune des variables des différences entres les moyennes des deux groupes, idem pour les écarts-types. Ces remarques peuvent être confirmées ou invalidées par des tests statistiques. On vérifie s'il existe bien de différences entre les groupes grâce au test de F ou test ANOVA uni varié, et le test de LAMBDA de Wilks...

3.3.1 Test ANOVA uni varié

L'analyse de la variance conduit à décomposer la somme des carrés des écarts à la moyenne globale de l'échantillon entre la somme des carrés des écarts à la moyenne locale pour les observations de chacun des groupes (intra-classes, ou W comme Within) et la somme des carrés des écarts des moyennes locales à la moyenne globale pour chacun des groupes (interclasses, ou B comme Between).

$$SCE_V = SCE_W + SCE_{BQ}$$

En divisant chacun des termes de cette équation par les degrés de libertés correspondant nombre de valeurs indépendantes dans les sommations effectuées, on aboutit à la notion de carré moyen

intra-classes CM_W et carré moyen interclasses CM_W permettant une comparaison de la variance intra-classes et de la variance interclasses :

$$SC_W = \frac{SCE_W}{(n-q)} \qquad \text{et} \qquad CM_B = \frac{SCE_B}{(q-1)}$$

Alors, la statistique F du rapport de variance définie par :

$$F = \frac{CM_B}{CM_W}$$

Suit une distribution de Fisher à $(q-1)$ et $(n-q)$ degrés de liberté sous l'hypothèse nulle H_0 d'égalité des moyennes entre les q groupes.
Au seuil de risque choisi α, l'hypothèse nulle sera rejetée si F est supérieur au $(1-\alpha)$ quantile $F_{(1-\alpha)}[(q-1)(n-q)]$ d'une distribution de Fisher.
Le calcul du F uni varié pour chaque variable j s'effectue alors directement à partir des termes diagonaux des matrices B et W.

$$F_j = \frac{(v_{jj}-w_{jj})(n-q)}{w_{jj}(q-1)} = \frac{b_{jj}(n-q)}{w_{jj}(q-1)}$$

Le tableau 4 tests d'égalité des moyennes des groupes (tableau d'ANOVA)

Variable	F-value	P-value	Signification
FRCAR	3.9099	0.0511	non significative
INCAR	83.084	2.171e-14	significative***
INSYS	68.566	1.152e-12	significative***
PRDIA	26.735	1.423e-06	significative***
PAPUL	22.993	6.484e-06	significative***
PVENT	7.8394	0.006269	significative**
REPUL	70.251	7.129e-13	significative***

TABLE 3.4 – Tests d'égalité des moyennes des groupes

• Ainsi pour une valeur $F = 3.9099$ et un risque $0.0511 > 0.05$, nous sommes conduits à accepter l'hypothèse nulle d'égalité des moyennes pour la variable FRCAR donc il n'y a pas une différence entre les moyenne c-à-d la variable FRCAR est non statiquement significative.
• Pour une valeur $F = 83.084$ est plus grand et un risque $2.171e-14 < 0.05$, nous sommes conduits à accepter l'hypothèse H_1 d'égalité des moyennes pour la variable INCAR donc il y a

une différence entre les moyenne c-à-d la variable INCAR est très statiquement significative.

L'examen du F dans notre exemple nous confirme que ce sont les variables : INCAR, INSYS et REPUL, qui sont les plus discriminant et plus élevés que les autres (puisque F représente le rapport entre la variance la variance interclasses et la variance intra classe alors on prend la valeur maximale).

3.3.2 Test de LAMBDA de wilks et ces extensions

Test de wilks : Le critère de Wilks est le produit des pourcentages de variance intra-classes

$$W = \prod_{k=1}^{r} (1 - \lambda_k)$$

La valeur du lambda de Wilks uni varié varie entre 0 et 1. La valeur 1 signifie l'égalité des moyennes pour l'ensemble des groupes. Une valeur quasi-nulle est associée à de très faibles variabilités intra-classes donc à de très fortes variabilités interclasses et des moyennes de groupes manifestement différentes. Il convient d'étudier globalement pour l'ensemble des descripteurs la décomposition de la variance é afin d'obtenir une vision synthétique des différences intergroupes et des différences interindividuelles (ou intra-classes), ce qui nécessite de passer de l'étude partielle de chacun des descripteurs à celle globale des matrices. Dans les analyses multivariées, on observe des liaisons entre les variables explicatives. Cependant, de fortes corrélations entre ces les variables peuvent conduire à une forte variabilité dans les estimations des coefficients de la fonction discriminante. Il est donc indispensable d'examiner les matrices de variance-covariance afin de détecter de semblables situations. En effet on a 5 matrices en tout dans notre exemple :

- La matrice globale notée V
- La matrice intra classe notée W
- La matrice intra de la première classe notée V_1
- La matrice intra de la deuxième classe notée V_2
- Et enfin la matrice inter classe notée B.

Test de pillai : Le critère de Pillai est la somme des valeurs propres de l'analyse discriminante

$$P = \sum_{k=1}^{r} \lambda_k$$

Test de Hotelling-Lawley : Le critère de Hotelling-Lawley est la somme des valeurs propres de l'analyse discriminante

$$W = \sum_{k=1}^{r} \mu_k$$

Test de Roy : Le critère de Roy est la plus grande valeur propre

$$W = \max \mu_k$$

Les valeurs propres de l'analyse permettent alors de tester la valeur discriminante de une ou toutes les combinaisons ainsi construites. C'est la MANOVA (Multivariate Analysis Of VAriance) qui généralise l'ANOVA. Évidemment, ici la MANOVA est extrêmement significative puisque les anova simples sont très significatives. Il y a plusieurs variantes.
Le tableau 5 montre les tests des moyennes.

Test	LAMBDA	F	P-value
Wilks	0.43484	15.411	9.099e-13
Pillai	0.56516	15.411	9.099e-13
Hotelling	1.2997	15.411	9.099e-13
Roy	1.2997	15.411	9.099e-13

TABLE 3.5 – Différents tests des moyennes

• La valeur de Lambda de Wilks étant faible, et est égale à 0.43484, et donc plus proche de 0 que de 1, avec une valeur $F = 15.411$ et un risque $9.099e - 13 < 0.05$, nous sommes conduits à rejeter l'hypothèse nulle d'égalité des moyennes pour tous les variables. Cela veut dire qu'au niveau global, la différence des moyennes entre les deux groupes est significative.
• Pour les autres test (Pillai, Hotelling et Roy) de valeur $F = 15.411$ et un risque $9.099e - 13 < 0.05$, nous sommes conduits à rejeter l'hypothèse nulle d'égalité des moyennes pour tous les variables. Cela veut dire qu'au niveau global, la différence des moyennes entre les deux groupes est significative.
• Donc d'après le test d'ANOVA uni varié et le test de Wilks, il existe bien une différence entre les groupes, donc on va commencer d'appliquer l'AD.

3.4 Approche géométrique

3.4.1 Valeurs propres

Les méthodes factorielles d'analyse multivariée reposent sur les concepts de la géométrie euclidienne faisant intervenir la notion d'inertie, produit de la masse (somme des pondérations

des observations) par la distance au carré (variance des descripteurs), correspondant à une somme pondérée des écarts au carré.
Cherchant les **combinaisons linéaires** :

$$a = \sum_{j=1}^{p} a_j(x_{ij} - g_j)$$

des variables susceptibles de séparer du mieux possible les q groupes, on sélectionne celles présentant le meilleur compromis entre deux objectifs distincts : représenter les groupes à la fois comme homogènes (minimiser l'inertie intra-classes) et comme bien séparés (maximiser leur inertie inter-classes).
La recherche de variables discriminantes revient à trouver une combinaison linéaire a qui maximise le rapport de ces deux inerties :

$$\frac{a'Ba}{a'Va}$$

on obtient comme solution le vecteur a défini par : $V^{-1}Ba = \lambda a$ est le vecteur propre associé à la plus grande **valeur propre** λ de la matrice $V^{-1}B$.
Comme on a vu dans le chapitre 2, a est le vecteur propre de $V^{-1}B$ et il l'est aussi de $W^{-1}B$. Mais par contre les valeurs propres respectives correspondantes diffèrent et sont liées par la relation :

$$\mu = \frac{\lambda}{1-\lambda}$$

Les valeurs propres μ à valeurs sur l'intervalle $[0, +\infty[$ sont définies par les valeurs propres λ à valeurs sur l'intervalle [0,1[.

