

REPUBLIQUE ALGERIENNE DEMOCRATIQUE et POPULAIRE.
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique.



UNIVERSITE MOULoud MAMMERI, TIZI-OUZOU
Faculté des Sciences Département de Mathématiques

MEMOIRE DE MASTER en Mathématiques

Spécialité : Mathématiques Appliquées à la Gestions

Etude des indicateurs du développement et classification des pays africains : Approche Statistique

présenté par

Rachid BOUNSIAR & Sabrina TEBANI

Devant le jury composé de

Fazia BEDOUHENE	Professeur	U.M.M.T.O	Présidente
Abdelghani HAMAZ	MCA	U.M.M.T.O	Rapporteur
Farida ACHEMINE	MCA	U.M.M.T.O	Examinatrice

soutenu le 23/ 09 / 2019

Remerciements

C'est avec une certaine émotion et beaucoup de sincérité que nous voulons remercier toutes les personnes qui nous ont soutenu durant la réalisation de ce mémoire. Nous tenons tout particulièrement à remercier M. Hamaz Abdelghani qui a dirigé ce travail, pour sa disponibilité, son soutien et ses conseils qui ont été nombreux et importants.

M. DAIF Hakim et AMEUR Chnini qui ont beaucoup œuvré pour la mise en valeur de mon travail, et nous avons apprécié l'ambiance d'équipe qui s'est formée au laboratoire LMPA. Nous avons beaucoup appris à ses côtés et nous lui adressons toute nos reconnaissances.

Nous remercions Mme Bedouhene Fazia, Mme Achemine Farida pour l'honneur qu'ils nous font en acceptant de faire partie du jury.

Nos plus profonds remerciements vont à nos parents. Tout au long de notre cursus, ils nous ont toujours soutenu, encouragé et aidé. Ils ont su nous donner toutes les chances pour réussir. Qu'ils trouvent, dans la réalisation de ce travail, l'aboutissement de leurs efforts ainsi que l'expression de notre plus affectueuse gratitude.

Ces remerciements ne seraient pas complets sans mentionner nos frères, sœurs et amis. Nous les remercions pour nous avoir fait partager leurs joie de vivre et de nous avoir ainsi soutenu dans nos efforts.

Introduction	1
1 Analyse en Composantes Principales	3
1.1 Analyse des composantes principales	3
1.1.1 Principe de L'ACP	3
1.1.2 Tableaux de données et leurs caractéristiques	4
1.1.3 Le problème de l'ACP	7
1.1.4 Interprétation et qualité de représentation d'une ACP	12
1.1.5 Avantage et inconvénient de l'ACP	14
2 Analyse discriminante et fonction score	15
2.1 Analyse discriminante descriptive	15
2.1.1 Variance intraclasse et interclasse	16
2.1.2 L'analyse factorielle discriminante (AFD)	17
2.1.3 Cas particulier de deux groupe	23
2.2 Analyse discriminante décisionnelle	24
2.2.1 Règles géométriques d'affictation	25
2.2.2 Analyse discriminante probabiliste	27
2.2.3 Analyse discriminante quadratique	30
2.2.4 Analyse discriminante linéaire	31
2.2.5 Cas particulier de deux groupe	32
2.2.6 Régression logistique	33
2.3 La méthode scoring	33
2.3.1 Différents types de score	34
2.3.2 Les différentes étapes de la construction d'un score	35
2.3.3 Validation d'une méthode de scoring	36
2.3.4 Les limites de la méthode scoring	41
3 Etude des facteurs du développement de l'Algérie, la Tunisie et le Maroc et classemnt des pays africains	43
3.1 Présentation et outils	43
3.1.1 Logiciel R et Packages	45
3.2 Application : Etude des indicateurs de développement des pays : Algérie, Tunisie, Maroc	46
3.2.1 Etude des indicateurs de développement du Maroc	47

3.2.2	Cas de Tunisie	55
3.2.3	Etude comparative	58
3.3	Classement des pays africains en fonction de leurs développement économique	59
	Annexe	68
	Bibliographie	75

TABLE DES FIGURES

1.1	Cercle de corrélation(variable)	10
1.2	Cercle de corrélation (variables)	11
1.3	Projection des variables	12
1.4	Figure des valeurs propres	13
2.1	Représentation des groupes	18
3.1	Représentation des valeurs propres(Algérie)	49
3.2	Graphe des individus et cercle de corrélation (Maroc)	50
3.3	Représentation des valeurs propres (Maroc)	53
3.4	Graphes des corrélations et des individus (Maroc)	54
3.5	Représentation des valeurs propres (Tunisie)	56
3.6	Cercle de corrélation et graphe des individus (Tunisie)	57
3.7	Représentation de la corrélation entre les variables	61
3.8	La courbe ROC	64

LISTE DES TABLEAUX

3.1	Tableau des valeurs propres(Algérie)	48
3.2	Matrice de corrélation(Algérie)	48
3.3	Tableau des valeurs propres (Maroc)	52
3.4	Tableau de corrélation des variables (Maroc)	52
3.5	Tableau des valeurs propres(Tunisie)	55
3.6	Tableau des corrélations(Tunisie)	57
3.7	Tableau de corrélation	60
3.8	Tableau de données de la Maroc	68
3.9	Données après estiamtion des valeurs manquantes Maroc	69
3.10	Tableau de données de la Tunisie	70
3.11	Données après estiamtion des valeurs manquantes Tunisie	71
3.12	Tableau de données de l'Algérie	72
3.13	Données après estiamtion des valeurs manquantes de l'Agérie	73
3.14	Tableau de donnés des pays africaines	74

De nos jours, l'information nous parviennent de plusieurs canons et différents moyen tels que la presse, les médias lourd, les réseaux sociaux, ...etc. Souvent nous sommes noyés dans ces données.

L'excès d'information sous forme de données multidimensionnelles générées par différents domaines tels que l'industrie, marketing, médical, banques. . . et stockées dans des bases de données, est aussi un problème concret et embarrassant. Ce problème nécessite une exploration détaillée basée sur des normes scientifiques pour pouvoir les comprendre et les exploiter.

L'analyse de données recouvre un ensemble de techniques ayant pour objectif la description statistique des grands tableaux. Ces techniques permettent de mettre en évidence les structures cachées dans les données telles que les corrélations, les dépendances, les causalités ..., etc. Parmi ces techniques, nous citons l'analyse en composante principale pour un tableau de données de variables quantitatives, l'analyse factorielle des correspondances pour les tableaux de contingence, l'analyse factorielle multiple pour les variables qualitatives, l'analyse discriminante pour la prise d'une répartition des individus en groupes ...etc. Cet excès d'information existe aussi dans les données fournies par la Banque Mondiale où les bases de données disponibles dans son site donne pratiquement tous les indices économiques et financier. Cet excédent d'information entraine souvent de la confusion et des difficultés dans l'interprétation. C'est pour cette raison que l'utilisation des techniques statistiques avancés et puissantes sont indispensables. Nous avons dans ce mémoire traité les données relatives aux indicateurs du développement des trois pays : Algérie, Maroc et Tunisie ainsi que le classement d'une quarantaine de pays africains par rapport à la caractéristique du développement. Pour ce faire, nous nous sommes basé sur le plan suivant : Le premier chapitre est une présentation des diverses techniques d'analyse multidimensionnelles en mettant l'accent sur l'analyse en composante principale. Le second aborde l'analyse discriminante en développant la démarche statistique et surtout la fonction Score qui sert d'outil de discrimination. Quant au troisième et dernier chapitre, nous mettons en application les biens fondés théorique des deux chapitres précédents afin de traiter les indicateurs du développement des pays Algérie, Maroc et Tunisie. Aussi, nous avons établi une fonction Score afin de répertorier le mieux possible les pays africains en deux catégories. Notre travail sera achevé par une conclusion contenant des perspectives futures.

L'analyse multidimensionnelle étudie des situations où un ensemble de variables doit être étudié simultanément sur un ensemble d'observations statistiques. Par nature, ces données se modélisent dans un espace à plusieurs dimensions.

L'analyse multidimensionnelle regroupe un ensemble de méthodes qui permettent de visualiser, décrire les n observations à p variables quantitatives ou qualitatives afin de dégager des liens entre les variables et les ressemblances entre individus.

Dans ce chapitre, nous allons présenter une méthode de l'analyse multidimensionnelle :
L'analyse en composante principale (ACP).

1.1 Analyse des composantes principales

L'analyse en composantes principales est la plus ancienne des méthodes d'analyse de données, elle a été mise au point par Hotelling (1933).

C'est une méthode descriptive qui consiste à représenter sous forme graphique plus d'information possible contenu dans un tableau rectangulaire de données comportant les valeurs de p variables quantitatives pour n unités.

Elle permet ainsi de visualiser un espace de p dimension à l'aide de dimension plus petit et d'explorer la liaison entre les variables et les ressemblances entre individus.

1.1.1 Principe de L'ACP

On cherche une représentation de n individus, dans un sous-espace F_k de \mathbb{R}^p de dimension, autrement dit on cherche à définir K nouvelles variables combinaison linéaire des p variables initial qui feront perdre le moins d'information possible.

Ces variables seront appelées "**composantes principales**".

Les axes qu'elles déterminent sont "**les axes principaux**".

Les formes linéaires associées sont "**les facteurs principaux**".

F_k est le sous-espace tel que le nuage projeté ait une inertie (dispersion) maximale.

F_k devra être " ajusté " le mieux possible au nuage des individus : La somme des carrés des distances des individus à F_k doit être minimale.

1.1.2 Tableaux de données et leurs caractéristiques

Les données sont les mesures effectuées sur n individus ou unités $\{e_1, e_2, e_3, \dots, e_n\}$.

Les p variables quantitatives qui représentent ces mesures sont $\{x^1, x^2, \dots, x^p\}$.

On possède ainsi un tableau rectangulaire de mesure noté X dont les colonnes sont les variables et les lignes représentent les individus statistiques

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1j} & \cdot & \cdot & \cdot & x_{1p} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2j} & \cdot & \cdot & \cdot & x_{2p} \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ x_{i1} & x_{i2} & \cdot & \cdot & \cdot & x_{ij} & \cdot & \cdot & \cdot & x_{ip} \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ \cdot & & & & & & & & & \cdot \\ x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{nj} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}$$

X : Tableau des données.

X_i : La i ème observation du tableau.

X^j : La j ème variable du tableau.

X_{ij} : Valeur de la i ème unité pour la j ème variable.

La matrice des poids

Lorsque les données ont été recueillies d'une manière aléatoire et équiprobables

$$p_i = \frac{1}{n}; \forall i = \overline{1, 2, \dots, n}.$$

Par contre, lorsque il n'ya pas équiprobabilités, les poids P_i sont tels que : $\sum_{i=1}^n p_i = 1$

On note "D" la matrice diagonale où sont regroupés les poids P .

$$D = \begin{pmatrix} p_1 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & p_n \end{pmatrix} = \frac{1}{n} * I$$

Où I est la matrice identité dans le cas $p_i = \frac{1}{n}$.

Centre de gravité

Le vecteur g des moyennes arithmétiques de chaque variables $g' = (\overline{x^1}, \overline{x^2}, \dots, \overline{x^p})$ définit le point moyen.

On a $g = X'Dh = \sum_{i=1}^n p_i e_i e_i'$ où h désigne le vecteur de \mathbb{R}^n dont toutes les composantes sont égales à 1.

Le tableau Y tel que $y_i^j = x_i^j - \overline{x^j}$ est le tableau centré associé à X .

On a :

$$Y = X - hg' = (I - hh'D)X. \quad (1.1)$$

Matrice de variance-covariance et de corrélation

La formule de la matrice de variance-covariance avec des poids égaux à $\frac{1}{n}$ se généralise comme suit :

$$V = X'DX - gg' = Y'DY. \quad (1.2)$$

On a également :

$$X'DX = \sum_{i=1}^n p_i e_i e_i'. \quad (1.3)$$

On note $D_{1/s}$ la matrice diagonale des inverses des écarts-types :

$$D_{1/s} = \begin{pmatrix} 1/s_1 & & & 0 \\ & \cdot & & \\ & & \cdot & \\ & & & \cdot \\ 0 & & & & 1/s_p \end{pmatrix}$$

et D_{1/s^2} la matrice diagonale des inverses des variances, le tableau des données centrées et réduites Z tel que :

$$z_i^j = \frac{x_i^j - \overline{x^j}}{s_j}.$$

est donc :

$$Z = YD_{1/s}. \quad (1.4)$$

La matrice regroupant tous les coefficients de corrélation linéaire entre les p variables prises deux à deux est notée R :

$$R = \begin{pmatrix} 1 & \cdot & r_{1p} & \cdot & \cdot & \cdot & r_{1p} \\ \cdot & 1 & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ r_{p1} & & & & & & 1 \end{pmatrix}$$

$$R = D_{1/s} V D_{1/s} = Z' D Z.$$

R est la matrice de variance-covariance des données centrées et réduites et résume la structure des dépendances linéaires entre les p variables.

L'espace des individus

Chaque individu étant un point défini par p coordonnées est considéré comme un élément d'un espace vectoriel F appelé l'espace des individus. L'ensemble des n individus est un " nuage " de points dans F et g en est le centre de gravité. L'espace F est muni d'une structure euclidienne afin de pouvoir définir des distances entre individus.

Le rôle de la métrique

Pour faire une représentation géométrique, il faut choisir une distance entre deux points de l'espace

La distance utilisée par l'analyse des composantes principales dans un espace où sont représentées les unités (individus) est la distance euclidienne classique. La distance entre deux unités e_i et e_j est défini par :

$$d^2(e_i; e_j) = (e_i - e_j)' M (e_i - e_j).$$

où M est une matrice symétrique de taille p définie positive.

Dans le cas d'une ACP centrée réduite, les métriques les plus utilisées sont les métriques diagonales des inverses des variances $M = D_{1/s^2}$ car elles rendent les distances entre les individus invariantes par transformation linéaire séparée de chaque variable et de s'affranchir des unités de mesure ce qui est particulièrement intéressant lorsque les variables sont hétérogènes .

$$M = D_{1/s^2} = \begin{pmatrix} 1/s^2 & . & . & . & . & 0 \\ . & 1/s^2 & & & & . \\ 0 & & . & & & . \\ . & & & . & & . \\ . & & & & . & . \\ 0 & 0 & . & . & . & 1/s^2 \end{pmatrix}$$

L'inertie

L'inertie est la somme pondérée des carrés des distances des individus au centre de gravité g .

$$I_g = \sum_{i=1}^p d^2(e_i; g), \quad (1.5)$$

tel que : $p_i = 1$.

L'inertie est donc aussi égale à la somme des variances des variables étudiées.

$$I_g = \text{tr}(MV). \quad (1.6)$$

Remarque 1.1 Dans le cas où les variables sont centrées réduites, la variance de chaque variable vaut 1. L'inertie totale est alors égale à p (nombre de variable).