Fonction	**valeur propre**	**% de la variance**	**% cumulé**
1	λ=0.565 μ=1.299	100	100

TABLE 3.6 – Valeurs propres

Comme on dispose de deux classes alors on aura $2 - 1 = 1$ valeur propre. La valeur propre associée à la fonction linéaire discriminante permet de juger le pouvoir discriminant de cette fonction. En effet la valeur propre est égale à la variance inter-classe de la fonction linéaire discriminante, On a $\lambda = 0.565$.

Le pourcentage de variance expliquée rapporte la valeur propre à la somme totale des valeurs propres. Le pourcentage de la variance inter-groupe expliquée par la fonction discriminante est de 100 %.

3.4.2 Fonctions discriminantes

Coefficients de la fonction discriminante :

Ces coefficients peuvent être utilisés pour calculer les coordonnées d'une observation dans l'espace des facteurs discriminants à partir de ses coordonnées dans l'espace des variables initiales.
Ces coefficients sont les estimations des coefficients de l'équation :

$$a_1 = A_{11}X_1 + \cdots + A_{p1}X_p.$$

On observe le pouvoir discriminant des axes grâce au tableau 3.7 ci-dessous donnant les estimations des coefficients des fonctions discriminantes.

Variables	**coefficients des fonctions discriminantes**	
	Coefficients de fct lda(MASS) valeur propre λ	**Coefficients de fct discrimin (ade4) valeur propre μ**
FRCAR	-0.0225044651	-0.24948661
INCAR	**2.4155744459**	**1.06861408**
INSYS	-0.0671347703	-0.40109786
PRDIA	0.0457155053	0.17547304
PAPUL	**-0.0841139970**	**-0.41515026**
PVENT	-0.0219299850	-0.06232694
REPUL	-0.0003009566	-0.15230269

TABLE 3.7 – Coefficients des fonctions discriminantes

Et donc les fonctions discriminantes peuvent s'écrire comme suit :

$a_\lambda(i) = -0.0225044651 * FRCAR + 2.4155744459 * INCAR - 0.0671347703 * INSYS + 0.0457155053 * PRDIA - 0.0841139970 * PAPUL - 0.0219299850 * PVENT - 0.0003009566 * REPUL.$

$a_\mu(i) = -0.24948661 * FRCAR + 1.06861408 * INCAR - 0.40109786 * INSYS + 0.17547304 * PRDIA - 0.41515026 * PAPUL - 0.06232694 * PVENT - 0.15230269 * REPUL.$

Pour l'individu 1 de notre échantillon d'apprentissage est égale à : (on le lit dans le tableau des données)

- $a_\lambda(1) = -0.0225044651(110) + 2.4155744459(0.96) - 0.0671347703(8.8) + 0.0457155053(15.0) - 0.0841139970(19.0) - 0.0219299850(16.0) - 0.0003009566(1583.00) = -1.05468675.$

$a_\mu(1) = -0.24948661(110) + 1.06861408(0.96) - 0.40109786(8.8) + 0.17547304(15.0) - 0.41515026(19.0) - 0.06232694(16.0) - 0.15230269(1583.00) = -1.581736852.$

• Sans en avoir l'air, les deux fonctions sont cohérentes selon le score frontière.

• Les coordonnées factorielles discriminantes sont centrées (de moyenne nulle) relativement à l'ensemble de l'échantillon d'apprentissage.

• Pour interpréter une fonction linéaire discriminante, l'analyse des coefficients standardisés(sans constant) est plus pertinente. Pour la fonction linéaire discriminante, les mesures sur l'index cardiaque(INCAR) et la pression artérielle pulmonaire (PAPUL) s'opposent aux mesures effectuées sur la fréquence cardiaque (FRCAR), l'index systolique (INSYS), la pression diastolique (PRDIA), la pression ventriculaire (PVENT) et la résistance pulmonaire (REPUL), Notons que les signes sont arbitraires : seules les oppositions de signes ont un sens.

3.4.3 Critère d'affectation géométrique

Les coordonnées factorielles des barycentres de groupe sur les axes discriminants sont évaluées comme valeurs moyennes des groupes.

Les tableaux 3.8 et 3.9 ci-dessous donnent l'estimation des valeurs moyennes des groupes.

FONCTION	
PRONO	**Scores moyens**
DECES	-0.7600799
SURVIE	0.7435565

TABLE 3.8 – Barycentres moyennes de chaque groupe

Affectation selon notre modèle	**Valeur du score**
DECES	$a< -0.0082617$
SURVIE	$a \geq -0.0082617$

TABLE 3.9 – Affectation avec score frontière

• Les axes factoriels discriminants fournissent un système de représentation maximisant la variance inter-classes et minimisant la variance intra-classes de la partition des n individus en q groupes.

• Pour classer une observation X_0 parmi les q groupes, l'affectation géométrique consiste à projeter cette observation dans l'espace défini par les axes factoriels discriminants puis à calculer la distance de cette observation à chacun des q centres de gravité des groupes. La règle d'affectation est alors définie par la métrique utilisée, soit dans notre cas W^{-1} la métrique de Mahanalobis.

• Chaque score individuel discriminant est ensuite comparé aux deux scores moyens et affecté au groupe dont-il est le plus proche. Mais la question qui se pose est la suivante : à partir de quel score peut-on affecter les individus au groupe 1 **(SURVIE)** et non pas au groupe 2 **(DECES)** ? Pour ce faire, on doit déterminer un score qui joue le rôle de frontière entre les groupes. le score critique est égal à la moyenne des moyennes des scores des groupes. Dans notre cas, $(-0.7600799 + 0.7435565)/2 = -0.0082617$, ce score est égal à -0.0082617.

Donc le score frontière = -0.0082617

Cette situation nous emmène à constater que chaque groupe peut se classer selon la règle de décision (tableau de score frontière).

• Chaque individu est classée selon le score obtenu :

si la fonction scoring est supérieure à -0.0082617, on affecte l'individu au groupe des SURVIE sinon on l'affecte au groupe des DECES.

3.4.4 Validation par Resubstitution -échantillon d'apprentissage-

La première manière d'évaluer (Validation) un classificateur est d'étudier la table de confusion qui consiste à confronter les valeurs observées de la variable dépendante Y (classe originale) avec les valeurs prédites $\hat{Y}$ (l'affectation).

Nous avons appliqué le modèle sur notre échantillon d'apprentissage comportant n=91 individus. Le tableau 3.10 représente les deux fonctions scoring (1 er avec lda, 2 ème avec discrimin),

la classe originale et l'affectation(en gras les mal classés).

Individus	fct scoring(avec lda)	fct scoring(avec discrimin)	l'affectation	la classe original
57	-1.581736852	-1.05468675	**DECES**	SURVIE
3	-1.260866417	-0.84073346	DECES	DECES
72	-2.404767157	-1.60347535	DECES	DECES
88	-0.135304427	-0.09021968	DECES	DECES
19	2.562789801	1.70884331	SURVIE	SURVIE
91	-1.647035755	-1.09822742	DECES	DECES
46	2.936865613	1.95827343	SURVIE	SURVIE
87	-0.730323087	-0.48697233	DECES	DECES
74	1.110021378	0.74015146	SURVIE	SURVIE
96	-0.946285315	-0.63097384	DECES	DECES
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
31	0.747893762	0.49868829	SURVIE	SURVIE
47	-2.219017759	-1.47961946	DECES	DECES
39	-1.381157483	-0.92094238	DECES	DECES
14	1.882915245	1.25550957	SURVIE	SURVIE
2	-0.578934265	-0.38602774	DECES	DECES
60	1.910087096	1.27362750	SURVIE	SURVIE
69	1.626819192	1.08474721	SURVIE	SURVIE
100	1.328605369	0.88590114	SURVIE	SURVIE
48	-1.158842440	-0.77270487	DECES	DECES
37	0.456904060	0.30465918	SURVIE	SURVIE

TABLE 3.10 – Affectation des individus par la Re substitution

(Voir Table A.1 pour le tableau COMPLET)

Pour s'assurer que la fonction discriminante classifie bien les individus (victimes d'infarctus de myocarde) en sous-groupes, on analyse la matrice de confusion qui regroupe les individus bien classées et les mal classées. C'est le moyen le plus utilisé est aussi le plus « parlant ». La

matrice de confusion (tableau 3.11) de notre fonction score se présente comme suit :

			Classe(s) d'affectation prévue(s)		Total
			DECES	SURVIE	
Original	Effectif	DECES	41	4	45
		SURVIE	6	40	46
	%(le taux)	DECES	91.11%	8.89%	100%
		SURVIE	13.04%	86.96%	100%
		Total	47	44	

TABLE 3.11 – Matrice de confusion

Interprétation de matrice de confusion -échantillon d'apprentissage-

- On analyse un échantillon de 91 victimes d'infarctus de myocarde avec 46 survivront et 45 décèderont :

• sûr les 46 survivront, 40 seront affectés comme tels et 6 seront affectés comme décès.

• sûr les 45 décèderont, 41 seront affectés comme décès et 4 seront affectés comme survie.

- D'autre part, concernant les taux : on constate que le taux de bon classement s'établit donc à 89.01%((40+41)/91), ce taux peut se décortiquer ainsi :

• Le pourcentage de bien classé pour les décès est égal 91.11% (41/45).

• Le pourcentage de bien classé pour les survies est égal 86.96% (40/46).

- Par contre, le taux d'erreurs (mal classé) est égal seulement 10.99% (10/91). Toutefois, on distingue pour ce taux entre :

• l'erreur du premier type (classer les décès parmi les survies) : ce taux est égal à 8.89% (4/45).

• l'erreur du second type (classer les survies parmi les décès) : ce taux est égal à 13.04% (6/46).

Déduction :

On constate que le taux de bon classement est plus grand que le taux d'erreurs (89.01% $>$ 10.99%) ça signifie que ce modèle est bon, autrement dit plus que les taux bien classés sont élevés et les taux mal classés sont faibles plus que le modèle est bon.

3.4.5 Validation par échantillon test

Nous avons appliqué le modèle sur un échantillon test comportant n=10 individus.
Le tableau 3.12 représente les deux fonctions scoring (1 er avec lda, 2 ème avec discrimin), la classe originale et l'affectation(en gras les mal classés).

Individus	fct scoring(lda)	Affectation	classe originale
16	1.52626447	SURVIE	SURVIE
24	-2.09913063	DECES	DECES
25	0.08606245	**SURVIE**	DECES
34	-0.02184022	**DECES**	SURVIE
35	-2.35722768	DECES	DECES
44	0.59968230	SURVIE	SURVIE
56	0.31306582	**SURVIE**	DECES
61	-0.90696025	DECES	DECES
66	1.51137405	SURVIE	SURVIE
73	-0.28016512	DECES	DECES

TABLE 3.12 – Table de l'échantillon test

La matrice de confusion (tableau 3.13) de notre fonction score se présente comme suit :

			Classe(s) d'affectation prévue(s)		Total
			DECES	SURVIE	
Original	Effectif	DECES	4	2	6
		SURVIE	1	3	4
	%(le taux)	DECES	66.67%	33.33%	100%
		SURVIE	25%	75%	100%
		Total	5	5	

TABLE 3.13 – Matrice de confusion

Interprétation de matrice de confusion -échantillon test-

- On analyse un échantillon de 10 victimes d'infarctus de myocarde avec 6 décèderont et 4 survivront :

• sûr les 6 décèderont, 4 seront affectés comme décès et 2 seront affectés comme survie.
• sûr les 4 survivront, 3 seront affectés comme tels et 1 sera affecté comme décès.
- D'autre part, concernant les taux : on constate que le taux de bon classement s'établit donc à 70%((4+3)/10), ce taux peut se décortiquer ainsi :
• Le pourcentage de bien classé pour les décès est égal 66.67% (4/6).
• Le pourcentage de bien classé pour les survies est égal 75% (3/4).
- Par contre, le taux d'erreurs (mal classé) est égal seulement 30% (3/10). Toutefois, on distingue pour ce taux entre :
• l'erreur du premier type (classer les décès parmi les survies) : ce taux est égal à 33.33% (2/6).
• l'erreur du second type (classer les survies parmi les décès) : ce taux est égal à 25% (1/4).

Déduction

On constate que le taux de bon classement est plus grand que le taux d'erreurs ($70\% > 30\%$) ça signifie que ce modèle est bon.

Si les dispersions des groupes sont très différentes à la fois en taille et en orientation, la règle géométrique de classement peut conduire à des taux de mal-classés importants. Pour pallier ces insuffisances inhérentes au point de vue géométrique, il est parfois nécessaire d'adopter une démarche probabiliste susceptible de fournir des règles de classement optimales.

3.5 Approche probabiliste

L'approche décisionnelle en analyse discriminante est construite sur un raisonnement probabiliste qualifié de bayésien car il s'appuie sur le théorème de Bayes utilisant les probabilités conditionnelles et les probabilités a priori pour calculer les probabilités a posteriori.

3.5.1 Probabilités à priori

La probabilité a priori, $\mathbb{P}_j$ est la probabilité qu'un individu appartienne au groupe G_j c-a-dire au groupe des survivons en l'absence de tout autre information. Les proportions observées dans l'échantillon d'apprentissage peuvent fournir une estimation de ces probabilités a priori. Ces probabilités a priori peuvent également être estimées d'après d'autres sources statistiques comme par exemple le recensement de la population pour des classes d'âge ou bien le recensement de l'agriculture pour les catégories d'exploitation agricole. En l'absence de toute information, on

choisira les groupes équiprobables.

Le tableau 3.14 représente les probabilités a priori des groupes

• La probabilité a priori, $\mathbb{P}_1 = \mathbb{P}(G_1) = \frac{46}{91} = 50.55$ est la probabilité qu'un individu appartienne au groupe G_1 c-à-d au groupe « SURVIE ».

• La probabilité a priori, $\mathbb{P}_2 = \mathbb{P}(G_2) = \frac{45}{91} = 49.45$ est la probabilité qu'un individu appartienne au groupe G_2 c-à-d au groupe des DECES

Variable	Modalités	Effectifs	%
PRONO	SURVIE	46	50.55%
	DECES	45	49.45%

TABLE 3.14 – Probabilités à priori

3.5.2 Validation par Resubstitution -échantillon d'apprentissage-

• Affectation des individus selon la règle bayésienne :

Dans le cas présent, il s'agit de la règle linéaire correspondant à une métrique unique W^{-1}.

Le tableau 3.15 représente les résultats du classement pour chaque individu de l'échantillon.

L'individu i est affecté au classe SURVIE si $\mathbb{P}(G_1) > \mathbb{P}(G_2)$

Individus	P-DECES	P-SURVIE	Affectation	La classe originale
57	0.9711663554	0.0288336446	**DECES**	SURVIE
3	0.9423170584	0.0576829416	DECES	DECES
72	0.9953808662	0.0046191338	DECES	DECES
88	0.5634552228	0.4365447772	DECES	DECES
19	0.0029320880	0.9970679120	SURVIE	SURVIE
91	0.9750157078	0.0249842922	DECES	DECES
46	0.0012634326	0.9987365674	SURVIE	SURVIE
87	0.8315966766	0.1684033234	DECES	DECES
74	0.0722212976	0.9277787024	SURVIE	SURVIE
...	...	...	...	...
...	...	...	...	...
...	...	...	...	...
31	0.1497647616	0.8502352384	SURVIE	SURVIE
47	0.9929946183	0.0070053817	DECES	DECES
39	0.9554104305	0.0445895695	DECES	DECES
14	0.0134403535	0.9865596465	SURVIE	SURVIE
2	0.7782674662	0.2217325338	DECES	DECES
60	0.0126516451	0.9873483549	SURVIE	SURVIE
69	0.0236962674	0.9763037326	SURVIE	SURVIE
100	0.0453916434	0.9546083566	SURVIE	SURVIE
48	0.9284627862	0.0715372138	DECES	DECES
37	0.2534577915	0.7465422085	SURVIE	SURVIE

TABLE 3.15 – Affectation des individus par Resubstitution

• **La première colonne** donne le numéro de séquence i_0 de l'individu permettant son identification.

• **La seconde colonne** affiche la probabilité à posteriori de le groupe DECES

• **La troisième colonne** affiche la probabilité à posteriori de le groupe SURVIE.

• **La quatrième colonne** affiche le groupe d'affectation de l'individu i_0.

• **La cinquième colonne** affiche la classe originale de l'individu i_0 c-à-dire le groupe à la-quel il appartient réellement.