L'espace des variables

Chaque variable x^j est en fait une liste de n valeurs numériques : On la considère comme un vecteur x^j d'un espace E à n dimension appelé espace des variables.

Pour étudier la proximité des variables entre elles il faut munir cet espace d'une métrique, c'est-à-dire trouver une matrice d'ordre n définie positive symétrique. Ici il n'y a pas d'hésitation comme pour l'espace des individus et le choix se porte sur la matrice diagonale des poids D pour les raisons suivantes :

- Le produit scalaire de deux variables x^j et x^k qui vaut $x^j D x^k = \sum_{i=1}^n p_i x_i^k x_i^j$ n'est autre que la covariance s_{jk} si les deux variables sont centrées.
- La norme d'une variable $\|x^j\|_D^2 = s_j^2$; en d'autres termes "la longueur" d'une variable est égale à son écart-type.

L'angle θ_{jk} entre deux variables centrées est donné par :

$$\cos \theta_{jk} = \frac{\langle x^j, x^k \rangle}{\|x^j\| \|x^k\|} = \frac{s_{jk}}{s_j s_k}. \quad (1.7)$$

Le cosinus de l'angle entre deux variables centrées n'est autre que leur coefficient de corrélation linéaire.

Si dans l'espace des individus on s'intéresse aux distances entre points, dans l'espace des variables on s'intéressera plutôt aux angles.

1.1.3 Le problème de l'ACP

La recherche d'axes portant le maximum d'inertie équivaut à la construction de nouvelles variables (auxquelles sont associés ces axes) de variance maximale.

En d'autres termes, on effectue un changement de repère dans \mathbb{R}^p de façon à se placer dans un nouveau système de représentation où le premier axe apporte le plus possible de l'inertie totale du nuage, le deuxième axe le plus possible de l'inertie non prise en compte par le premier axe, ainsi de suite.

Les axes principaux

Nous devons chercher une droite de \mathbb{R}^p passant par le centre de gravité g , et maximisant l'inertie du nuage projeté sur cette droite.

Soit a un vecteur porté par cette droite. La matrice MVM est dite matrice d'inertie du nuage, elle définit la forme quadratique d'inertie qui à tout vecteur \vec{a} de M -normé à 1, associé l'inertie projetée sur cet axe.

$$VMa = \lambda a,$$

Où a est donc vecteur propre de matrice VM et λ sa valeur propre.

Le premier axe est celui qui va correspondre à la plus grande valeur propre.

Le deuxième axe est celui de la deuxième valeur propre.

On appelle axes principaux d'inertie les vecteurs propres de VM normé à 1, ils sont au nombre de p .

Facteurs principaux

A chaque axe principale a tel que M -normé à 1 est associé le facteur principal $u = Ma$. Puisque a est un vecteur propre de VM , on a :

$$\begin{aligned} VMa &= \lambda a \\ \Leftrightarrow MVMa &= \lambda Ma \\ \Leftrightarrow MVu &= \lambda u \end{aligned}$$

Donc les facteurs principaux u seront aussi les vecteurs propres de la matrice MV .

Composantes principales

A chaque axe est associé une variable appelée composante principale. La composante principale C_i est définie par les facteurs principaux

$$\begin{aligned} C_i &= Xu_i, \\ C_1 &= X_1^1 u_1 + \dots + X_{P_1}. \end{aligned}$$

C_1 est le vecteur renfermant les coordonnées des projecteurs des individus sur l'axe 1.

C_2 est le vecteur renfermant les coordonnées des projecteurs des individus sur l'axe 2.

Propriétés des composantes principales

$$V(C_i) = \lambda_i$$

En effet,

$$\begin{aligned}
 V(C) &= C^t DC \\
 &= U' X' D X U \\
 &= U' V U \\
 &= \lambda U' M^{-1} \\
 &= \lambda.
 \end{aligned}$$

Les C_i sont non corrélées deux à deux, car les axes associés sont orthogonaux. Les composantes principales sont elles-mêmes vecteurs propres d'une matrice de taille n :

En effet :

$$MVMu = \lambda u$$

s'écrit

$$MX'DX = \lambda u.$$

En multipliant à gauche par X et en remplaçant Xu par C on a :

$$XMX'DC = \lambda C.$$

La variance d'une composante principale est égale à l'inertie apportée par l'axe principal qu'il est associé.

première composante C^1	variance : λ_1
deuxième composante C^2	variance : λ_2
troisième composante C^3	variance : λ_3

Représentation des individus

$$C^j = \begin{pmatrix} C_1^j \\ \vdots \\ C_n^j \end{pmatrix}$$

La j ème composante principale fournit les coordonnées des n individus sur le j ème axe principal. Si

on désire une représentation plane des individus, la meilleure sera celle réalisée grâce aux deux premières composantes principales. Dans le cas où, on travaille avec un tableau centré réduit Z associé à

X , on utilisera la métrique $M = 1$, ce qui implique que la matrice de variance-covariance et les facteurs principaux sont tout simplement les vecteurs propres de la matrice de corrélation C rangé selon l'ordre décroissant des valeurs propres.

Représentation des variables

La méthode la plus naturelle pour donner une signification à la composante principale est de la relier aux variables initiales X^j , en calculant son coefficient de corrélation.

$$r(C, X^j) = r(C, Z^j) = \frac{C' DZ^j}{S_C S_{Z^j}}.$$

Comme $V(C) = \lambda$ alors :

$$r(C, X^j) = \frac{C' DZ^j}{\sqrt{\lambda}}. \quad (1.8)$$

On calcule pour chaque composante principale cette corrélation.

$$\begin{aligned} r(X^j, C) &= \frac{1}{\sqrt{\lambda}} (Z^j)' DC \\ &= \frac{1}{\sqrt{\lambda}} (Z^j)' DZU \\ &= \frac{1}{\sqrt{\lambda}} C^j U \\ &= \sqrt{\lambda} U_j. \end{aligned}$$

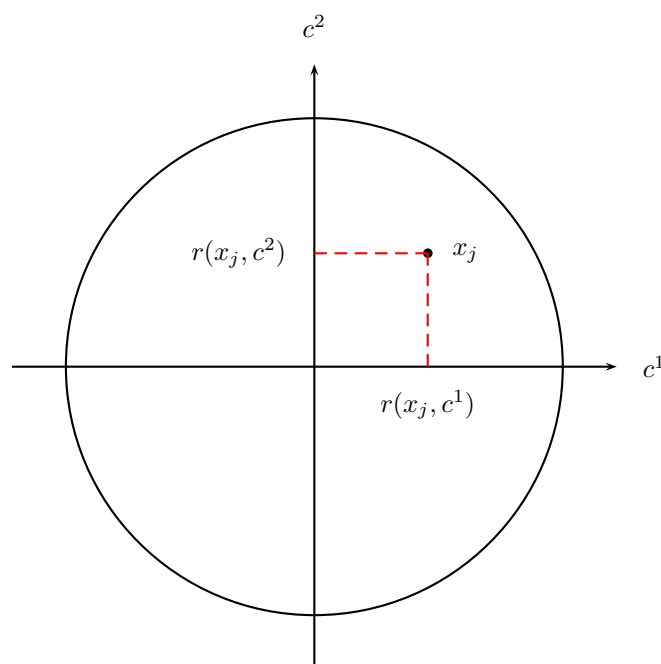


Figure 1.1 – Cercle de corrélation(variable)

$(Z^j)' DZ$ représente la jème ligne de $Z^j DZ$, donc $(Z^j)' DZU$ est la jème composante de CU .

Ces calculs s'effectuent pour chaque composante principale. Pour un couple de composantes principales c^1 et c^2 par exemple on synthétise usuellement les corrélations sur une figure appelée "cercle de corrélation" où chaque variable x^j est repérée par un point d'abscisse $r(c^1; x^j)$ et d'ordonnée $r(c^2; x^j)$.

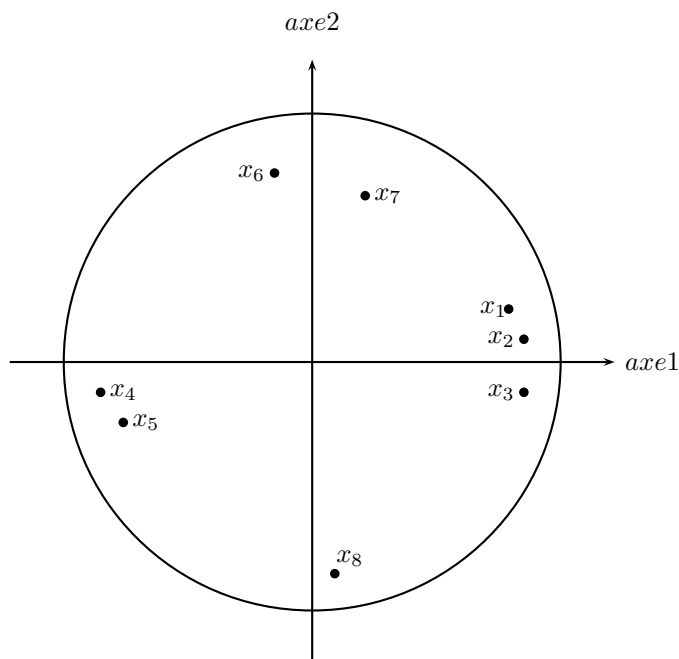


Figure 1.2 – Cercle de corrélation (variables)

Ainsi la figure 1.2 montre une première composante principale très corrélée positivement avec les variables 1 , 2 et 3, anticorrélée avec les variables 4 et 5 et non corrélée avec 6 , 7 et 8.

Par contre la deuxième composante principale oppose la variable numéro 8 aux variables 6 et 7.

On se gardera d'interpréter des proximités entre points variables, si ceux-ci ne sont pas proche de la circonférence.

Dans le cas de la métrique D_{1/s^2} c'est-à-dire, rappelons-le de l'ACP sur données centrées réduites, le cercle de corrélation n'est pas seulement une représentation symbolique commode : c'est la projection de l'ensemble des variables centrées-réduites sur le sous-espace engendré par c^1 et c^2 . En effet les Z^j étant de variance un, sont situées sur la surface de la sphère unité de l'espace des variables (isomorphe à \mathbb{R}^n).

Projetons les extrémités des vecteurs Z^j sur le sous-espace de dimension 2 engendré par c^1 et c^2 (qui sont orthogonale) les projection tombent à l'intérieur du grand cercle intersection de la sphère avec le plan $c^1 ; c^2$. La projection se faisant avec la métrique D de l'espace des variables, Z se projette sur l'axe engendré par c^1 en un point d'abscisse $\cos(z^j ; c^1)$ ce qui n'est autre que le coefficient de corrélation linéaire $r(x^j ; c^1)$.

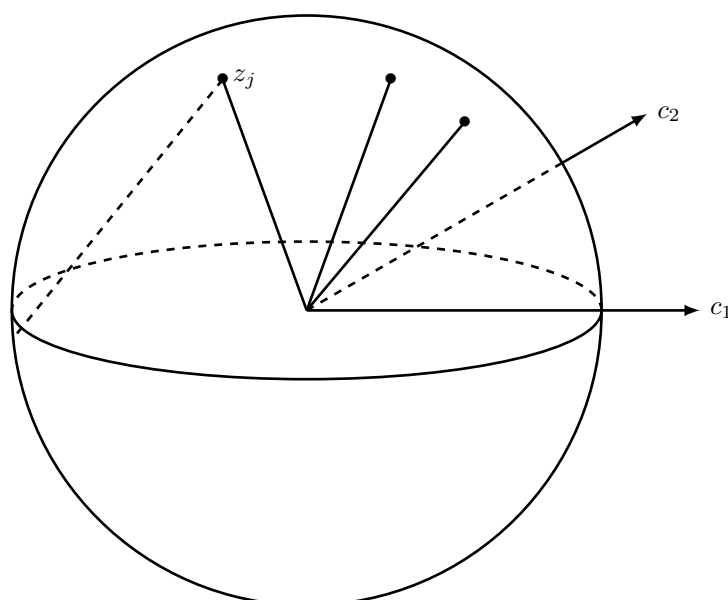


Figure 1.3 – Projection des variables

Le cercle de corrélation est donc, dans l'espace des variables, le pendant exact de la projection des individus sur le premier plan principale.

1.1.4 Interprétation et qualité de représentation d'une ACP

L'ACP construit de nouvelles variables dites artificielles, et des représentations graphiques permettant de visualiser les relations entre variables, ainsi que l'existence d'éventuelle groupes d'individus et de groupes de variables.

L'objectif est d'obtenir la représentation des individus dans un espace de dimension plus faible que p .

Nombre d'axe à retenir

Différentes procédures sont complémentaires :

Règle de KAISER : On considère que, si tous les éléments de Y sont indépendants, les composantes principales sont toutes de variances égales (égales à 1 dans le cas de l'ACP réduite). On ne conserve alors que les valeurs propres supérieures à leur moyenne car seules jugées plus "informatives" que les variables initiales ; dans le cas d'une ACP réduite, ne sont donc retenues que celles plus grandes que 1.

Part d'inertie : Diviser l'inertie totale par le nombre de variables initiales, ce qui donne l'inertie moyenne par variable notée IM .

Il faut conserver tous les axes apportant une inertie supérieure à cette valeur IM

Si les variables sont centrées et réduites, on prend l'inertie supérieure à 1.

Critère de CATTEL : Le critère de Cattell (la règle du coude) préconise de détecter sur un diagramme des valeurs propres, l'existence du coude. On doit conserver les axes associés aux valeurs propres

situées avant le coude.

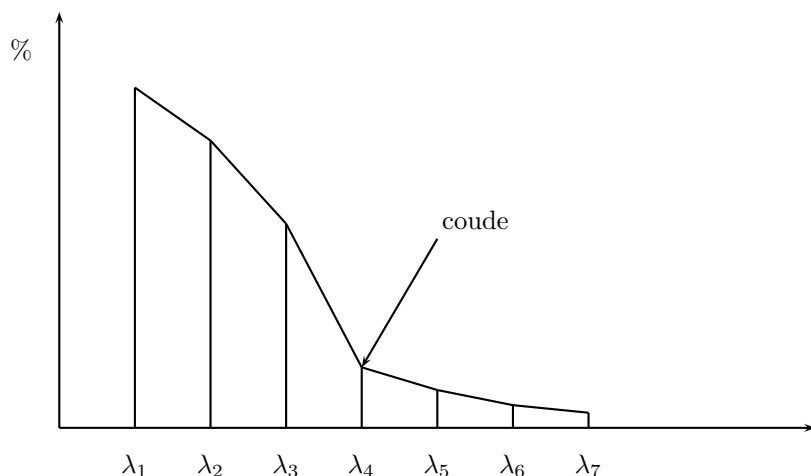


Figure 1.4 – Figure des valeurs propres

Interprétation des axes

Pour chaque axe retenu, il est nécessaire de vérifier les variables et les individus qui participent le mieux à la formation de l'axe.

La contribution de l'individu i à la composante c_k est défini par :

$$CTR_k(e_i) = \frac{p_i c_{ki}^2}{\lambda_k},$$

avec c_{ki} : est la valeur de la composante c_k pour le i ème individus.

En pratique, nous retenons pour l'interprétation les individus dont la contribution est supérieur à la contribution moyenne.

La sélection des individus suit la formule suivante :

$$|c_{ik}| > \sqrt{\lambda_k}$$

La contribution des variables j à la composante c_k est défini par :

$$CRT_k(x_j) = \frac{r^2(c^k; x^j)}{\lambda_k} = (u_j^k)^2,$$

Nous retenons les variables dont la contribution est supérieur à la contribution moyenne,

$$|u_{jk}| > \frac{1}{\sqrt{p}}.$$

Une fois les axes interpréter, nous regardons les graphiques et analyser les proximités entre les points.

La qualité de représentation de l'individu sur l'axe k :

$$QLT_k(e_i) = \cos^2(\Theta_{ik}) = \frac{c_{ik}^2}{\|e_i\|^2}$$

où $\|e_i\|^2 = \sum_{k=1}^p c_{ik}^2$

Lorsque l'angle est proche de 0, le cosinus est proche de 1, donc ce cas, l'individu est bien représenté. Dans le cas inverse, l'individu est mal représenté.

La qualité de représentation de la variable j sur l'axe k :

$$qlt_k(x_j) = \cos^2(\Theta_{kj}) = \frac{c_{jk}^2}{\|x_j\|^2}$$

Une variable est d'autant mieux représentée sur un axe qu'elle est proche du bord du cercle des corrélations de l'axe, d'autant plus mal représentée qu'elle est proche de l'origine.

Algorithme du L'ACP

1. Calculer les moyennes des variables \bar{X}^j .
2. Centrer le tableau X (réduire si les données sont hétérogènes).
3. Calculer la matrice de variance-covariance V (ou matrice des corrélations).
4. Calculer les valeurs propres et les vecteurs propres de V (ou matrice des corrélations).
5. Calculer les projections des individus et des variables sur les axes factoriels.
6. Représenter graphiquement les individus et les variables.
7. Interpréter les résultats de l'analyse.

1.1.5 Avantage et inconvénient de l'ACP

Avantage

- Méthode particulièrement bien adaptée pour traiter de grandes bases de données ;
- Synthèse de l'information contenue dans des variables ayant des unités de mesure variées et détection des variables fortement redondantes ;
- Aide à la décision quand au nombre d'indicateurs pertinents à retenir par la sélection des indicateurs les plus significativement corrélés ;
- Capacité de générer des paramètres (saturation, contribution) représentatifs du poids de chaque indicateur ;

Inconvénient

- L'ACP considère qu'un indicateur donné est parfaitement fiable et qu'il est mesuré sans erreur ;
- Perte d'information (de variance) par l'emploi fréquent de la 1^{ère} composante principale uniquement ;
- En fonction du mode de sélection des indicateurs, mise à l'écart, de l'information contenue dans les variables ne contribuant pas à la construction de la composante principale .

CHAPITRE 2

ANALYSE DISCRIMINANTE ET FONCTION SCORE

L'analyse discriminante (A.D) est une technique statistique connue et reconnue qui vise à décrire, expliquer et prédire l'appartenance à des groupes prédéfinis (classe, modalités de la variable à prédire...) d'un ensemble d'observation (individus, exemples...) à partir d'une série de variables prédictives.

On distingue classiquement deux aspects en analyse discriminante :

Analyse discriminante descriptive : consiste à chercher quelles sont les combinaisons linéaires de variables qui permettent de séparer le mieux possible les k catégories et donner une représentation graphique qui rende compte au mieux de cette séparation.

Analyse discriminante décisionnelle : un nouvel individu se présente pour lequel on connaît les valeurs des prédicteurs, il s'agit alors de décider dans quelle catégorie il faut l'affecter. C'est un problème de classement.

L'analyse discriminante est utilisée dans de nombreux domaines :

En biologie : Lorsque on veut affecter un objet à sa famille d'appartenance à partir de ses caractéristiques physiques, les iris de sir de RONALD FISHER qui est à l'origine de cette méthode en est un exemple très fameux, il s'agit de reconnaître le type d'iris (setosa, virginica, et versicolor) à partir de la longueur et la largeur de ses pétales et sépales.

En informatique : Pour la reconnaissance optique de caractère, l'analyse discriminante est utilisée pour reconnaître un caractère imprimé à partir d'images simples comme la présence ou l'absence de symétrie, le nombre d'extrémités, ...

Dans le domaine bancaire : Lorsque l'on veut évaluer la stabilité d'un demandeur de crédit à partir de ses revenus, du nombre de personnes à charge, des encours de crédit qu'il détient.

En médecine : par exemple pour détecter les groupes à hauts risques cardiaques à partir de caractéristiques telles que l'alimentation, le fait de fumer ou pas les antécédents familiaux etc.

2.1 Analyse discriminante descriptive

L'analyse discriminante descriptive est une technique de statistique exploratoire qui travaille sur un ensemble de n observations décrite par p variables, réparties en k groupes.

Elle vise à produire un nouveau système de représentation, constitué de combinaisons linéaires des

variables initiales, qui permet de séparer au mieux les k catégories.

L'analyse discriminante descriptive est une technique descriptive car elle propose une représentation graphique qui permet de visualiser les proximités entre les observations, appartenant au même groupe ou non.

C'est aussi une technique explicative car nous avons la possibilité d'interpréter les axes factoriels combinaisons linéaires des variables initiales, et ainsi comprendre les caractéristiques qui distinguent les différents groupes.

2.1.1 Variance intraclasse et interclasse

Définition 2.1 (Variance totale) : (V) Variance calculée sur toute les observations.

Variance interclasse : (B) variance des moyenne.

Variance intraclasse : (W) moyenne des variances.

$$V = B + W \quad (2.1)$$

Les calculs doivent être effectuées en pondérant les moyennes et variance des variables de chaque groupe. Les n individus \mathbf{x}_i de l'échantillon constituent un nuage E , de \mathbb{R}^p partagé en K sous-nuages $E_1; E_2; \dots; E_K$ de centres de gravité $g_1; g_2; \dots; g_K$ de matrices de variance $V_1; V_2; \dots; V_K$

Point moyen du groupe :

$$g_k = \frac{1}{n_k} \sum_{i \in I_k} X_i. \quad (2.2)$$

Donc la variance de chaque groupe est :

$$V_k = \frac{1}{n_k} \sum_{i \in I_k} (\mathbf{x}_i - g_k)(\mathbf{x}_i - g_k)', \quad (2.3)$$

point moyen général :

$$g = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i. \quad (2.4)$$

Donc la **variance totale** est :

$$V = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - g)(\mathbf{x}_i - g)', \quad (2.5)$$

Variance intra-groupe

$$W = \sum_{r=1 \dots k} \frac{n_k}{n} V_k, \quad (2.6)$$

Où I_k est ensemble des indices des individus appartenant au groupe $G_k (k = 1 \dots k)$.

Point moyen de l'ensemble des points moyens :

$$g = \sum_{k=1 \dots K} \frac{n_k}{n} g_k. \quad (2.7)$$

Variance inter-groupe

$$B = \sum_{k=1 \dots K} \frac{n_k}{n} (g_k - g)(g_k - g)' \quad (2.8)$$

Ces différences matrices sont liées par l'équation : $V = B + W$

Qui se démontre et constitue une généralisation de relation classique :

variance totale = moyenne des variances + variance des moyennes.

Nous supposerons désormais que $g = 0$, c'est-à-dire que les variables explicatives sont centrées.

Si l'on considère que le tableau de données à étudier se met sous la forme :

$$\begin{array}{c} 1 \\ 2 \\ \vdots \\ n \end{array} \begin{array}{c} 1 \ 2 \ \dots \ K \\ \left[\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & & & & \\ & & & & & & & \\ & & & A & & & & \\ 0 & 0 & \dots & 1 & & & & \end{array} \right] \end{array} \begin{array}{c} 1 \ 2 \ \dots \ P \\ X \end{array},$$

Où X est la matrice des P variables explicatives et A le tableau disjonctif associé à la variable qualitative, les K centres de gravité $g_1; g_2; \dots; g_k$ sont les lignes de la matrice $(A'DA)^{-1}(A'DX)$.

$A'DA$ est la matrice diagonale des poids q_j des sous-nuages est :

$$\begin{pmatrix} q_1 & & 0 \\ & q_2 & \\ & & \cdot \\ 0 & & & q_k \end{pmatrix}.$$

La matrice de variance interclasse s'écrit alors, si $g = 0$

$$B = ((A'DA)^{-1}A'DA((A'DA)^{-1}A'DX)) = X'DA(A'DA)^{-1}A'DX = (X'DAD)^{-1}(A'DX).$$

Dans le cas où $p_i = \frac{1}{n}$ les expressions précédentes se simplifient et en introduisant les effectifs $n_1; n_2; \dots; n_k$ des K sous-nuage, on obtient :

$$B = \frac{1}{n} \sum_j n_j g_j g_j'; g_j = \frac{1}{n} \sum_{E_j} x_i; W = \frac{1}{n} \sum_j n_j V_j.$$

2.1.2 L'analyse factorielle discriminante (AFD)

L'analyse factorielle discriminante (AFD) est essentiellement descriptive. Son objectif est de chercher quelles sont les combinaisons linéaires des variables quantitatives qui permettent de séparer le mieux possible les k catégories, et de donner une représentation graphique qui rende au mieux compte de cette opération. Cette visualisation sur les plans factoriels appropriés permet de décrire les liaisons entre la variable à expliquer et les p variables explicatives.

Exemple 2.1 Soit 100 élèves (observation) décrits par 3 variables (Math, histoire, anglais) ont été classés en 3 groupes selon leurs moyennes (Faible, moyenne, fort).

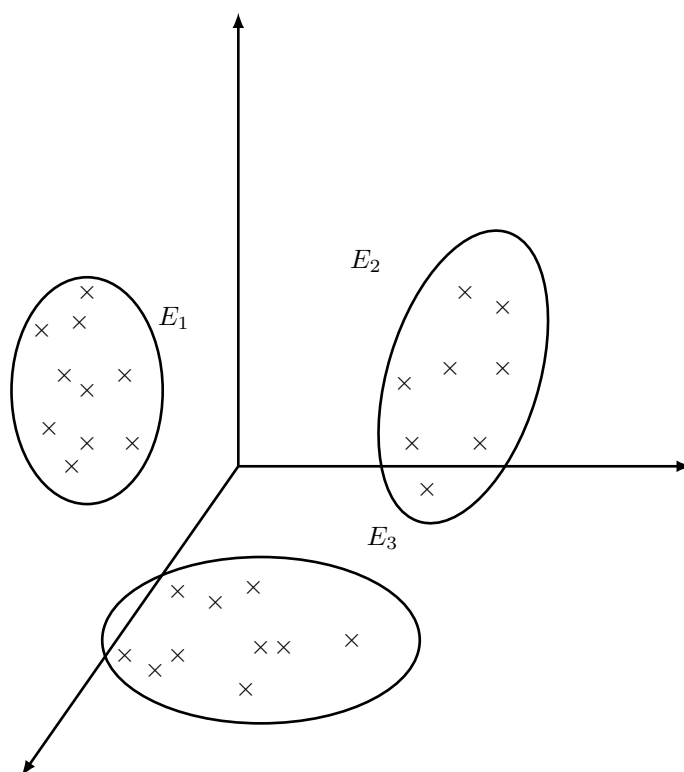


Figure 2.1 – Représentation des groupes

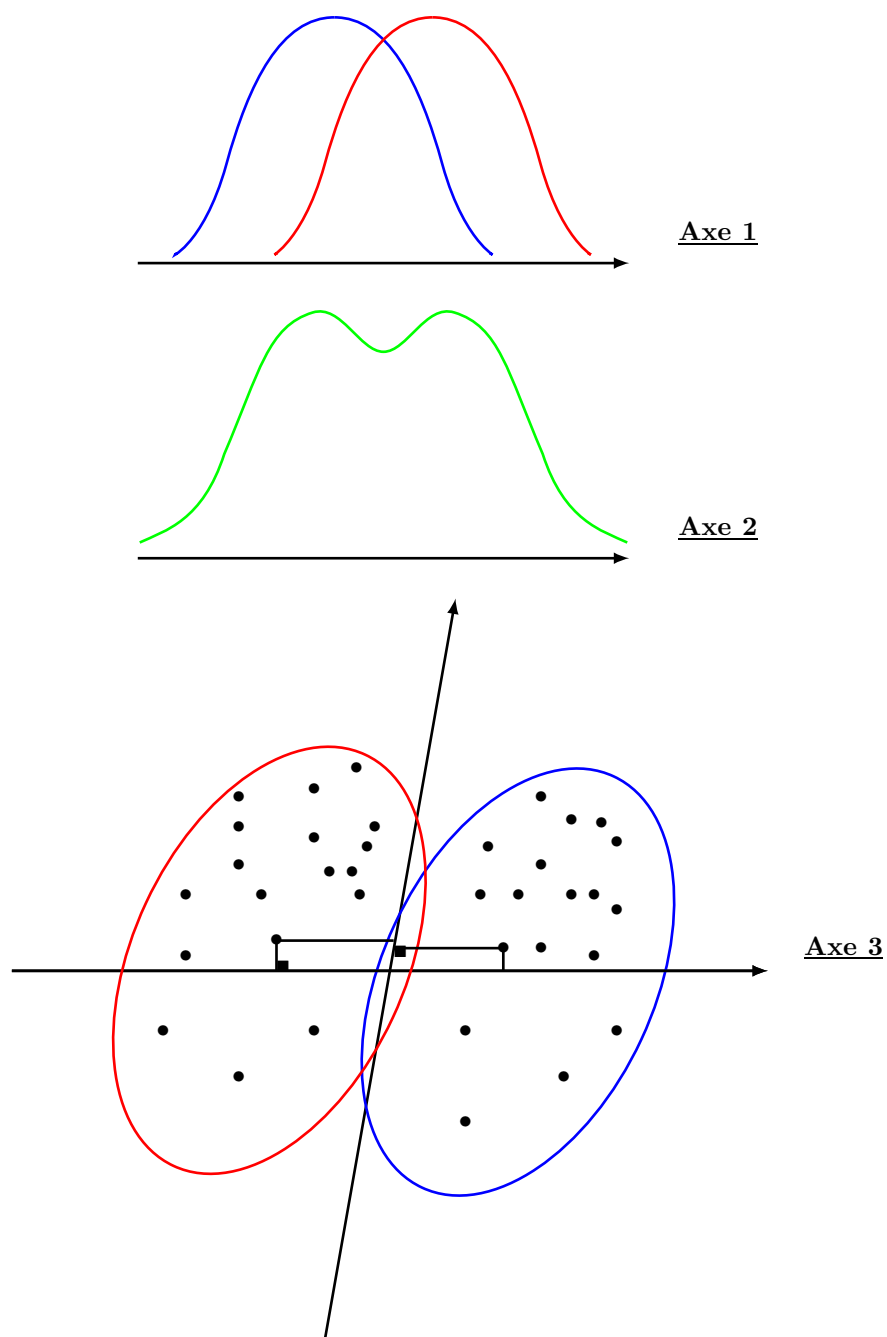
L'objectif de l'AFD

L'analyse factorielle discriminante consiste à trouver une nouvelle variable, combinaison linéaire des variables explicatives, qui "discrimine" au mieux les groupes définis par les modalités de la variable à expliquer. Cette variable notée s est définie ici comme un vecteur de \mathbb{R}^n , combinaison linéaire des vecteurs $x_1; \dots; x_p$ (colonnes de X) :

$$s = Xu = u_1x^1 + \dots + u_px^p,$$

où $u = (u_1, \dots, u_p)' \in \mathbb{R}^p$ est le vecteur des coefficients de cette combinaison linéaire. Il va donc falloir définir :

- comment mesurer pour "discrimine" bien,
- comment trouver u pour que $s = Xu$ "discrimine" au mieux.