Matrice de confusion par la validation de Resubstitution : La matrice de confusion de la validation d'échantillon de Resubstitutions représente comme suit :

			Classe(s) d'affectation prévue(s)		Total
			DECES	SURVIE	
Original	Effectif	DECES	40	5	45
		SURVIE	7	39	46
	%(le taux)	DECES	88.89%	11.11%	100%
		SURVIE	15.22%	84.78%	100%
		Total	47	44	

TABLE 3.16 – Matrice de confusion

Interprétation de tableau de confusion -échantillon d'apprentissage- :

• Le taux de bien classé de la classe 1 (SURVIE) est égale à $\frac{39}{46} = 0.8478$

• Le taux de bien classé de la classe 2 (DECES) est égale à $\frac{40}{45} = 0.8889$

Donc le taux total de bien classé est égale à $\frac{0.85+0.89}{2} = 0.8684$

Les résultats du classement effectué selon la règle bayésienne d'affectation montrent un taux apparent global de bien-classés élevé (86.84%), moyenne pondérée du taux apparent de bien-classés pour chacun des groupes qui varie de 84.78% pour le groupe SURVIE à 88.89% pour le groupe DECES, Le calcul des probabilités a posteriori utilisées dans la règle bayésienne d'affectation n'étant fonction que de la valeur des distances généralisées des individus aux barycentres des groupes.

• Le taux de mal classé de la classe 1 (SURVIE) est égale à $\frac{7}{46} = 0.1522$

Le taux de mal classé de la classe 2 (DECES) est égale à $\frac{5}{45} = 0.1111$

Donc le taux total de mal classé est égale à $\frac{0.1522+0.1111}{2} = 0.1316$.

Déduction

On constate que le taux de bon classement est plus grand que le taux d'erreurs (86.84% $>$ 13.16%) ça signifie que ce modèle est bon.

3.5.3 Validation par échantillon test

• Affectation des individus selon la règle bayésienne :

L'application de la règle bayésienne d'affectation à chaque individu de l'échantillon de test selon cette procédure conduit aux résultats listés dans le tableau 3.17.

Individus	La classe originale	P-DECES	P-SURVIE	Affectation
16	SURVIE	0.02954927	0.970450731	SURVIE
24	DECES	0.99083991	0.009160092	DECES
25	DECES	0.43930246	0.560697545	**SURVIE**
34	SURVIE	0.49983275	0.500167253	SURVIE
35	DECES	0.99486085	0.005139151	DECES
44	SURVIE	0.19746328	0.802536725	SURVIE
56	DECES	0.31953831	0.680461692	**SURVIE**
61	DECES	0.88030367	0.119696330	DECES
66	SURVIE	0.03052753	0.969472470	SURVIE
73	DECES	0.64149598	0.358504017	DECES

TABLE 3.17 – Affectation des individus d'échantillon test

L'individu i est affecté au classe SURVIE si $\mathbb{P}(G_1) > \mathbb{P}(G_2)$

Matrice de confusion pour la validation de l'échantillon Test :

La matrice de confusion de la validation d'échantillon de Resubstitution représente comme suit :

			Classe(s) d'affectation prévue(s)		Total
			DECES	SURVIE	
Original	Effectif	DECES	4	2	6
		SURVIE	0	6	6
	%(le taux)	DECES	66.67%	33.33%	100%
		SURVIE	0%	100%	100%
		Total	4	8	

TABLE 3.18 – Matrice de confusion

Interprétation de tableau de confusion -échantillon test-

• Le taux de bien classé de la classe 1 (SURVIE) est égale à $\frac{4}{4} = 1$

• Le taux de bien classé de la classe 2 (DECES) est égale à $\frac{4}{6} = 0.6667$

Donc le taux total de bien classé est égale à $\frac{1+0.6667}{2} = 0.8334$

Les résultats du classement effectué selon la règle bayésienne d'affectation montrent un taux apparent global de bien-classés élevé (83.34%), moyenne pondérée du taux apparent de bien-classés pour chacun des groupes qui varie de 100% pour le groupe SURVIE à 66.67% pour le groupe DECES, Le calcul des probabilités a posteriori utilisées dans la règle bayésienne d'affectation n'étant fonction que de la valeur des distances généralisées des individus aux barycentres des groupes.

• Le taux de mal classé de la classe 1 (SURVIE) est égale à $\frac{0}{4} = 0$

• Le taux de mal classé de la classe 2 (DECES) est égale à $\frac{2}{6} = 0.3333$

Donc le taux total de mal classé est égale à $\frac{0+0.3333}{2} = 0.1666$

Déduction :

On constate que le taux de bon classement est plus grand que le taux d'erreurs (83.34% $>$ 16.66%) ça signifie que ce modèle est bon.

3.5.4 Comparaison entre la validation de l'approche géométrique et l'approche probabiliste

	Taux de validation éch-appr		Taux de validation éch-test	
	%bien classés	%mal classés	%bien classés	%mal classés
Appr-géom	89.01%	10.99%	70%	30%
Appr-prob	87%	13%	83%	17%

TABLE 3.19 – Comparaison entre l'approche géométrique et l'approche probabiliste

D'après le tableau 3.19 ci-dessus qui compare la validation dans les deux approches géométrique et probabiliste :

• Dans la validation d'échantillon d'apprentissage

On constate que l'approche géométrique a un grand taux bien classés par rapport l'approche probabiliste (89.01% > 87%)et un faible taux mal classés (10.99% < 13%), cela signifie que dans la validation d'échantillon d'apprentissage, l'approche géométrique est la meilleure.

par contre.

• Dans la validation d'échantillon test

On constate que l'approche probabiliste a un grand taux bien classés par rapport l'approche géométrique (84% > 70%)et un faible taux mal classés (16% < 30%), cela signifie que dans la validation d'échantillon test, l'approche probabiliste est la meilleure.

La validation par Resubstitution sous-estime souvent le taux d'erreur (taux de mal classé). Il est en effet biaisé car est estimé sur les données qui ont servi à définir la règle de décision.

Pour ces raisons qu'il est recommandé de baser sur la validation pour échantillon test sur laquelle on peut considérer que l'approche probabiliste est le meilleure que celle géométrique.

3.6 Validation croisée

3.6.1 Courbe de ROC

La courbe ROC (Receiver Operating Characteristics) permet de visualiser la performance d'un modèle, et de la comparer cette performance à celle d'autres modèles. Les termes utilisés viennent de la théorie de détection du signal. On désigne par sensibilité (sensitivity) la proportion d'événements positifs (SURVIE) bien classés. La spécificité (specificity) correspond à la proportion d'événements négatifs (DECES) bien classés. L'aire sous la courbe (ou Area Under the Curve – AUC) est un indice synthétique calculé pour les courbes ROC. L'AUC correspond à la probabilité pour qu'un événement positif ait une probabilité donnée par le modèle plus élevée qu'un événement négatif. Pour un modèle parfait, on a $AUC = 1$, pour un modèle aléatoire, on a $AUC = 0.5$. On considère habituellement que le modèle est bon dès lors que la valeur de l'AUC est supérieure à 0.7. Un modèle bien discriminant doit avoir une AUC entre 0.87 et 0.9. Un modèle ayant une AUC supérieure à 0.9 est excellent.