L'axe1 possède un bon pouvoir discriminant.

L'axe2 ne permet pas de séparer en projection les deux groupes.

En AFD, on muni \mathbb{R}^p d'une métrique M et on projette les n points de \mathbb{R}^p $x_1; \dots; x_n$ (les n lignes de X) sur un axe Δ de vecteur directeur \mathbf{a} . On effectue des projections M -orthogonales et \mathbf{a} est M -normé à 1. La liste des coordonnées $s_i = x_i' M \mathbf{a}$ des individus sur Δ forme la nouvelle variable s et on a donc :

$$s = (s_1, \dots, s_n)' = X M \mathbf{a} = X u.$$

Le critère à optimiser

On définit maintenant le critère qui mesure la capacité d'un axe à discriminer les groupes. Un axe sera discriminant si les groupes sont bien séparés en projection. Pour cela, on utilise la variance intra-groupe et la variance inter-groupe de la variable discriminante s constituée des coordonnées des projections des individus sur l'axe. On a les définitions suivantes :

- La variance de s est définie par :

$$VAR(s) = \frac{1}{n} \sum_{i=1}^n (s_i - \bar{s})^2 = u'Vu, \quad (2.9)$$

où \bar{s} est la moyenne de s .

- La variance intra-groupe de s est définie par :

$$Intra(s) = \sum_{k=1}^k \frac{n_k}{n} \sum_{i \in E_k} \frac{1}{n_k} (s_i - s_k)^2 = u'Wu, \quad (2.10)$$

où \bar{s}_k est la moyenne de s dans le groupe k .

- La variance inter-groupe de s est définie par :

$$Inter(s) = \sum_{k=1}^k \frac{n_k}{n} (\bar{s} - \bar{s}_k)^2 = u'Bu. \quad (2.11)$$

On a donc la relation classique $Var(s) = Intra(s) + Inter(s)$ qui s'écrit :

$$u'Vu = u'Wu + u'Bu.$$

La quantité $u'Vu$ étant indépendante des groupes, minimiser $u'Wu$ est équivalent à maximiser $u'Bu$.

On aura une bonne discrimination des groupes si :

- Les centres de gravité projetés sont bien éloignés c'est-à-dire : $Inter(s) = u'Bu$ est maximum ;
- Les groupes projetés ne sont pas trop dispersés c'est-à-dire : $Intra(s) = u'Wu$ est minimum.

Le critère à maximiser

$$\frac{u'Bu}{u'Vu} \in [0, 1] \quad (2.12)$$

La formule (2.12) est la proportion de la variance de s expliquée par y encore appelé le rapport de corrélation entre s et y . Ce critère désigne également le pouvoir discriminant de l'axe Δ . Dans le cadre de la discrimination parfaite, ce rapport serait égal à 1. Plus généralement, la discrimination est d'autant meilleure que le ratio est proche de 1.

La solution du problème d'optimisation

La solution de ce problème d'optimisation linéaire passe par la résolution de l'équation $V^{-1}Bu = \lambda u$. La réponse nous est directement formé par le calcul des valeurs propres et vecteurs propres de la matrice $V^{-1}B$.

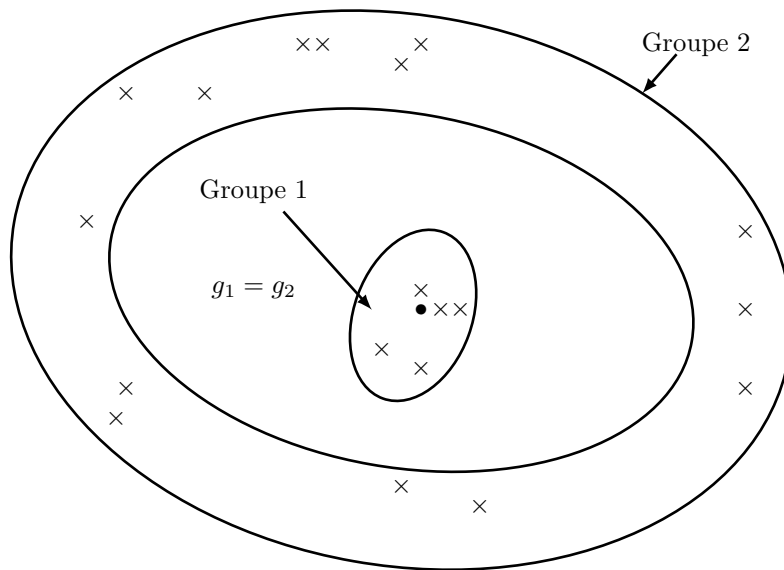
Le premier axe factoriel a_1 est donc obtenu à l'aide du vecteur propre correspondant à la plus grande

valeur propre λ .

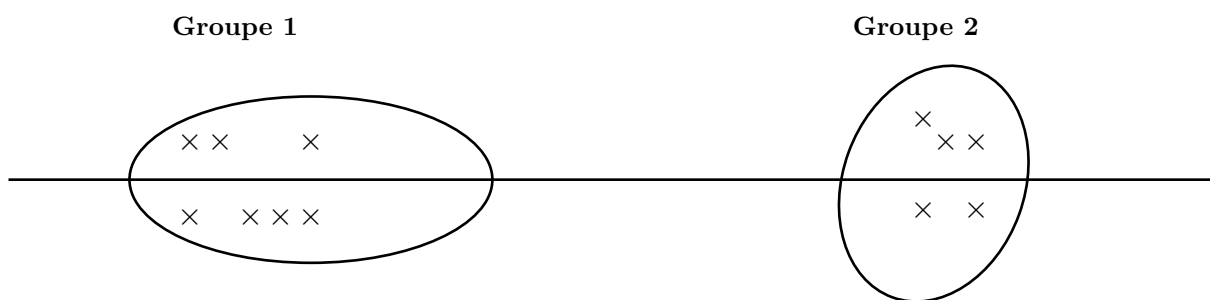
Le second axe factoriel est défini par le vecteur propre suivant

L'ensemble des axes factoriels sont déterminés par les valeurs non nulles de la matrice $V^{-1}B$ dans le cas usuel.

$\lambda = 0$ aucune séparation linéaire n'est possible, groupes concentriques Il correspond au cas où le meilleur axe ne permet pas de séparer les centres de gravités g_i .



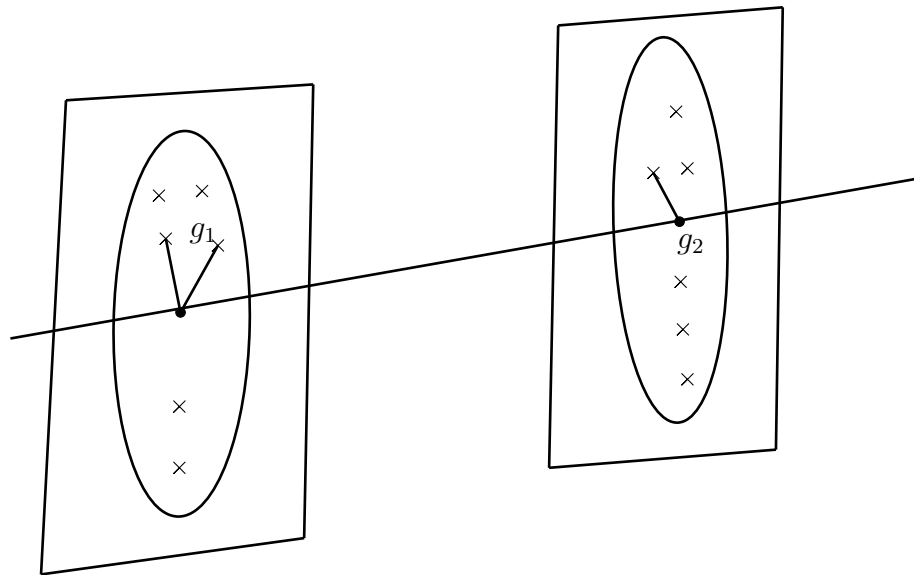
$0 \leq \lambda_1 \leq 1$: séparation possible avec groupes non recouvrants.



$\lambda_1 = 1$: La séparation est parfaite et elle correspond à deux cas :

Les k nuages sont donc chacun dans un hyperplan orthogonal à \mathbf{a}

Les centres de gravités se projettent en des points différents.



Equivalence des métriques W^{-1} et V^{-1}

Le vecteur u vérifié l'équation $Bu = \lambda Vu$ avec $u'Vu = 1$ en remplaçant V par $B + W$, cette équation devient :

$$Bu = \lambda(W + B)u \text{ avec } W^{-1}Bu = \mu u \text{ avec } \mu = \frac{\lambda}{1-\lambda}.$$

Si $0 \leq \lambda \leq 1$ on a en revanche $0 \leq \mu \leq \infty$ et $\lambda = \frac{\mu}{1+\mu}$.

L'utilisation de V^{-1} ou W^{-1} comme métrique est donc indifférent. La métrique W^{-1} est appelée métrique de Mahalanobis.

La convention usuelle dans la plupart des logiciels est d'avoir des variables discriminantes dont la variance intraclasse vaut 1.

On doit donc avoir $u'Wu = 1$, ce qui revient à $u'Bu = \frac{\lambda}{1-\lambda} = \mu$ et à $u'Vu = \frac{1}{1-\lambda}$
car $u'Bu = u'\lambda(W + B)u = \lambda u'Vu$

On pose $w = \sqrt{1+\mu}Wu$, donc $u = \frac{1}{\sqrt{1+\mu}}W^{-1}w$

En remplaçant u par sa valeur, il vient : $BW^{-1}w = \mu w$

On obtient : $w'W^{-1}w = 1$ norme de W

D'où l'équivalence :

$$BV^{-1}v = \lambda v(v'V^{-1}v = 1) \iff BW^{-1}w = \mu w(w'W^{-1}w = 1).$$

Une analyse en composante principale (ACP) particulière

D'après les équations précédentes, l'analyse factorielle discriminante n'est autre que l'ACP du nuage des centres de gravités avec la métrique V^{-1} .

On en déduit que les variables discriminantes sont non corrélées 2 à 2. S'il existe un second axe discriminant, il est possible de représenter le nuage des n observations en projection sur le plan défini par ces deux axes : Ce plan est alors celui qui permet le mieux de visualiser des observations en classe ainsi qu'en ACP, on pourra interpréter les variables discriminantes au moyen d'un cercle des corrélations.

Nous verrons plus loin que l'analyse factorielle discriminante équivaut aussi à l'ACP des g_i avec la métrique W^{-1} .

2.1.3 Cas particulier de deux groupe

En pratique, on étudie souvent le cas particulier où $K = 2$ (bon client/mauvais client, sain/malade...). On ne recherche ici qu'une seule variable discriminante puisque $K - 1 = 1$. Nous allons voir qu'il est possible de définir analytiquement l'expression de l'unique valeur propre de $V^{-1}B$ et du vecteur propre associé. Dans un premier temps, on montre que :

$$- B = \frac{n_1 n_2}{n^2} (g_1 - g_2)(g_1 - g_2)'$$

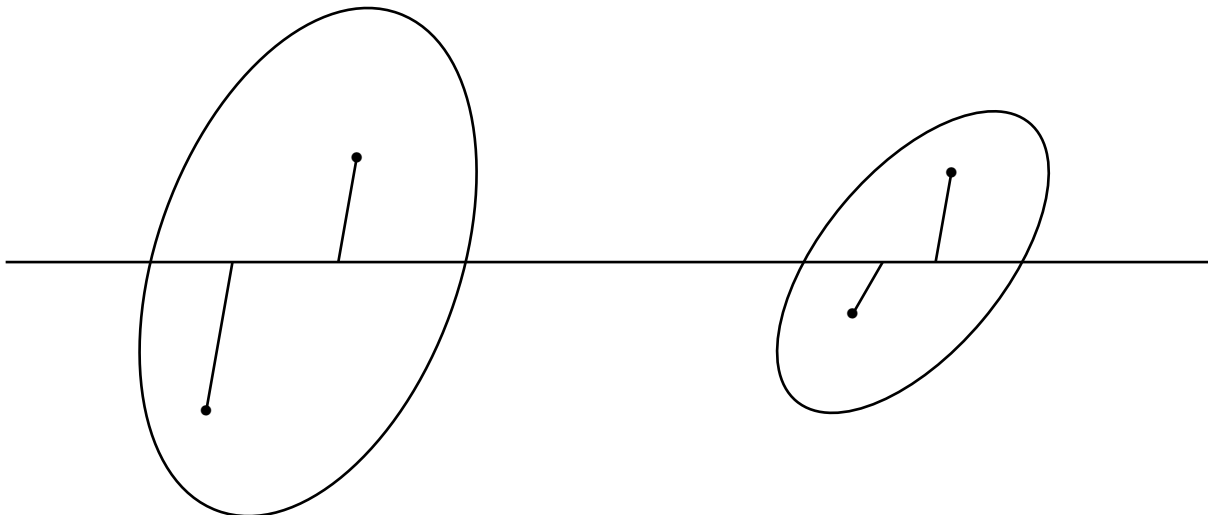
$$- V^{-1}(g_1 - g_2) \text{ est le vecteur propre de } V^{-1}B.$$

$$\lambda = \frac{n_1 n_2}{n^2} (g_1 - g_2)' V^{-1} (g_1 - g_2) \text{ est la valeur propre de } W^{-1}B.$$

On en déduit qu'avec la métrique $M = V_{-1}$ et le critère $\frac{u' B u}{u' V u} = \frac{\text{Intra}(s)}{\text{Var}(s)}$ on a :

$$\begin{cases} u = V^{-1}(g_1 - g_2), & \text{est le facteur discriminant.} \\ a = (g_1 - g_2), & \text{est l'axe discriminant.} \\ s = X V^{-1}(g_1 - g_2), & \text{est la variable discriminante.} \\ \lambda = \frac{n_1 n_2}{n^2} d_{V^{-1}}^2(g_1, g_2), & \text{est le pouvoir discriminant } \frac{\text{Intra}(s)}{\text{Var}(s)}. \end{cases}$$

où $d_{V^{-1}}^2(g_1, g_2) = (g_1 - g_2)' V^{-1} (g_1 - g_2)$ est le carré de la distance entre g_1 et g_2 .



Remarque 2.1 *il existe une infinité de vecteur propres colinéaire à u . le carré deu = $V^{-1}(g_1 - g_2)$ est :*

$$\|V^{-1}(g_1 - g_2)\|_V^2 = (g_1 - g_2)'V^{-1}(g_1 - g_2) = d_{V^{-1}}^2(g_1, g_2)$$

on en déduit q'on divisant u , a et s par $d_{V^{-1}}(g_1, g_2)$, le facteur discriminant est V -normé à 1 ($u'Vu = 1$), l'axe discriminant est V^{-1} -normé à 1 $a'Va = 1$, et la variance de la variable discriminante égale a 1 ($Var(s) = 1$)

de la même manière, avec la métrique W^{-1} et le critère $\frac{u'Bu}{u'Wu} = \frac{Intra(s)}{Var(s)}$ on a :

$$\left\{ \begin{array}{ll} u = W^{-1}(g_1 - g_2), & \text{est le facteur discriminant.} \\ a = (g_1 - g_2), & \text{est l'axe discriminant.} \\ s = XW^{-1}(g_1 - g_2), & \text{est la variable discriminante.} \\ \mu = \frac{n_1n_2}{n^2}d_{W^{-1}}(g_1, g_2), & \text{est le pouvoir discriminant } \frac{Intra(s)}{Var(s)}. \end{array} \right.$$

où $d_{W^{-1}}^2(g_1, g_2) = (g_1 - g_2)'W^{-1}(g_1 - g_2)$ est le carré de la distance entre g_1 et g_2 .

Remarque 2.2 *Il faut diviser u , a et s par $d_{W^{-1}}(g_1, g_2)$ pour que le facteur discriminant soit W -normé à 1, l'axe discriminant soit W^{-1} -normé à 1 et la variance intra-groupe de la variable discriminante soit égale à 1 ($Intra(s) = 1$).*

Finalement, dans le cas particulier de deux groupes, tout se calcule analytiquement à partir des centres de gravité g_1 et g_2 et l'axe discriminant engendre la droite reliant g_1 et g_2 . Ceci n'a rien de surprenant vu que l'on a préconisé que l'axe discriminant doit être tel qu'en projection sur Δ , les K centres de gravité doivent être aussi séparés le mieux possible. Le choix de la métrique $M = V^{-1}$ (ou $M = W^{-1}$) a une interprétation géométrique. En effet, on projette le nuage des individus sur la droite Δ (passant par g_1 et g_2) en tenant compte de l'orientation des nuages par rapport à cette droite.

La règle de Fisher (1936)

La première variable discriminante peut être utilisée comme score pour construire une règle de décision simple en fixant un seuil c . On fixe comme seuil c le milieu des scores des deux centres de gravité g_1 et g_2 sur la première (et unique....) variable discriminante (les coordonnées de leur projection sur l'axe discriminant). En d'autre termes, le seuil c est le milieu des moyennes des 2 groupes sur la variable discriminante $s = Xu$. On note \bar{s}_1 et \bar{s}_2 ces deux moyennes et on utilise ici la métrique $M = W^{-1}$

$$c = \frac{\bar{s}_1 + \bar{s}_2}{2} = \frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2).$$

On peut alors définir la règle de décision :

$s_i \geq c \Rightarrow$ **l'individu i affecté au groupe 1.**

$s_i < c \Rightarrow$ **l'individu i affecté au groupe 2.**

avec $s_i = \mathbf{x}_i W^{-1}(g_1 - g_2) - \frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2)$.

on trouve la règle de Mahalanobis-Fisher dans le cas de deux groupes.

2.2 Analyse discriminante décisionnelle

L'analyse prédictive est à la fois la plus fondamentale et la plus populaire des techniques de classification. Il s'agit dans ce cas de construire une fonction de classement (règle d'affectation . . .) qui permet

de prédire le groupe d'appartenance d'un individu à partir des valeurs prise par les variables prédictives.

2.2.1 Règles géométriques d'affectation

Nous présentons dans cette section une approche purement géométrique (sans aucune hypothèse probabiliste) de l'analyse discriminante décisionnelle. Ayant trouvé la meilleure représentation de la séparation en K groupes des n individus, on va chercher ici à affecter une nouvelle observation à l'un de ces groupes. Notons \mathbf{x} le vecteur des valeurs des p variables quantitatives mesurées sur ce nouvel individu. La règle géométrique consiste à calculer les distances de cette observation à chacun des K 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 qui consiste à prendre la métrique W^{-1} (ou V^{-1} ce qui est équivalent). La distance du nouvel individu au groupe k est alors :

$$d^2(\mathbf{x}; g_k) = (\mathbf{x} - g_k)'W^{-1}(\mathbf{x} - g_k) \quad (2.13)$$

Fonctions linéaires discriminantes

La règle géométrique classe la nouvelle observation \mathbf{x} dans le groupe k^* tel que :

$$k^* = \arg \min_{k=1\dots K} d^2(\mathbf{x}, g_k), \quad (2.14)$$

ce qui se réécrit :

$$k^* = \arg \max_{k=1\dots K} L_k(\mathbf{x}), \quad (2.15)$$

où

$$L_k(\mathbf{x}) = \mathbf{x}'W^{-1}g_k - \frac{1}{2}g_k'W^{-1}g_k. \quad (2.16)$$

$L_k(\mathbf{x})$ est la fonction linéaire discriminante du groupe k (encore appelée fonction linéaire de classement).

Preuve. W^{-1} étant symétrique on a :

$$\begin{aligned} d^2(\mathbf{x}, g_k) &= (\mathbf{x} - g_k)'W^{-1}(\mathbf{x} - g_k) \\ &= \underbrace{\mathbf{x}'W^{-1}\mathbf{x}}_{\text{indépendant de } k} + g_k'W^{-1}g_k - 2\mathbf{x}'W^{-1}g_k. \end{aligned}$$

Donc minimiser d^2 est équivalent à maximiser $\frac{2\mathbf{x}'W^{-1}g_k - g_k'W^{-1}g_k}{2} = L_k(\mathbf{x})$.

Chaque fonction linéaire discriminante définit une fonction score qui donne une "note" à l'observation \mathbf{x} dans chaque groupe. Cette observation est donc affectée au groupe pour lequel le score est le plus grand. ■

Le cas particulier de deux groupes

On peut dans ce cas particulier où $K = 2$, calculer la différence entre les fonctions linéaires discriminantes du groupe 1 et du groupe 2 pour définir une nouvelle fonction score que l'on comparera à zéro pour

affecter une nouvelle observation. On définit ainsi la nouvelle fonction linéaire discriminante (fonction score) :

$$\Delta_{1/2} = L_1(\mathbf{x}) - L_2(\mathbf{x}) = \mathbf{x}'W^{-1}(g_1 - g_2) - \frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2). \quad (2.17)$$

On retrouve alors la règle de Mahalanobis-Fisher introduite en AFD :

$\mathbf{x}'W^{-1}(g_1 - g_2) - \frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2) \geq 0 \Rightarrow$ l'individu i est affecté au groupe 1,
 $\mathbf{x}'W^{-1}(g_1 - g_2) - \frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2) < 0 \Rightarrow$ l'individu i est affecté au groupe 2.

Dans le cas particulier de 2 groupes, la règle géométrique revient donc à projeter l'observation \mathbf{x} avec la métrique W^{-1} sur le premier axe discriminant (calculé du score $\mathbf{x}'W^{-1}1(g_1 - g_2)$) et à classer \mathbf{x} selon un seuil c qui est le milieu des moyennes des groupes sur ce score ($c = \frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2)$)

De plus, $\Delta_{1/2}(\mathbf{x})$ s'écrit :

$$\Delta_{1/2}(\mathbf{x}) = \left(\mathbf{x} - \frac{g_1 + g_2}{2}\right)'W^{-1}(g_1 - g_2).$$

Donc la frontière entre les groupes 1 et 2 (aussi appelée frontière de décision) est l'hyperplan d'équation :

$$\left\langle \mathbf{x} - \frac{g_1 + g_2}{2}, g_1 - g_2 \right\rangle W^{-1} = 0$$

L'hyperplan séparateur est donc l'hyperplan W^{-1} -orthogonal à la droite reliant les points g_1 et g_2 et passant par le milieu de g_1 et g_2 .

Insuffisance des règles géométriques

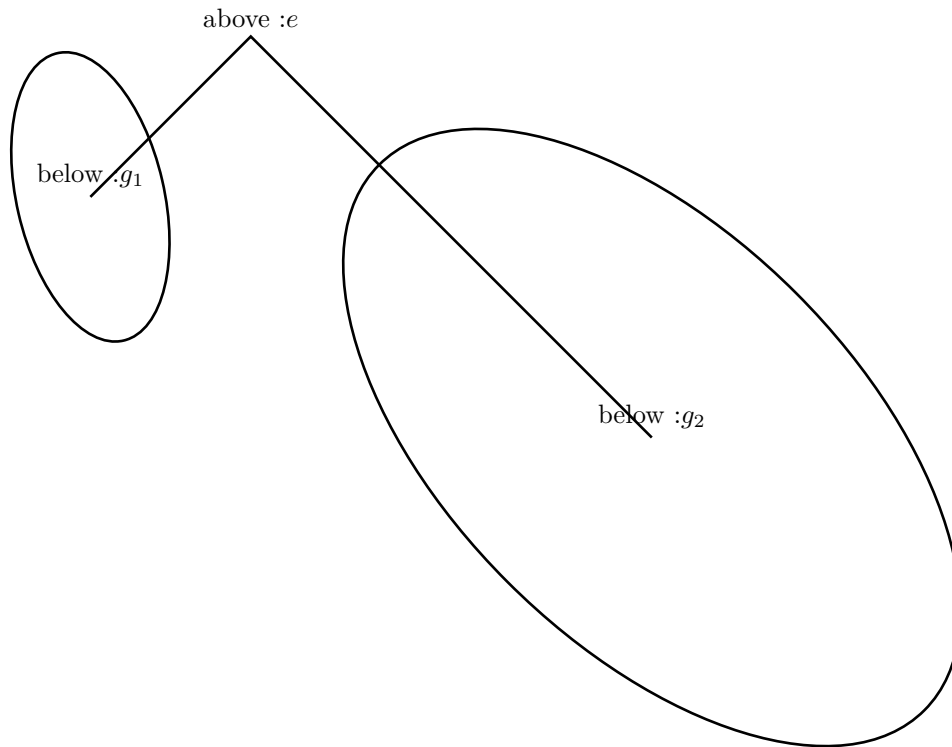
Soit un individu se présente pour lequel on connaît les valeurs des prédicteurs. Il s'agit alors de décider dans quelle catégorie il faut l'affecter. Donc l'utilisation de la règle précédente conduit à des affectations incorrectes lorsque les dispersions des groupes sont très différentes entre elles.

Soit la figure, bien que e est plus proche de g_1 que de g_2 au sens habituel il est plus naturel d'affecter \mathbf{x} à la deuxième classe qu'à la première (pouvoir d'attraction classe 2 plus pouvoir d'attraction classe 1).

Diverses solutions utilisant des métriques locales M_i telles que :

$$d^2(\mathbf{x}; g_i) = (\mathbf{x} - g_i)'M_i(\mathbf{x} - g_i),$$

ont été proposées, la plupart prenant M_i proportionnel à V_i^{-1} .



La question de l'optimalité d'une règle de décision géométrique ne peut cependant être résolue sans référence en modèle probabiliste. En effet le problème est de savoir comment cette règle se comportera de nouvelles observations ce qui impose de faire des hypothèses distributionnelles sur la répartition dans l'espace de ces nouvelles observations. On atteint donc ici les limites des méthodes descriptives. Nous verrons plus loin dans quelles conditions elles conduisent à des règles optimales.

2.2.2 Analyse discriminante probabiliste

On dispose d'un échantillon de taille n sur lequel est mesurées simultanément les p variables explicatives quantitatives et la variable qualitative à expliquer. Généralement cet échantillon est appelé échantillon d'apprentissage. Lorsque on dispose d'un nouvel individu sur lequel on mesure uniquement les p variables explicatives (et dont on ne connaît pas sa valeur de la variable qualitative), on doit décider du groupe k auquel appartient ce nouvel individu. On a alors défini une règle de décision (d'affectation) fondée seulement sur des arguments géométriques (calcul des distances de cet individu aux centres de gravité de chaque groupe et affectation au groupe "le plus proche"). On va définir des règles de décision bayésienne qui vont permettre d'affecter le nouvel individu à la classe "la plus probable". Pour cela, il va être nécessaire de faire des hypothèses probabilistes sur les données, d'où l'appellation des méthodes probabilistes.

Le cadre probabiliste

On suppose maintenant que l'échantillon d'apprentissage est issu d'une population en K groupe G_1, G_2, \dots, G_k tel que :

Y est une variable aléatoire qui prend ses valeurs dans $\{1, \dots, K\}$.

$X = (X_1, X_2, \dots, X_p)$ est un vecteur de variables aléatoires réelles.

(p_1, p_2, \dots, p_K) la distribution de Y ou $p_K = P(Y = K)$ est la proportion théorique de G_k encore appelée probabilité à priori de G_k .

$f_k : R_p \rightarrow [0, 1]$ la densité de X dans le groupe k .

la densité X est une densité de mélange :

$$X \sim \sum_{k=1}^K p_k f_k(x).$$

on supposera que l'on dispose d'un échantillon i.i.d, $(X_1, Y_1), \dots, (X_n, Y_n)$ de même que (X, Y) .

Règle du maximum à posteriori

La règle du maximum a posteriori affecte une nouvelle observation x dans le groupe k^* le plus probable sachant que :

$$k^* = \arg \max_{k=1, \dots, K} P(G_k | \mathbf{x}), \quad (2.18)$$

où $P(G_k | \mathbf{x}) = P(Y = k | X = \mathbf{x})$ est la probabilité conditionnelle appelée la probabilité à posteriori de G_k . Les probabilités à posteriori $P(G_k | \mathbf{x})$ sont parfois qualifiées de scores et on affecte donc une nouvelle observation au groupe pour lequel le score est le plus grand.

Cette règle se réécrit :

$$k^* = \arg \max_{k=1, \dots, K} p_k f_k(\mathbf{x}) \quad (2.19)$$

On parle alors souvent de règle de Bayes.