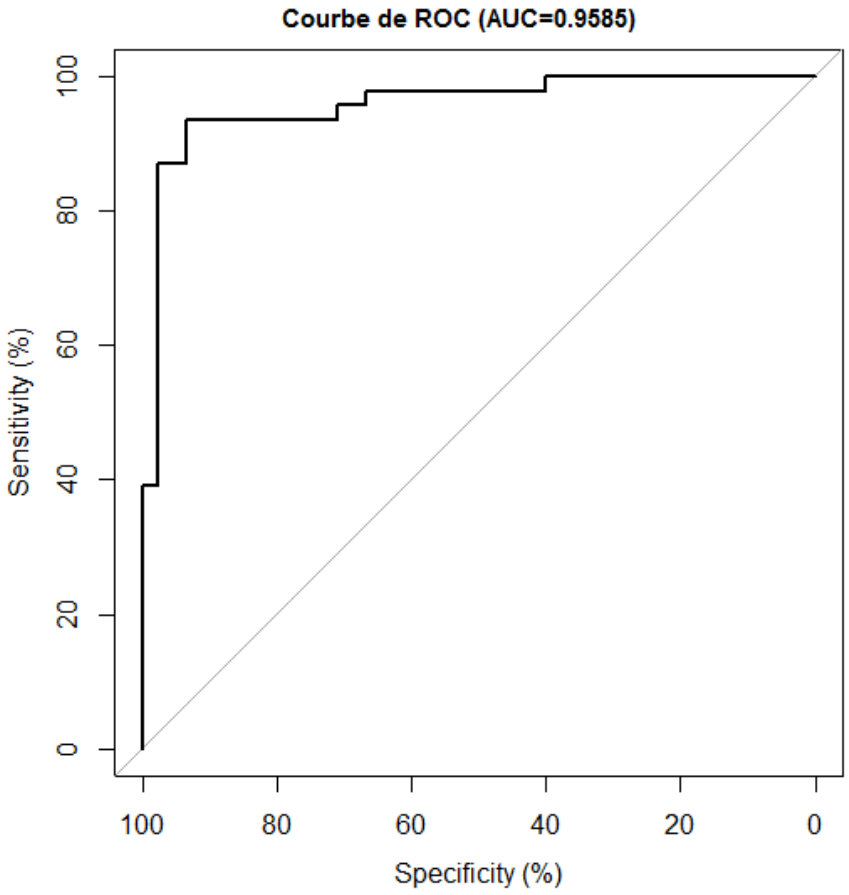


FIGURE 3.2 – Courbe de ROC

3.6.2 Interprétation de La courbe de ROC

On constate que l'aire sous la courbe est grande ($AUC = 0.9585$), elle est proche de 1 en s'éloignant au diagonale, alors notre modèle est excellent (discrimination excellente),autrement dit plus que l'aire est grande en s'éloignant au diagonale plus que le modèle est bon.

Conclusion

L'analyse Discriminante est depuis des décennies la technique de référence en classification. Elle est simple, bien comprise sur le plan théorique, et raisonnablement efficace sur la plupart des problèmes ordinaires.

Contrairement aux méthodes de classification dont le but est de construire des groupes ou des classes les plus homogènes et les plus distincts dans un échantillon possible, les méthodes d'analyse discriminante connues aussi sous le nom de classification supervisée d'apprentissage statistique ou de reconnaissance statistique des formes ou encore méthodes de classement, sont basées sur une division des individus en un certain nombre de groupe définis à priori, à l'aide des modalités d'une des variables de l'échantillon. Le problème est le classement ou l'affectation d'un nouvel individu dans l'un de ces groupes.
C'est une méthode de description de groupes. Elle fournit un outil d'évaluation et d'interprétation des résultats (valeurs propres, vecteurs propres significativités des axes, coefficients ...). Elle est connectée à d'autres méthodes factorielles telles que l'ACP. Mais aussi avec des méthodes prédictives en particulier l'analyse discriminante prédictive. Notre problème -comme il est déjà signalé à l'introduction générale- consiste à analyser les relations existant entre la variable à expliquer et les autres variables du tableau qui constituent l'ensemble des variables explicatives. Dans le cas où la variable à expliquer est de nature quantitative, la méthode paramétrique explicative correspondante est connue sous le nom de régression multiple.

On s'est intéressé dans notre mémoire au cas où la variable à expliquer est qualitative nominale binaire (variable PRONO à deux modalité : Survie et Décès) et 07 variables explicatives quantitatives. On a modélisé le phénomène étudié par les deux approches géométrique et probabiliste sur lesquelles nous avons appliquée les différentes règles de validation à savoir la méthode de Ré-substitution et l'échantillon test, la méthode probabiliste a donnée de résultats meilleurs comparativement à la méthode géométrique.

Nous espérons avoir répondu au problème posé et résultats trouvés seront d'une utilité pertinente.

Bibliographie

[1] Allain Baccini et PHilippe Besse, **Statistique descriptive multidimensionnelle**, université de Paul Sabatier, Toulouse, 1999.

[2] Benzekri J-P, **Analyse discriminante et analyse factorielle, les cahiers de l'analyse des données, tome 2, n°4(1977) p.369-406.**

[3] Gérard Gouaet, **Analyse de données**, Lavoisier, 2003.

[4] Gilbert Saporta, **Probabilité analyse des données et statistique**, paris, Technip, 1990.

[5] Jean Jacques Droesbeke, Michel L ejeune, Gilbert Saporta, **Modèle statistiques pour données qualitatives**, Technip, 2005.

[6] Jean-Pienne Nakache, Josiane Confais, **statistique explicative appliquée**, Technip, 2003.

[7] Ludovic Lebart, Alain Morineau et Marie Piron, **Statistique exploratoire multidimensionnelle**.

[8] Mireille Bardos, **Analyse discriminante application au risque et soring financier**, Dunod, 2001.

[9] Pierre Lafaye, Mecheaux Rémy et Drouihet Benoit Liquet, **Le logiciel R, Maitriser le langage Effectuer des analyses statistiques.**

[10] Ricco RAKOTOMALALA, **Analyse discriminante linéaire**, Laboratoire ERIC.