Preuve. On utilise le théorème de Bayes qui donne :

$$P(G_k | \mathbf{x}) = P(Y = k | X = \mathbf{x}) = \frac{f_k(\mathbf{x})P(Y = k)}{f_X(\mathbf{x})} = \frac{f_k(\mathbf{x})p_k}{f_X(\mathbf{x})},$$

or $f_X(\mathbf{x})$ est indépendant de k donc il suffit de maximiser $f_k(\mathbf{x})p_k$ ■

Plusieurs approches sont possibles :

Les approches paramétriques ou on peut supposer que $f_k(\mathbf{x})$ a une forme paramétrique et estimer les paramètres sur l'échantillon d'apprentissage. Par exemple, $f_k(\mathbf{x})$ est une densité $N(\mu_k, \Sigma_k)$ pour

l'analyse discriminante linéaire et l'analyse discriminante quadratique. On peut aussi supposer que la probabilité à posteriori $P(G_k|\mathbf{x})$ a une expression paramétrique et estimer directement. Par exemple $P(G_1|\mathbf{x}) = \frac{\exp(\beta_0 + \beta' \mathbf{x})}{1 + \exp(\beta_0 + \beta' \mathbf{x})}$ pour la régression logistique.

Les approches non paramétriques ou on cherche à estimer directement à partir des données les densités $f_k(\mathbf{x})$. on parle d'estimation non paramétrique lorsque le nombre de paramètres à estimer est infini. L'objectif fixé est alors une fonction que l'on supposera par exemple continue et dérivable. Cette approche très souple a donc l'avantage de ne pas nécessiter d'hypothèses particulières sur la densité $f_k(\mathbf{x})$ (seulement la régularité de $f_k(x)$ pour avoir de bonnes propriétés de convergence). En revanche, elle n'est applicable d'un point de vue pratique qu'avec des échantillons de grandes taille d'autant plus que la dimension p augmente.

Le cas gaussien

On suppose maintenant que X suit une loi normal de moyenne μ_k et l'écart type Σ_k pour chaque groupe G_k

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} (\det(\Sigma_k))^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_k)' \Sigma_k^{-1} (\mathbf{x} - \mu_k)\right), \quad (2.20)$$

avec $\mu_k \in R^p$ le vecteur moyennes théoriques et Σ_k la matrice $p * p$ des varianaces-covariances théoriques.

Dans ce cas, la règle de Bayes se réécrit :

$$k^* = \arg \min_{k=1, \dots, K} D_k^2(\mathbf{x}) \quad (2.21)$$

$$D_k^2 = (\mathbf{x} - \mu_k)' \Sigma_k^{-1} (\mathbf{x} - \mu_k) - 2 \ln(p_k) + \ln(\det(\Sigma_k)),$$

D_k^2 est appelé le carré de la distance de Mahalanobis théorique.

Preuve. Maximiser $p_k f_k(\mathbf{x})$ revient à maximiser $\ln(p_k f_k(\mathbf{x}))$ tel que :

$$\ln(p_k f_k(\mathbf{x})) = \ln(p_k) - \frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Sigma_k)) - \frac{1}{2} (\mathbf{x} - \mu_k)' \Sigma_k^{-1} (\mathbf{x} - \mu_k),$$

Donc avec $\frac{p}{2} \ln(2\pi)$ qui est indépendant de k , maximiser $\ln(p_k f_k(\mathbf{x}))$ est équivalent à minimiser D_k^2 ■
A partir de l'échantillon d'apprentissage, on veut estimer les paramètres

$$\theta = (p_1, \dots, p_k, \mu_1, \dots, \mu_k, \Sigma_1, \dots, \Sigma_k).$$

La méthode du maximum de vraisemblance peut être utilisée.

$$L(\theta) = \prod_{i=1}^n f_k(x_i) = \prod_{k=1}^K \prod_{x_i \in E_k} p_k f_k(x_i),$$

et on note que la log-vraisemblance s'écrit :

$$\ln(L(\theta)) = \sum_{k=1}^K \sum_{\mathbf{x}_i \in E_k} (\ln(p_k) - \frac{P}{2} \ln(2\pi) - \frac{1}{2} \ln(\det(\Sigma_k)) - \frac{1}{2} (\mathbf{x} - \mu_k) \Sigma_k^{-1} (\mathbf{x} - \mu_k)),$$

on obtient alors les estimateurs du maximum de vraisemblances suivant :

$$\hat{p}_k = \frac{n_k}{n}$$

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{\mathbf{x}_i \in E_k} \mathbf{x}_i$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^K \sum_{\mathbf{x}_i \in E_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)' \text{ dans le cas homoscédastique.}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{k=1}^K \sum_{\mathbf{x}_i \in E_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)' \text{ dans le cas hétéroscédastique.}$$

Les estimateurs de $\hat{\Sigma}_k$ étant biaisés, on a les estimateurs suivant :

$$\hat{\Sigma} = \frac{1}{n - K} \sum_{k=1}^K \sum_{\mathbf{x}_i \in E_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)',$$

$$\hat{\Sigma}_k = \frac{1}{n_k - 1} \sum_{\mathbf{x}_i \in E_k} (\mathbf{x} - \mu_k)(\mathbf{x} - \mu_k)',$$

2.2.3 Analyse discriminante quadratique

On se place dans le cas où $\exists k \neq k^*$ tel que $\Sigma_k \neq \Sigma_{k^*}$ appelé cas hétéroscédastique . On estime alors les paramètres sur l'échantillon d'apprentissage et en reprenant les notations :

– μ_k est estimée par $g_k = \frac{1}{n_k} \sum_{i \in E} \mathbf{x}_i$.

– Σ_k est estimée par $V_k = \frac{1}{n_k} \sum_{\mathbf{x}_i \in E_k} (\mathbf{x}_i - g_k)(\mathbf{x}_i - g_k)'$ ou encore par sa version sans biais :

$$V_k = \frac{1}{n_k - 1} \sum_{\mathbf{x}_i \in E_k} (\mathbf{x}_i - g_k)(\mathbf{x}_i - g_k)'.$$

– p_k est estimée par $\pi_k = \frac{n_k}{n}$.

avec ces estimateurs la règle de Bayes se réécrit :

$$k^* = \arg \min_{k=1, \dots, K} Q_k(X), \tag{2.22}$$

où :

$$Q_k(X) = (\mathbf{x} - g_k)' V_k^{-1} (\mathbf{x} - g_k) - 2 \ln(\pi_k) + \ln(\det(v_k))_k, \tag{2.23}$$

$Q_k(X)$ est la fonction quadratique discriminante du groupe k . Chaque fonction quadratique discriminante définit une fonction score et une nouvelle observation \mathbf{x} sera affectée au groupe pour lequel le score sera le plus petit.

2.2.4 Analyse discriminante linéaire

L'analyse discriminante linéaire fait partir des techniques d'analyse discriminante prédictive. Il s'agit d'expliquer et de prédire l'appartenance d'un individu à une classe (groupe) prédéfinie à partir de ses caractéristiques mesurées à l'aide de variable prédictive.

On se place dans le cas où $\Sigma_1 = \dots = \Sigma_K = \Sigma$ appelé cas homoscedastique. Dans ce cas la règle de Bayes se réécrit :

$$k^* = \arg \max_{k=1, \dots, K} \mathbf{x}' \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k' \Sigma^{-1} \mu_k + \ln(p_k). \quad (2.24)$$

Preuve. $\Sigma_k = \Sigma$ pour tout $k=1, \dots, K$ donc

$$\begin{aligned} D_k^2(\mathbf{x}) &= (\mathbf{x} - \mu_k)' \Sigma^{-1} (\mathbf{x} - \mu_k) - 2 \ln(p_k) + \ln(\det \Sigma) \\ &= \underbrace{\mathbf{x}' \Sigma^{-1} \mathbf{x}}_{\text{indép. de } k} - \underbrace{2 \mathbf{x}' \Sigma^{-1} \mu_k + \mu_k' \Sigma^{-1} \mu_k}_{\text{car } \Sigma^{-1} \text{ sym}} - 2 \ln(p_k) + \underbrace{\ln(\det(\Sigma))}_{\text{indép. de } k} \quad \blacksquare \end{aligned}$$

Donc minimiser $D_k^2(\mathbf{x})$ est équivalent à maximiser $-\frac{1}{2}(2 \mathbf{x}' \Sigma^{-1} \mu_k + \mu_k' \Sigma^{-1} \mu_k - 2 \ln(p_k))$

On estime alors les paramètres sur l'échantillon d'apprentissage et les paramètres qu'on a déjà estimés :

– μ_k est estimée par $g_k = \frac{1}{n_k} \sum_{i \in E_k} \mathbf{x}_i$.

– La matrice de variance-covariance Σ commune aux différents groupes est estimée par $W = \frac{1}{n} \sum_k \sum_{i \in E_k} (\mathbf{x}_i - g_k)(\mathbf{x}_i - g_k)'$ ou encore par la version sans biais :

$$W_k = \frac{1}{n - K} \sum_{\mathbf{x}_i \in E_k} (\mathbf{x}_i - g_k)(\mathbf{x}_i - g_k)';$$

– p_k est estimée par $\pi_k = \frac{n_k}{n}$.

avec ces estimateurs, la règle de Bayes se réécrit :

$$k^* = \arg \max_{k=1, \dots, K} L_k(\mathbf{x}). \quad (2.25)$$

$L(\mathbf{x})$ est la fonction linéaire discriminante du groupe k . Chaque fonction linéaire discriminante définit une fonction score et une nouvelle observation sera au groupe pour lequel le score sera le plus grand.

Remarque 2.3 On retrouve la fonction linéaire discriminant de l'analyse discriminant géométrique avec le terme $\ln(\pi_k)$ en plus. Dans le cas où l'on fait l'hypothèse d'égalité des probabilités à priori ($p_1 = \dots = p_k$), la règle de l'analyse discriminante linéaire est équivalente à la règle de l'analyse discriminante géométrique.

Estimation des probabilités à posteriori

On peut montrer que dans le cas gaussien, on a :

$$P(G_k|\mathbf{x}) = \frac{\exp(-\frac{1}{2}D_k^2(\mathbf{x}))}{\sum_{l=1}^K \exp(-\frac{1}{2}D_l^2(\mathbf{x}))}. \quad (2.26)$$

en pratique, on estime les probabilités a posteriori par :

$$\hat{P}(G_k|\mathbf{x}) = \frac{\exp(-\frac{1}{2}\hat{D}_k^2(\mathbf{x}))}{\sum_{l=1}^K \exp(-\frac{1}{2}\hat{D}_l^2(\mathbf{x}))} \quad (2.27)$$

où :

$$D_k^2 = (\mathbf{x} - \mu_k)' \hat{\Sigma}_k^{-1} (\mathbf{x} - \mu_k) + g_k(1) + g_2(\mathbf{x}) \quad (2.28)$$

telque

$g_1(\mathbf{x}) = \ln(\widehat{\Sigma}_k)$ dans le cas hétéroscédastique (quadratique) et égale à 0 dans le cas homoscedastique (linéaire)

$g_2(\mathbf{x}) = -2\ln(\hat{p}_k)$ si toutes les probabilités a priori ne sont pas égales, et égale à 0 si elles sont toutes égales (equiprobabilité).

2.2.5 Cas particulier de deux groupes

On se place dans le cadre linéaire (homoscedastique) et dans le cas où $K = 2$. On définit alors la nouvelle fonction linéaire discriminante :

$$\Delta_{\frac{1}{2}}(x) = L_1(\mathbf{x}) - L_2(\mathbf{x}) = \mathbf{x}' W^{-1} (g_1 - g_2) - \frac{1}{2} (g_1 - g_2)' W^{-1} + \ln\left(\frac{\pi_1}{\pi_2}\right). \quad (2.29)$$

On compare $\Delta_{\frac{1}{2}}(\mathbf{x})$ à 0 pour affecter un nouvel individu à l'un des deux groupes. La règle de Bayes se réécrit :

$$\mathbf{x}' W^{-1} (g_1 - g_2) - \frac{1}{2} (g_1 - g_2)' W^{-1} + \ln\left(\frac{\pi_1}{\pi_2}\right) \geq 0 \Rightarrow \text{l'individu } i \text{ sera affecté au groupe 1.}$$

$$\mathbf{x}' W^{-1} (g_1 - g_2) - \frac{1}{2} (g_1 - g_2)' W^{-1} + \ln\left(\frac{\pi_1}{\pi_2}\right) \leq 0 \Rightarrow \text{l'individu } i \text{ sera affecté au groupe 2.}$$

Dans le cas particulier de deux groupes avec hypothèse d'égalité des matrices de variances-covariances des groupes :

- la règle de Bayes revient à projeter l'observation \mathbf{x} avec la métrique W^{-1} sur le premier axe discriminant (calcul du score $\mathbf{x}' W^{-1} (g_1 - g_2)$) et à classer \mathbf{x} selon un seuil c qui est le milieu des moyennes des groupes sur ce score $\ln\left(\frac{\pi_1}{\pi_2}\right)$. cela veut dire qu'on décale le seuil vers la moyenne du groupe 2 si $\pi_1 > \pi_2$ et vers la moyenne du groupe 1 dans le cas contraire.

- on peut montrer que :

$$\hat{P}(G_1|\mathbf{x}) = \frac{\exp(\Delta_{\frac{1}{2}}(\mathbf{x}))}{1 + \exp(\Delta_{\frac{1}{2}}(\mathbf{x}))}. \quad (2.30)$$

On dit que $\hat{P}(G_1|\mathbf{x})$ est une fonction logistique du score $\Delta_{\frac{1}{2}}$ qui s'écrit aussi :

$$\Delta_{\frac{1}{2}} = \hat{\beta}_0 + \hat{\beta}' \mathbf{x}. \quad (2.31)$$

où :

$$\hat{\beta}_0 = -\frac{1}{2}(g_1 + g_2)'W^{-1}(g_1 - g_2) + \ln\left(\frac{\pi_1}{\pi_2}\right) \in R \quad (2.32)$$

et

$$\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p) = W^{-1}(g_1 - g_2). \quad (2.33)$$

2.2.6 Régression logistique

On suppose que l'échantillon d'apprentissage est issu d'une population en 2 groupes G1 et G2 et que :

- Y est une variable aléatoire binaire ($Y = 1$ ou $Y = 0$).
- $X = (X_1, \dots, X_p)$ est un vecteur de variables réelles supposées cette fois non aléatoires.

Le modèle de régression logistique est alors :

$$Y = f(\beta_0 + \beta'X) + \epsilon$$

et f est la fonction logistique $f(\mathbf{x}) = \frac{\exp(\mathbf{x})}{1+\exp(\mathbf{x})}$ de \mathbb{R} dans $[0,1]$.

En supposant que le terme d'erreur ϵ est centré, on aura $E(\epsilon) = 0$ et donc

$$E(Y/X) = \frac{\exp(\beta_0 + \beta'X)}{1 + \exp(\beta_0 + \beta'X)}.$$

Et comme

$$E(Y/X) = 1 \times P(Y = 1/X = \mathbf{x}) + 0 \times P(Y = 0/X = \mathbf{x}) = P(G_1/\mathbf{x})$$

Les paramètres β_0 et β sont estimés par maximum de vraisemblance sur un échantillon **i.i.d.** pour obtenir $\hat{\beta}_0$ et $\hat{\beta}$ et donc $\hat{P}(G_1/\mathbf{x}) = \frac{\exp(\hat{\beta}_0 + \hat{\beta}'X)}{1 + \exp(\hat{\beta}_0 + \hat{\beta}'X)}$.

En régression logistique, on utilisera généralement la règle de classement :

$$\hat{P}(G_1/\mathbf{x}) \geq 0.5 \implies \mathbf{x} \text{ est affecté au groupe 1,}$$

$$\hat{P}(G_1/\mathbf{x}) < 0.5 \implies \mathbf{x} \text{ est affecté au groupe 2.}$$

2.3 La méthode scoring

Selon le langage courant, le terme score peut signifier \acute{n} classement \acute{z} , \acute{n} résultat \acute{z} , \acute{n} marque \acute{z} etc. En statistique, c'est l'idée de \acute{n} classement \acute{z} qui est surtout retenue.

Le scoring (statistique) se présente en effet comme un ensemble de méthodes conduisant à un classement d'individus au sein de groupes préalablement définis.

La notion de classement mérite à son tour d'être élucidée compte tenu des confusions souvent constatées avec le terme de classification. Ce dernier terme signifie en effet la mise en évidence de groupements inconnus dans une population. En revanche, un classement désigne toute méthode d'affectation des individus d'une population dans des groupes définis à priori.

Formellement, étant donné un ensemble d'individus pouvant être décrits par un certain nombre de variables. Ces individus se répartissent entre quelques groupes définis à priori. Un individu se présente. On ne connaît pas son groupe d'appartenance. Peut-on, sur la base des observations qu'il présente vis-à-vis des variables considérées, prévoir le groupe auquel il appartient ?

C'est le problème auquel les méthodes de scoring cherchent à donner une solution. Une méthode de scoring se présente en effet comme une technique statistique permettant de classer un individu dans l'un des quelques groupes définis à priori et ce au vu de certaines caractéristiques de cet individu.

Il s'agit bien d'une méthode de classement statistique car elle est basée d'abord sur un traitement statistique des données issues d'un échantillon d'individus. D'autre part, les facteurs derrière l'appartenance d'un individu à tel ou tel autre groupe reste en partie inconnus.

En fait, les méthodes de scoring relèvent de méthodes générales de classement statistiques (et non de classification) comme l'analyse discriminante, la régression logistique, etc.

2.3.1 Différents types de score

Score d'appétence

Utilisé notamment en marketing, le score d'appétence est une mesure de la propension d'achat d'un client.

En pratique, on utilise notamment ce type de score pour apprécier la probabilité d'un client d'être intéressé par un nouveau produit.

En effet, dans les domaines où il est possible d'atteindre directement (par email, SMS, ou par voie postale par exemple) un grand nombre de clients potentiels (téléphonie, services bancaires, ...), la promotion de nouveaux produits gagnerait à être ciblée auprès des clients les plus intéressés.

Dans ce contexte, on est ainsi en présence de deux groupes : celui des clients qui sont intéressés par le nouveau produit et celui des clients qui ne le sont pas.

La construction d'un score d'appétence permet alors de prévoir le groupe d'appartenance des différents clients de l'entreprise au vu de leur caractéristiques et partant :

- De n'atteindre que les clients les plus réceptifs
- D'éviter d'importuner les clients non intéressés.

Ce qui assure une utilisation optimale du budget alloué à cet effet.

Score de comportement (risque)

Le score de risque ou de comportement est une mesure de la probabilité pour un client de subir un certain événement défavorable pour l'entreprise.

L'exemple typique est le crédit scoring utilisé par les banques pour apprécier les risques de non remboursement des crédits accordés à leurs clients.

Dans ce contexte, les groupes en présence sont le groupe des bons clients et celui des mauvais clients.

Une méthode de scoring se présente alors comme un précieux outil d'aide à la décision à la disposition des banquiers leur permettant lors de demandes de crédit par leur clients de détecter si ces derniers présentent ou non un grand risque de non remboursement.

Ce type de score peut être aussi utilisé par les compagnies d'assurance pour apprécier le niveau de sinistralité d'un nouveau client. L'utilisation d'un score de risque permet ainsi de réduire les impayés. Elle permet également de fournir les bases d'une tarification du risque.

Score d'octroi (ou score d'acceptation)

C'est un score de risque calculé pour un client qui est nouveau ou a une faible activité avec la banque. Le risque est calculé en temps réel, au moment où le client sollicite la banque, sur la base de données déclaratives éventuellement croisées avec des données de géomarketing.

2.3.2 Les différentes étapes de la construction d'un score

La construction d'un modèle de scoring est un cheminement logique qui se scinde en plusieurs étapes :

Le choix de la population à étudié

Pour bâtir un modèle de score, il convient de choisir la variable à expliquer Y . On parle aussi de critère à modéliser ou encore de variable cible. La variable Y sera qualitative généralement binaire à deux modalités. Chaque modalité définit un groupe.

Le choix des variables explicatives

On doit choisir un vecteur de variables explicatives $X = (X^1, \dots, X^p)$ qui sont les caractéristiques de chaque individu. Les caractéristiques peuvent être quantitatives ou qualitatives.

Le choix des données et de l'échantillon

Il faut savoir comment choisir les données et l'échantillon pour faire une bonne étude de la population. L'échantillon doit être représentatif de la population. On dispose alors d'un échantillon de taille n (nombre d'individus) sur lequel sont mesurées simultanément les p variables explicatives $X^1; \dots; X^p$ et la variable à expliquer Y .

Les différentes méthodes de construction d'un score

Comme on a déjà vu précédemment la construction d'un score est basée sur les méthodes d'analyse discriminante à savoir le cas qu'on veut étudier.

On va faire une récapitulation sur les méthodes de l'analyse discriminante :

1. L'analyse factorielle discriminante (AFD) :

- variables explicatives quantitatives.
- outil descriptif de visualisation des données.

2. L'analyse discriminante géométrique :

- variables explicatives quantitatives.
 - affectation d'un nouvel individu au groupe "le plus proche".
3. L'analyse discriminante probabiliste :
- variables explicatives quantitatives
 - Y et $X = (X^1; \dots; X_p)$ aléatoires.
 - affectation d'un nouvel individu au groupe "le plus probable".
 - analyse discriminante linéaire (ADL) ou discriminante linear analysis (LDA) : cas homoscédastique gaussien.
 - analyse quadratique discriminante (AQD) ou quadratic discriminant analysis (QDA) : cas hétéroscédastique gaussien.
4. La régression logistique :
- variables explicatives quantitatives ou qualitatives.
 - Y binaire et aléatoire et $X = (X^1; \dots; X_p)$ non aléatoire.

2.3.3 Validation d'une méthode de scoring

Une fois un modèle ou plusieurs modèles de scoring sont estimés, il convient d'analyser leurs performances avant de les valider pour être utilisés comme outil d'aide à la décision.

L'analyse de performances, à l'issue de la quelle une méthode de scoring est validée, permet notamment

- D'améliorer un modèle en comparant plusieurs de ses variantes (ajout ou retrait de variables explicatives, etc.)
- De choisir entre plusieurs types de modèles candidats

L'analyse des performances d'un modèle gagnerait à être conduite sur un jeu de données différent de celui qui a été utilisé pour l'estimation. On doit en effet, lorsque cela est possible, distinguer entre l'échantillon d'apprentissage et l'échantillon de test ou de validation. Ce dernier doit nécessairement contenir les valeurs réelles de la variable cible (appartenance aux groupes). D'une manière générale, il s'agit de comparer entre les valeurs réelles de la variable cible avec celles prédites par le modèle.

Concepte de base

Positifs et négatifs

Soit une population partitionnée en deux sous groupes G_1 et G_2 . On appelle (par convention) les positifs les individus de G_1 et les négatifs les individus de G_2 .

On dispose par ailleurs d'une fonction de score (issue d'un modèle) notée S et d'un seuil s définies tels que :

- On affecte l'individu présentant l'observation x au groupe G_1 si $S(x) > s$.
Autrement dit, on considère cet individu comme positif.
- Sinon, on l'affecte au groupe G_2 , on le considère donc comme négatif.
On appelle :

- Faux positif, un individu négatif considéré par la méthode de score comme positif;
- Faux négatif, un individu positif considéré par la méthode de score comme négatif.

Spécificité

On appelle coefficient de spécificité et on note $1 - \alpha$ la probabilité suivante :

$$1 - \alpha = P(S(\mathbf{x}) < s/\mathbf{x} \in G_2)$$

C'est donc la probabilité de bien détecter un négatif ou encore c'est la proportion des négatifs dans la population pouvant être détecté par la méthode.

La quantité $\alpha = P(S(\mathbf{x}) \geq s/\mathbf{x} \in G_2)$ désigne donc la probabilité de considérer un individu comme positif alors qu'il est négatif (faux positif). Pour une méthode de score, c'est un premier type de risque d'erreur d'affectation.

Sensibilité

On appelle coefficient de sensibilité et on note $1 - \beta$ la probabilité suivante :

$$1 - \beta = P(S(\mathbf{x}) > s/\mathbf{x} \in G_1)$$

C'est donc la probabilité de bien détecter un positif ou encore c'est la proportion des positifs dans la population pouvant être détecté par la méthode.

La quantité $\beta = P(S(\mathbf{x}) \leq s/\mathbf{x} \in G_1)$ désigne par conséquent la probabilité de considérer un individu comme négatif alors qu'il est positif (faux négatif). Il s'agit pour une méthode de score d'un deuxième type de risque d'erreur d'affectation

Remarque 2.4 – *On peut aussi considérer la quantité $\gamma = P(S(\mathbf{x}) > s)$ qui est la probabilité de considérer un individu comme positif. C'est la proportion d'individus supposés être intéressés par un nouveau produit dans une campagne marketing par exemple.*

- *Le meilleur modèle (et donc la meilleure fonction de score) est celui qui minimise les deux types de risque d'affectation (les quantités b et a).*
- *Les coefficients α et β changent lorsque le seuil s change. On les exprime comme des fonctions de s : $\alpha(s)$ et $\beta(s)$. Le seuil s est déterminé à l'extérieur du modèle notamment par des considérations d'ordre économique*

NB : En augmentant s , on réduit la probabilité de tomber sur des faux positifs (α) mais on augmente la probabilité d'avoir des faux négatifs (β). A noter aussi que le modèle le plus performant est celui pour lequel les deux distributions sont séparées. En revanche, lorsque les deux distributions sont confondues, le modèle correspondant est le moins performant.

Principaux outils de mesure de performance

Plusieurs outils de mesure de performance sont proposés dans la littérature statistique. On présente dans ce qui suit trois de ces outils qui sont les plus connus : la matrice de confusion, la courbe ROC et la courbe LIFT.

Matrice de confusion

On l'appelle aussi tableau de classement. Elle prend la forme suivante :

	Positifs	Négatifs	Total
Considérés positifs	n_{11}	n_{12}	$n_{1.}$
Considérés négatifs	n_{21}	n_{22}	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	n

Ces résultats sont issus d'un traitement de l'échantillon de validation. A partir de ce tableau, on calcule :

- Le taux d'erreur de classement donné par : $MC = (n_{12} + n_{21})/n$
- Le taux de biens classés donné par : $BC = (n_{11} + n_{22})/n$

Remarque 2.5

Ces taux doivent être interprétés en tenant compte des fluctuations d'échantillonnage

Il n'ya pas de référence servant de base pour la comparaison.

La courbe ROC

L'appellation ROC vient des abréviations du nom anglais donné à cette courbe : (Receiver operating characteristics) qui mesure la performance d'un classificateur binaire, présente un outil graphique qui permet d'évaluer et de comparer le modèle.

Définition 2.2 La courbe ROC est défini par la représentation graphique de la proportion $(1 - \beta(s))$ des positifs détectés par la méthode en fonction de la proportion des faux positifs $\alpha(s)$ lorsque s varie.

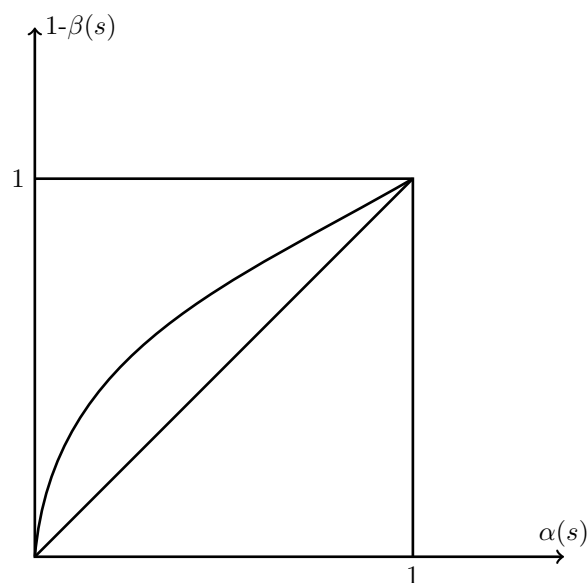
On peut noter d'après le graphique ci-dessus, que :

Lorsque $\alpha(s) = 0, 1 - \beta(s) = 0$

Lorsque $\alpha(s) = 1, 1 - \beta(s) = 1$.

$1 - \beta(s)$ et $\alpha(s)$ évoluent dans le même sens.

D'une manière générale, la courbe ROC a cette allure



En pratique, la courbe ROC étant construite à partir des données de l'échantillon de validation, les quantités $1 - \beta(s)$ et $\alpha(s)$ sont remplacées par leur estimations issues de ces données, soit respectivement la proportion des faux positifs dans l'ensemble des négatifs et la proportion des individus considérés positifs dans l'ensemble des positifs.

Utilisation

On peut noter que :

- Lorsque les deux distributions de $S(\mathbf{x})$ (dans les deux groupes) sont bien distinctes, la courbe ROC est horizontale. En effet, lorsque $\alpha(s)$ passe de 0 à 1, $1 - \beta(s)$ prend toujours la valeur 1 (Modèle le plus performant).
- Lorsque les deux distributions de $S(\mathbf{x})$ sont confondues, la courbe ROC coïncide avec la première bissectrice (Modèle le moins performant). Ces remarques conduisent à considérer la surface sous la courbe ROC (AUC) comme indicateur synthétique de la performance d'un modèle :
 - $AUC \approx 1$ modèle très performant.
 - $AUC \approx 1/2$ modèle non performant.

Cet indicateur permet ainsi de choisir entre modèles : On retient le modèle ayant le AUC le plus élevé.

La surface AUC peut être calculée en utilisant la méthode de trapèzes une fois que la courbe ROC est tracée. Mais en pratique, on utilise la méthode des paires concordantes. On démontre en effet que : $AUC = P(S_1 > S_2)$ où S_1 et S_2 sont respectivement les scores de deux individus tirés d'une manière indépendante dans le groupe des positifs puis dans le groupe des négatifs.