[11] Vincent Richard, **S'initier à l'analyse des données avec le logiciel R.**

Annexe A

Annexe

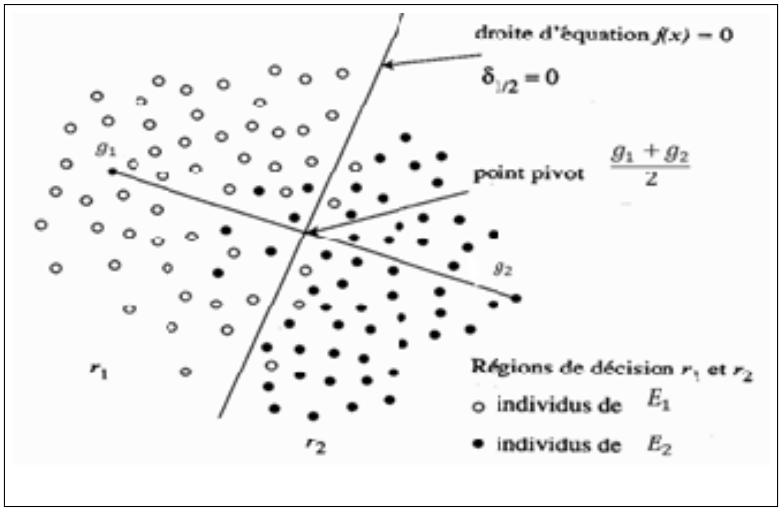


FIGURE A.1 – Mahalanobis-Fisher cas de deux groupes

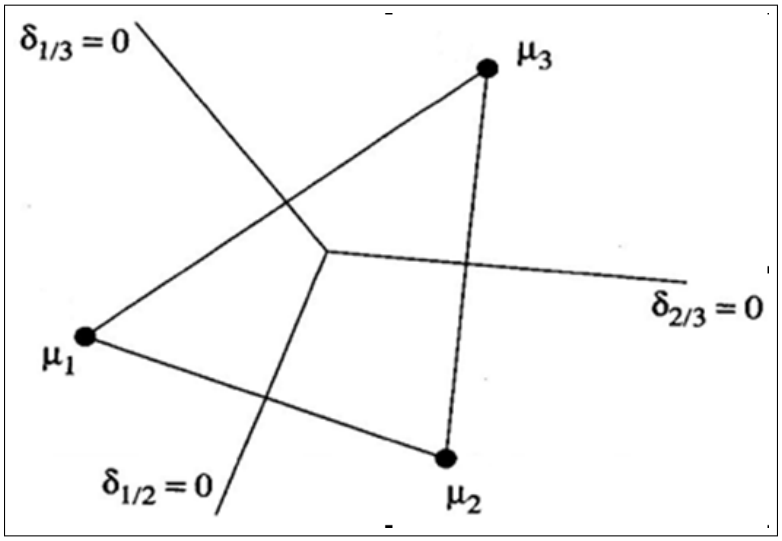


FIGURE A.2 – Mahalanobis-Fisher cas de trois groupes

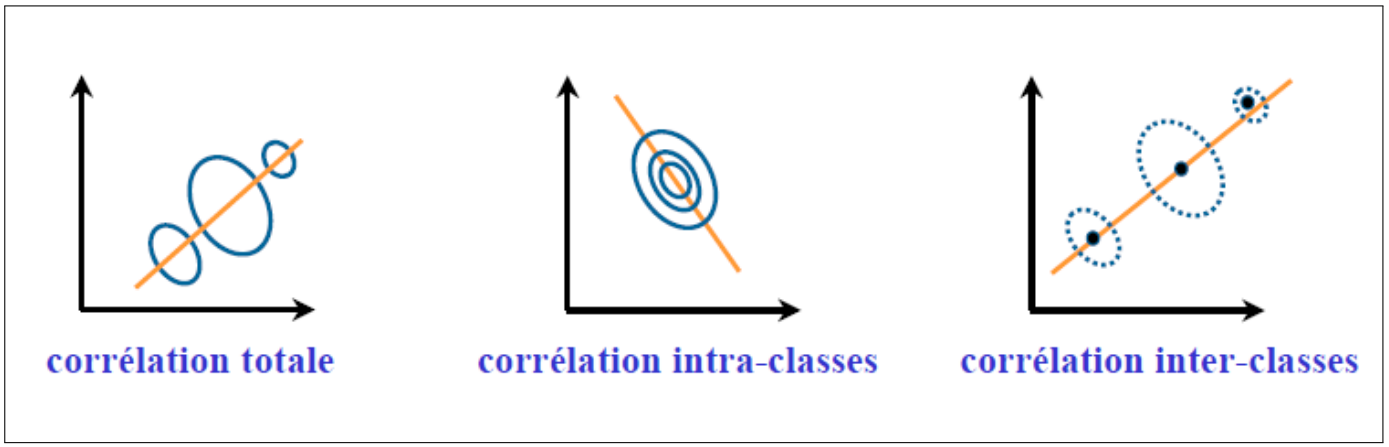


FIGURE A.3 – Décomposition de corrélation totale -intraclasses-interclasses

Individus	fct scoring(avec lda)	fct scoring(avec discrimin)	l'affectation	la classe original
57	-1.581736852	-1.05468675	**DECES**	SURVIE
3	-1.260866417	-0.84073346	DECES	DECES
72	-2.404767157	-1.60347535	DECES	DECES
88	-0.135304427	-0.09021968	DECES	DECES
19	2.562789801	1.70884331	SURVIE	SURVIE
91	-1.647035755	-1.09822742	DECES	DECES
46	2.936865613	1.95827343	SURVIE	SURVIE
87	-0.730323087	-0.48697233	DECES	DECES

74	1.110021378	0.74015146	SURVIE	SURVIE
96	-0.946285315	-0.63097384	DECES	DECES
1	0.200859017	0.13393084	SURVIE	SURVIE
22	-0.722307719	-0.48162776	DECES	DECES
92	-1.964802387	-1.31011112	DECES	DECES
95	-4.036925025	-2.69178234	DECES	DECES
63	-2.294510081	-1.52995701	DECES	DECES
30	-0.168825841	-0.11257143	**DECES**	SURVIE
41	-1.782381115	-1.18847439	DECES	DECES
85	-0.911547362	-0.60781091	DECES	DECES
17	0.611482333	0.40773047	SURVIE	SURVIE
4	0.633664589	0.42252138	SURVIE	SURVIE
40	1.129834006	0.75336233	SURVIE	SURVIE
86	-0.588150826	-0.39217325	DECES	DECES
36	0.922196560	0.61491170	SURVIE	SURVIE
6	-0.964003587	-0.64278822	DECES	DECES
77	0.081698912	0.05447604	SURVIE	SURVIE
52	2.762229606	1.84182784	SURVIE	SURVIE
78	-1.079273274	-0.71964892	DECES	DECES
65	3.225408796	2.15067121	SURVIE	SURVIE
83	-0.657437211	-0.43837274	DECES	DECES
33	-0.389073352	-0.25943033	DECES	DECES
9	0.988750388	0.65928914	SURVIE	SURVIE
18	0.598104207	0.39881007	SURVIE	SURVIE
84	-0.971072109	-0.64750144	DECES	DECES
11	1.775561609	1.18392721	SURVIE	SURVIE
82	-2.068359429	-1.37916195	DECES	DECES
97	2.403758350	1.60280269	SURVIE	SURVIE
15	2.686912847	1.79160735	SURVIE	SURVIE
89	-2.024275617	-1.34976729	DECES	DECES
50	-0.656022756	-0.43742959	DECES	DECES
94	-2.059781342	-1.37344216	DECES	DECES
81	-0.366445738	-0.24434245	DECES	DECES

7	0.503209692	0.33553533	SURVIE	SURVIE
10	1.097940947	0.73209634	SURVIE	SURVIE
38	0.650294122	0.43360980	SURVIE	SURVIE
67	-2.625307658	-1.75052961	DECES	DECES
62	0.071725145	0.04782563	SURVIE	SURVIE
76	-0.636181396	-0.42419957	DECES	DECES
64	-0.494891351	-0.32998874	**DECES**	SURVIE
55	0.762580305	0.50848113	SURVIE	SURVIE
70	1.120918915	0.74741783	**SURVIE**	DECES
58	2.229601071	1.48667631	SURVIE	SURVIE
26	-0.384488392	-0.25637312	DECES	DECES
93	0.389080661	0.25943520	**SURVIE**	DECES
98	0.546048620	0.36409991	SURVIE	SURVIE
68	-1.413370012	-0.94242138	DECES	DECES
23	0.101136665	0.06743694	**SURVIE**	DECES
51	1.308515417	0.87250535	SURVIE	SURVIE
27	-0.352959768	-0.23535014	DECES	DECES
42	-1.721978343	-1.14819841	DECES	DECES
49	0.004023454	0.00268280	**DECES**	SURVIE
53	-1.416006157	-0.94417913	DECES	DECES
54	-1.159635025	-0.77323336	DECES	DECES
8	-0.842556777	-0.56180866	**DECES**	SURVIE
75	1.016757028	0.67796370	SURVIE	SURVIE
43	0.380640608	0.25380745	SURVIE	SURVIE
21	2.718590200	1.81272950	SURVIE	SURVIE
80	-2.165016033	-1.44361163	DECES	DECES
13	2.164658898	1.44337350	SURVIE	SURVIE
71	1.892960705	1.26220779	SURVIE	SURVIE
29	0.946734645	0.63127345	SURVIE	SURVIE
101	-0.130334339	-0.08690567	**DECES**	SURVIE
99	0.964281906	0.64297380	SURVIE	SURVIE
79	-1.684707811	-1.12334678	DECES	DECES
28	0.277201755	0.18483543	**SURVIE**	DECES

5	-0.628615803	-0.41915490	DECES	DECES
90	-0.142041057	-0.09471160	DECES	DECES
20	0.914205692	0.60958346	SURVIE	SURVIE
59	2.036323814	1.35780091	SURVIE	SURVIE
32	-2.235803510	-1.49081203	DECES	DECES
45	-0.619209652	-0.41288297	DECES	DECES
12	1.721775811	1.14806336	SURVIE	SURVIE
31	0.747893762	0.49868829	SURVIE	SURVIE
47	-2.219017759	-1.47961946	DECES	DECES
39	-1.381157483	-0.92094238	DECES	DECES
14	1.882915245	1.25550957	SURVIE	SURVIE
2	-0.578934265	-0.38602774	DECES	DECES
60	1.910087096	1.27362750	SURVIE	SURVIE
69	1.626819192	1.08474721	SURVIE	SURVIE
100	1.328605369	0.88590114	SURVIE	SURVIE
48	-1.158842440	-0.77270487	DECES	DECES
37	0.456904060	0.30465918	SURVIE	SURVIE

TABLE A.1: Affectation par Resubstitution -méth géométrique- (TABLEAU COMPLET)

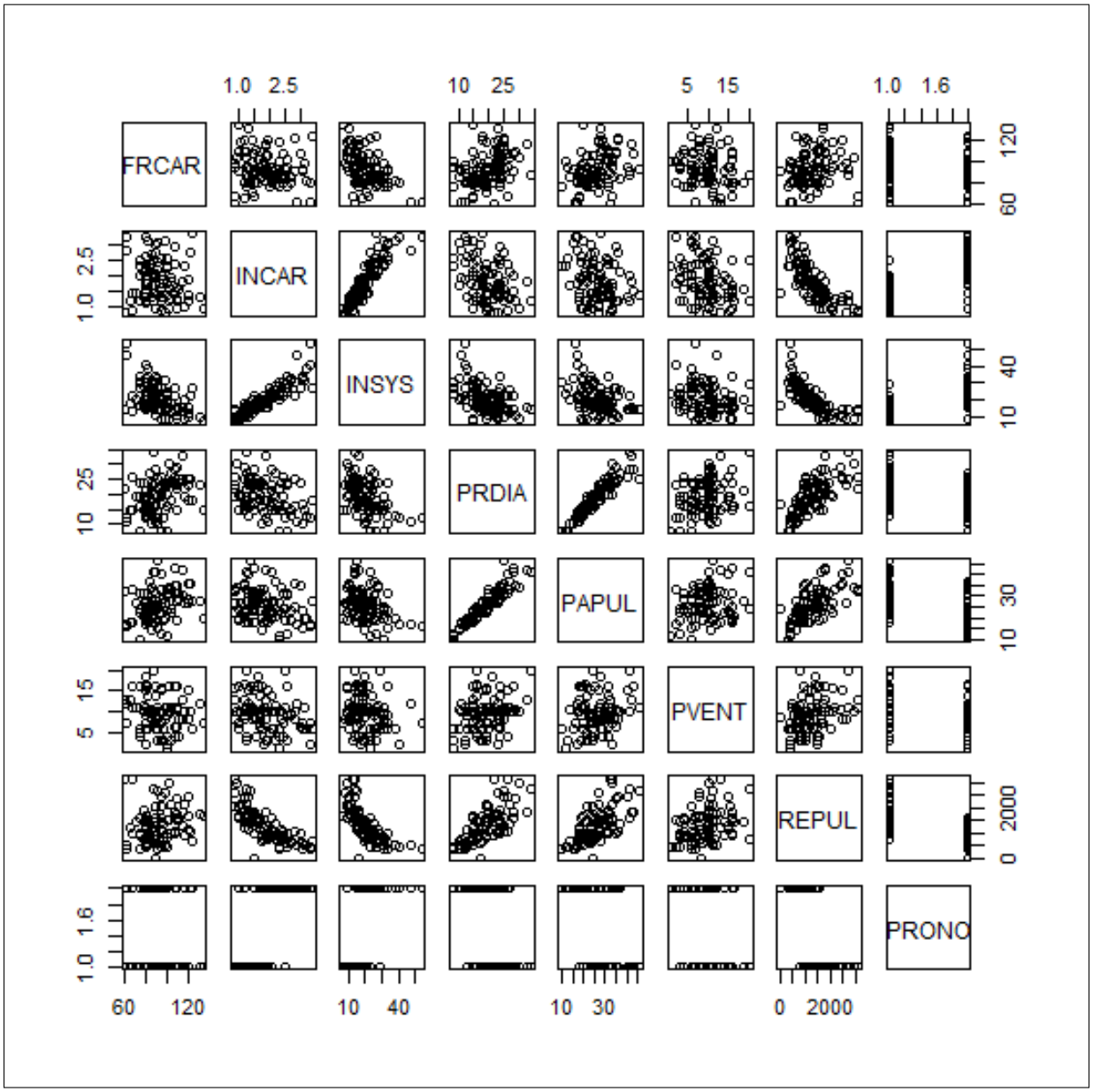


FIGURE A.4 – Représentation graphique bi varié des variables

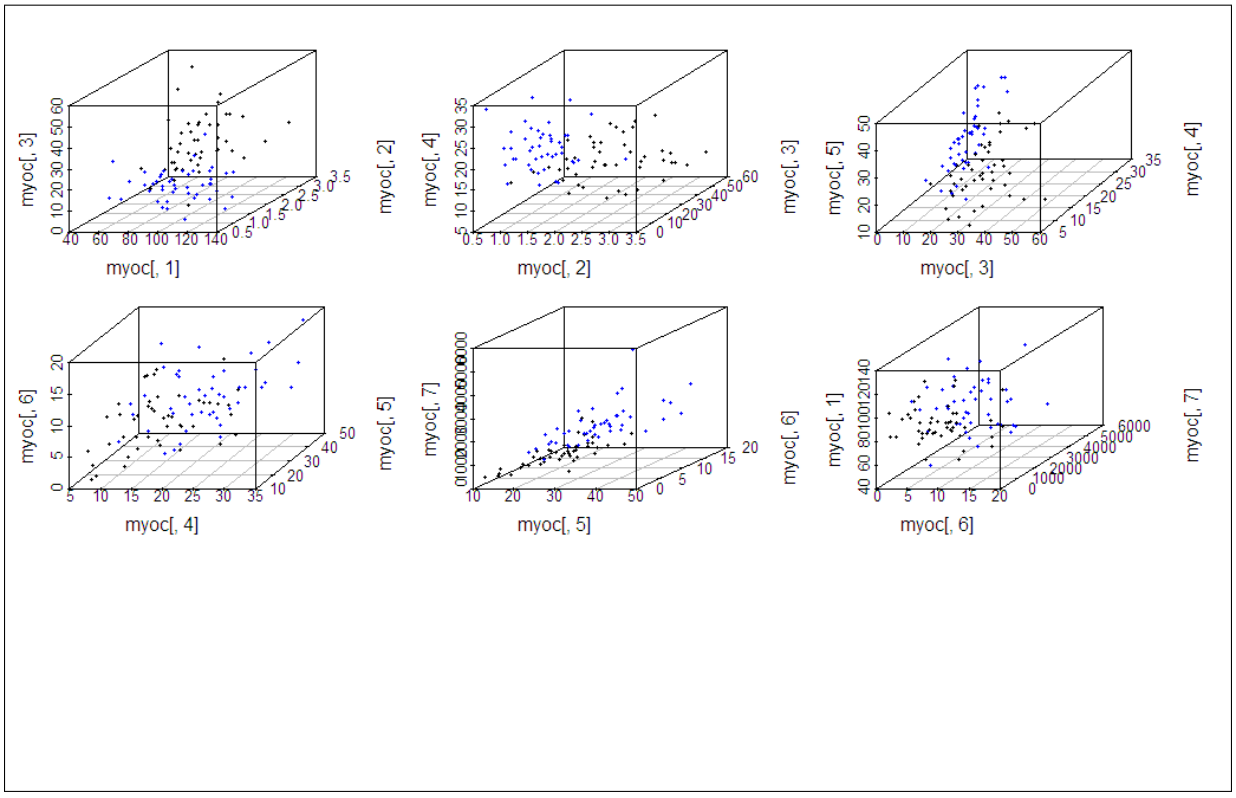


FIGURE A.5 – Représentation trois dimension des variables

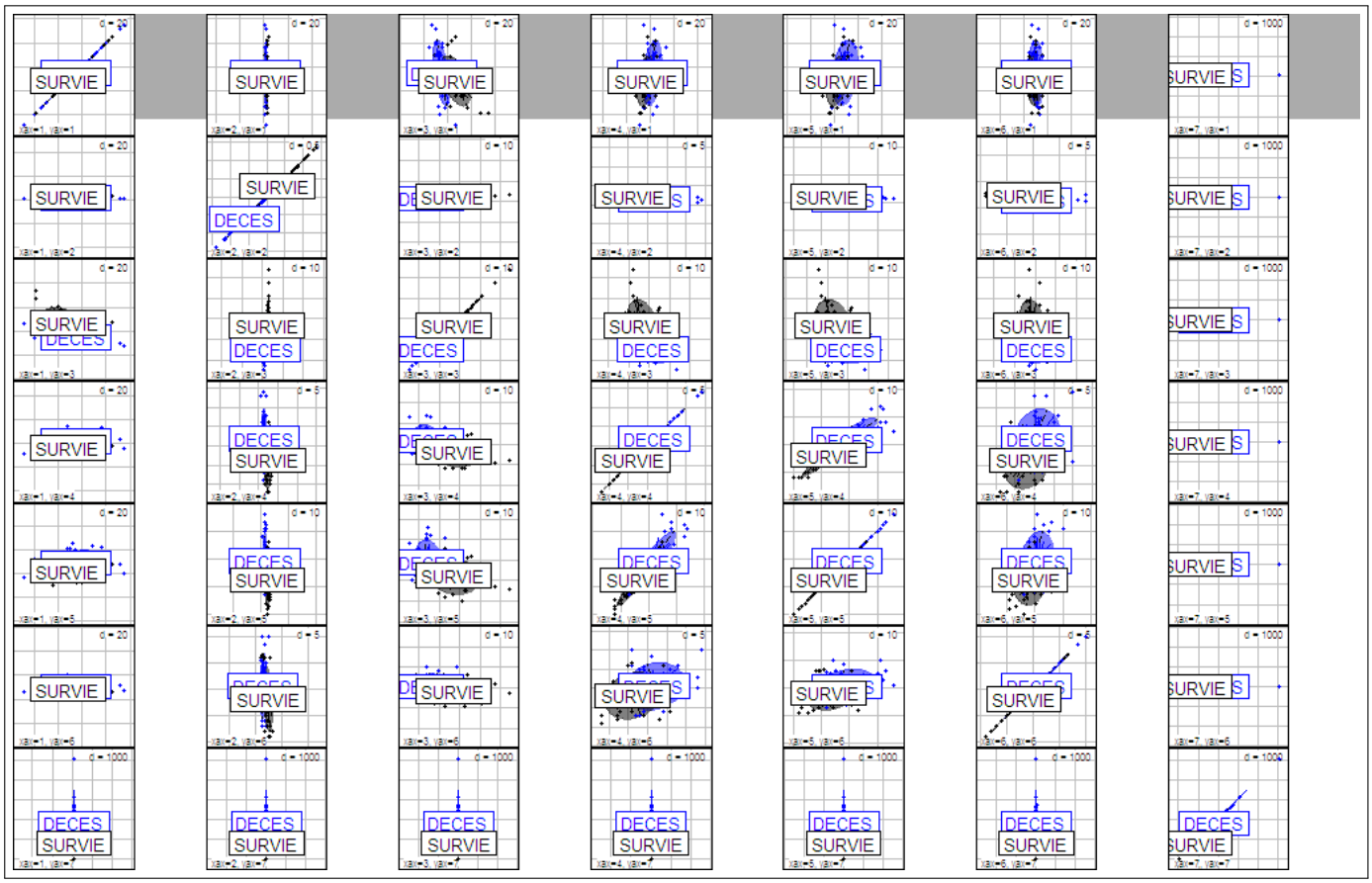


FIGURE A.6 – Représentation bi varié des groupes

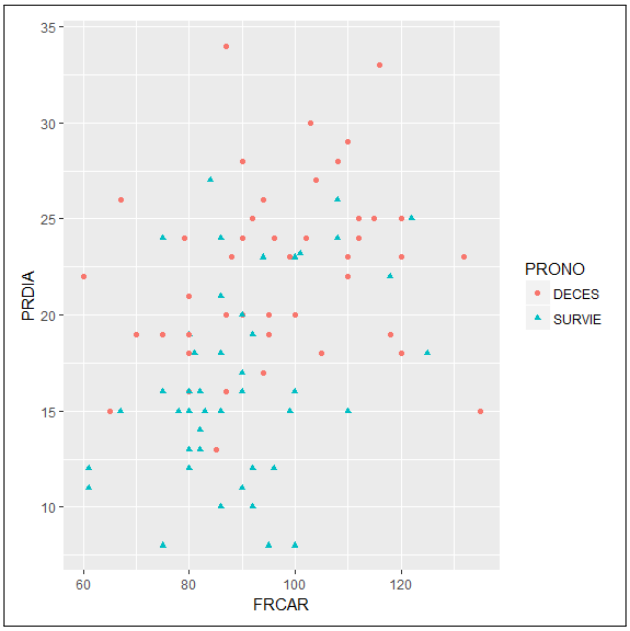


FIGURE A.7 – Nuage de points des individus pour les deux variables PRDIA et FRCAR

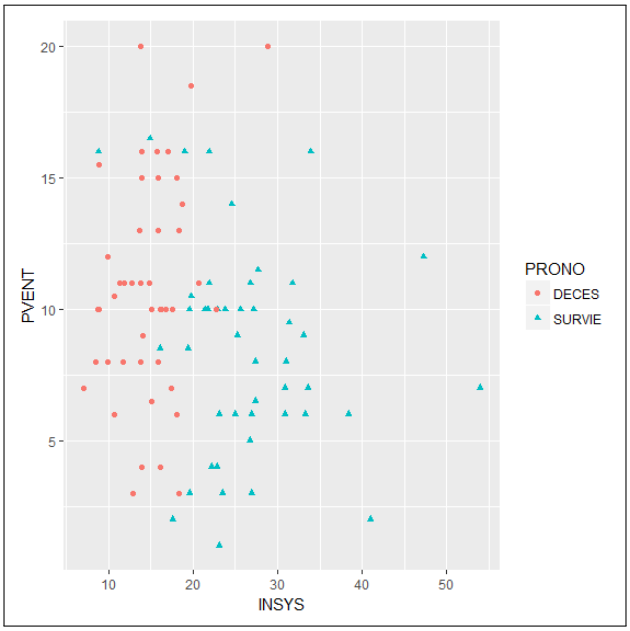


FIGURE A.8 – Nuage de points des individus pour les deux variables PVENT et INSYS

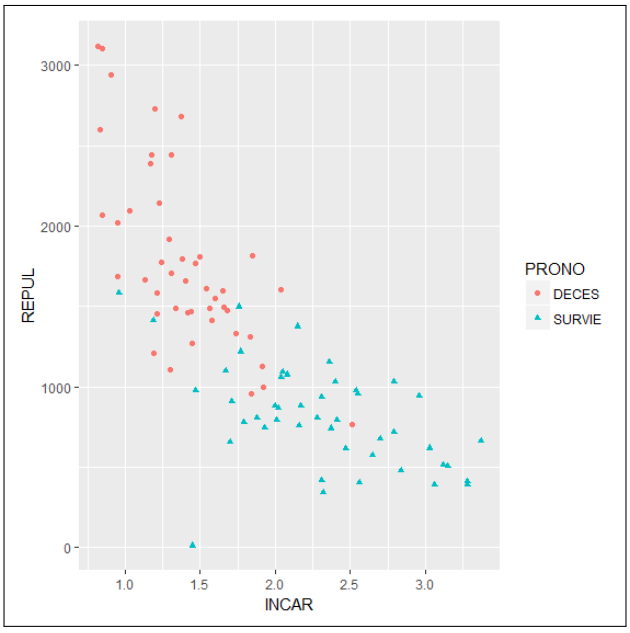


FIGURE A.9 – Nuage de points des individus pour les deux variables REPUL et INCAR

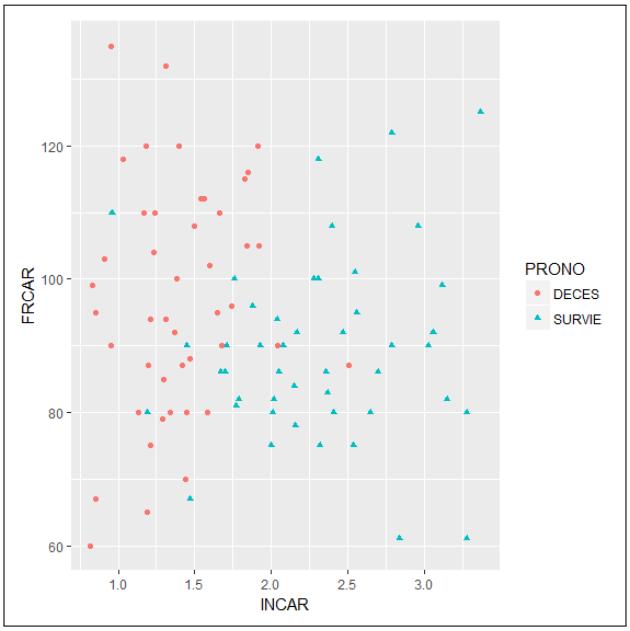


FIGURE A.10 – Nuage de points des individus pour les deux variables FRCAR et INCAR

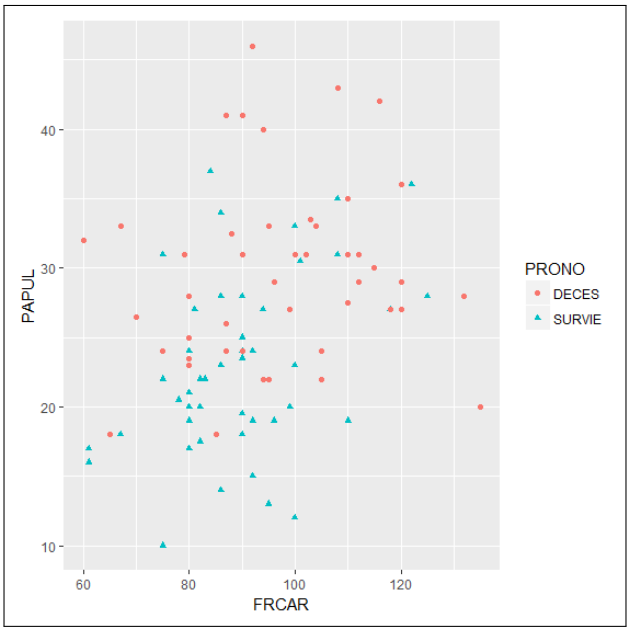


FIGURE A.11 – Nuage de points des individus pour les deux variables PAPUL et FRCAR