Dans les applications, cette probabilité est estimée par la proportion des paires concordantes. Si on note n_1 et n_2 le nombre d'individus dans le groupe des positifs et dans le groupe des négatifs. Le nombre de paires s'élève à $n_1 n_2$. Parmi ces paires, celles où le score du positif dépasse celle du négatif sont appelées paires concordantes.

La courbe LIFT

Cette courbe est très utilisée en marketing. Elle peut servir aussi à mesurer la performance d'une campagne de promotion d'un nouveau produit. Elle est alors déterminée à partir des données de la population.

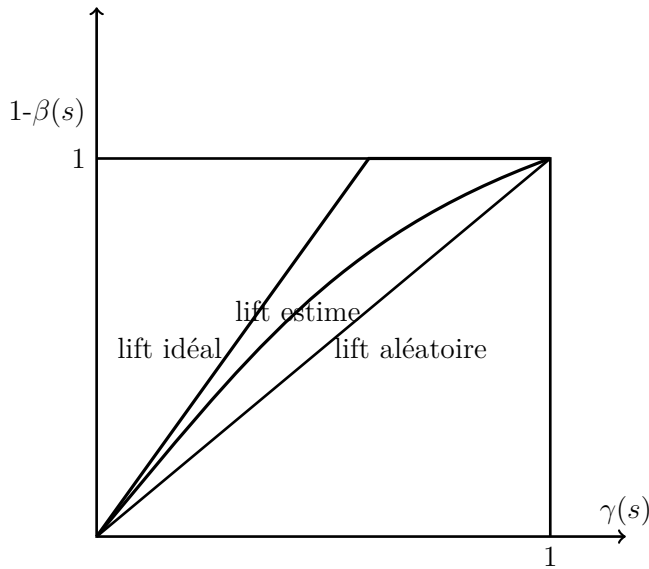
Définition 2.3 La courbe LIFT est défini par la représentation graphique de la proportion $(1 - \beta(s))$ des positifs détectés par la méthode en fonction de la proportion des individus considérés positifs par la méthode $\gamma(s)$ lorsque s varie.

On peut noter que :

- Suite à une réduction de s , $\gamma(s)$ augmente. Cette augmentation se décompose en une augmentation de vrais positifs et en une augmentation des faux positifs
- L'idéal est lorsque l'on enregistre que des augmentations des vrais positifs suite à une augmentation de $\gamma(s)$ (les distributions sont alors bien séparées). Cette situation est représentée par un LIFT idéal.

- Lorsqu'on enregistre autant de vrais positifs que de faux positifs suite à une augmentation de $\gamma(s)$, la méthode est non performante et est équivalente à un choix aléatoire. Cette situation est représentée par un LIFT coïncidant avec la première bissectrice (LIFT aléatoire).

D'une manière générale, la courbe LIFT a l'allure suivante :



Utilisation

Notons p_1 la proportion des positifs et donc $(1 - p_1)$ la proportion des négatifs. La probabilité γ s'écrit alors :

$$\gamma(s) = (1 - \beta)p_1 + \alpha(1 - p_1)$$

On peut démontrer que la surface sous la courbe LIFT notée AUL est donnée par :

$$AUP = p_1/2 + (1 - p_1)AUC.$$

Cette surface peut être utilisée comme une mesure de la performance d'un modèle. On lui préfère la quantité K_i donnée par : La division de **la surface entre lift estimé et lift aléatoire** sur **surface entre lift idéal et lift aléatoire**. En remplaçant on trouve :

$$K_i = 2AUC - 1$$

Illustration

Les données de l'échantillon de validation, une fois qu'on a calculé le score pour chaque individu, se présentent ainsi par ordre décroissant du score :

Identifiant	Groupe	Score	$\gamma(s)$	$\alpha(s)$	$1 - \beta(s)$
1	positif	1	0.1	0	0.2
2	positif	0.9	0.2	0	0.4
3	positif	0.85	0.3	0	0.6
4	négatif	0.7	0.4	0.2	0.6
5	positif	0.6	0.5	0.2	0.8
6	négatif	0.55	0.6	0.4	0.8
7	positif	0.45	0.7	0.4	1
8	négatif	0.3	0.8	0.6	1
9	négatif	0.2	0.9	0.8	1
10	négatif	0.1	1	1	1

Matrice de confusion

Pour un seuil s fixé à 0.4, la matrice de confusion se présente ainsi :

	Prévu	Positifs	Négatifs	Total
Réel				
Positifs		5	0	5
Négatifs		2	3	5
Total		7	3	10

On déduit :

$$BC = 8/10,$$

et

$$MC = 2/10.$$

La courbe ROC

La courbe ROC prend la forme suivante pour les données considérées :

On en déduit directement,
 $AUC = 0.88$

La courbe LIFT

En portant les points, on obtient :

On en déduit : $K_i = 0.76$

2.3.4 Les limites de la méthode scoring

En pratique, on pourrait utiliser d'autres méthodes plus ou moins subjectives pour apprécier la probabilité d'appartenance d'un individu à un groupe donné.

Par rapport à cette méthode, les techniques statistiques de scoring présentent un certain nombre d'avantages et inconvénients dont les principaux sont présentés ci après.

Principaux avantages

Par rapport à un système de notation subjective, le scoring statistique présente les avantages suivants :

- Evaluation quantitative de la probabilité d'appartenance ce qui permet d'ordonner les individus.
- Caractère immuable (fixe) des résultats
- Cohérence des résultats : deux individus ayant les mêmes caractéristiques auront le même score
- Caractère explicite : méthodologie d'évaluation pouvant être clairement présentée.
- Prise en compte de plusieurs facteurs de risque.
- Aptitude d'être testée au préalable
- Explication du lien existant entre le niveau de risque et les facteurs de risque
- Aptitude à donner lieu à des calculs sur les effets sur la rentabilité de l'entreprise.

Principaux inconvénients

Les méthodes statistiques de scoring souffrent néanmoins de quelques insuffisances dont entre autres :

- La décision pouvant être prise suite à l'utilisation des méthodes de scoring est basée sur une probabilité et non sur une certitude
- Les méthodes statistiques de scoring supposent comme toute autre méthode statistique que le futur est identique au passé.
- Le risque est expliqué par les seules variables disponibles
- Il existe un vrai problème de biais de sélection dans l'élaboration d'une méthode de crédit scoring. En effet, les dossiers refusés ne sont pas pris en considération.
- L'application d'un système de scoring nécessite un grand nombre de données et de variables statistiques et serait de ce fait impossible à réaliser sans l'outil informatique.
- La mise en place d'un système de scoring dans une entreprise n'est pas toujours facile à réaliser du fait de la nécessité de son intégration informatique avec les autres systèmes d'information.

CHAPITRE 3

ETUDE DES FACTEURS DU DÉVELOPPEMENT DE L'ALGÉRIE, LA TUNISIE ET LE MAROC ET CLASSEMENT DES PAYS AFRICAINS

L'objectif de ce chapitre est d'étudier les indicateurs du développement des pays : Algérie, Tunisie et Maroc. Le but est d'associer des caractéristiques qui permettent de synthétiser l'information relative au développement de ces pays. Aussi, nous nous sommes donnés l'objectif de discriminer certains pays d'Afrique en fonction de ces indicateurs. Plus précisément, nous réalisons à base de la théorie présentée aux chapitres précédents, une analyse discriminante en s'appuyant sur la fonction score afin de retenir les variables ayant un fort pouvoir discriminatoire.

Les données qui ont servi de support pour cette étude sont disponibles sur le site de la Banque Mondiale <https://donnees.banquemondiale.org/> et sont données en annexe de ce mémoire.

Pour mener à bien cette étude, nous procédons suivant le plan ci-dessous :

1. Présentation de la Banque Mondiale et du logiciel ainsi que les packages utilisés.
2. Traitement des indicateurs de développements de pays : Algérie, Tunisie et Maroc afin de visualiser un espace de p dimension à un espace réduit, et explorer les liaisons entre variables et les ressemblances entre individus en effectuant une analyse en composante principale.
3. A base de la fonction score, nous cherchons les variables ayant un pouvoir discriminatoire sur une quarantaine de pays africains.

3.1 Présentation et outils

La Banque Mondiale est une institution internationale créée le 27 décembre 1945. Elle se situe à Washington (Etas-Unis) et le président du groupe, depuis le 1^{er} juillet 2012, est Jim Youg Kim de nationalité Coréenne. Le nombre de pays membre est de 189.

La Banque Mondiale cible particulièrement les pays en voie de développement. Son principal rôle est de favoriser la croissance de ces pays en luttant contre la pauvreté.

Pour financer ses interventions, la Banque Mondiale procède à des emprunts obligataires sur les marchés financiers mondiaux. Le soutien des pays membres, qui lui versent leur quote-part tous les ans, ainsi que la solidité financière de l'institution, lui permettent de bénéficier de la meilleure notation des agences de rating et ainsi d'emprunter à des conditions financières avantageuses.

De nombreuses critiques ont été faites sur son action, notamment par l'ONU qui estime qu'elle ne prend pas en compte la sauvegarde des droits humains. Le nombre de pays pour lesquels elle est intervenue et qui ont réussi à passer du stade de pays en développement à celui de pays développés est très faible.

Les institutions qui forment le Groupe de la Banque mondiale sont spécialisées, chacune à sa manière, dans différents aspects du développement mais œuvrent toutes dans le même sens : lutter contre la pauvreté.

Les cinq institutions du groupe de la Banque mondiale.

1. **La Banque internationale pour la reconstruction et le développement (BIRD)** accorde des prêts aux gouvernements des pays à revenu intermédiaire et des États pauvres solvables.
2. **Association internationale pour le développement (IDA)** fournit des prêts à intérêt zéro, appelés crédits, aux gouvernements des pays les plus pauvres.
3. **La Société financière internationale (SFI)** octroie des prêts directement au secteur privé des pays en développement.
4. **L'agence multilatérale de garantie des investissements (MIGA)** offre des garanties contre les pertes occasionnées par des risques non commerciaux à ceux qui investissent dans les pays en développement.
5. **Le Centre international pour le règlement des différends relatifs aux investissements (CIRDI)** fournit les infrastructures internationales nécessaires à la conciliation et à l'arbitrage des conflits dans ce domaine.

Les termes "Banque mondiale" et "la Banque" ne désignent que la BIRD et l'IDA, tandis que les expressions "Groupe de la Banque mondiale" et "Groupe de la Banque" englobent les cinq institutions.

Mission du groupe de BM Les objectifs sont ambitieux : améliorer la qualité de vie et augmenter considérablement les profits économiques et l'autonomisation dans les sociétés.

Les différentes branches du Groupe de la Banque mondiale sont impliquées pratiquement dans tous les aspects du développement. Leurs activités vont de la recherche économique à l'analyse en vue d'apporter aux gouvernements et aux entreprises privées une aide financière et technique.

Grâce à ses cinq institutions, le Groupe de la Banque travaille dans plus de 100 économies en développement, y apportant idées et programmes de financement destinés à améliorer le niveau de vie et à éradiquer la pauvreté au pire sens du terme.

Au cours de ces dernières années, ce rôle a gagné en importance relative dans le monde des finances internationales, tandis que les flux financiers nets du secteur privé vers les pays en développement reculaient.

Le Groupe de la Banque est dirigé par ses pays membres (débiteurs, créanciers et donateurs), dont les représentants résident au siège du Groupe de la Banque à Washington, D.C., et dans les bureaux nationaux du monde entier.

Les pays membres

Chaque institution du Groupe de la Banque mondiale appartient à ses pays membres (ses actionnaires).

Leur nombre varie en fonction de l'institution, allant de 184 pour la Banque internationale pour la reconstruction et le développement (BIRD) à 139 pour le Centre international pour le règlement des différends relatifs aux investissements (CIRDI), selon les chiffres de mai 2003.

En pratique, les pays membres dirigent le Groupe de la Banque mondiale par le biais des Conseils des Gouverneurs et des Administrateurs. Ces organes prennent toutes les grandes décisions en matière de politique organisationnelle.

Le status de membre

Les cinq institutions du Groupe de la Banque appartiennent à leurs pays membres. Pour devenir membre de la Banque internationale pour la reconstruction et le développement (BIRD), un pays doit d'abord adhérer au Fonds monétaire international (FMI). Pour rejoindre l'Association internationale pour le développement (IDA), la Société financière internationale (SFI) et l'Agence multilatérale de garantie des investissements (MIGA), il faut faire partie de la BIRD. Les pays membres achètent des parts de l'institution et participent ainsi à la formation de son capital et de son pouvoir d'emprunt. Cette entente porte le nom de *né souscription au capital*. De plus, les pays membres signent le texte fondamental de chaque institution : les Statuts de la BIRD, de l'IDA et de la SFI, ainsi que la Convention de la MIGA. L'adhésion au Centre international pour le règlement des différends relatifs aux investissements (CIRDI) suppose la signature et la ratification de sa Convention et implique également une souscription au capital.

Les pays membres peuvent se retirer des institutions du Groupe de la Banque à n'importe quel moment, par notification. Un membre peut aussi être suspendu et même, après un an, être expulsé s'il ne remplit pas ses obligations envers une institution du Groupe. Le fait de quitter le FMI implique automatiquement le départ du Groupe de la Banque à moins que, dans les trois mois, ce dernier décide, à la majorité spéciale, d'autoriser le pays à rester. Lorsqu'un pays cesse d'être membre, il est encore tenu de respecter ses obligations contractuelles, par exemple de payer le service de la dette. De même, il doit toujours répondre de sa souscription impayée qui résulte des pertes subies par une institution du Groupe de la Banque sur les garanties et les emprunts non remboursés en date de son retrait.

3.1.1 Logiciel R et Packages

R est un logiciel gratuit et à code source ouvert (open-source). Il fonctionne sous UNIX (et Linux), Windows et Macintosh. C'est donc un logiciel multi-plates-formes. Il est développé dans la mouvance des logiciels libres par une communauté sans cesse plus vaste de bénévoles motivés.

Tout le monde peut d'ailleurs contribuer à son amélioration en y intégrant de nouvelles fonctionnalités ou méthodes d'analyse non encore implémentées. Cela en fait donc un logiciel en rapide et constante évolution.

C'est aussi un outil très puissant et très complet, particulièrement bien adapté pour la mise en œuvre informatique de méthodes statistiques. Il est plus difficile d'accès que certains autres logiciels du marché (comme SPSS ou Minitab par exemple), car il n'est pas conçu pour être utilisé à l'aide de "clics" de souris dans des menus.

R est un logiciel dans lequel de nombreuses techniques statistiques tel que l'analyse descriptive, tests d'hypothèses, etc sont implémentés dans le système.

Une des grandes forces de R réside dans ses capacités, bien supérieures à celles des autres logi-

ciels courants du marché, à combiner un langage de programmation avec la possibilité de réaliser des graphiques de qualité. Les graphiques usuels s'obtiennent aisément au moyen de fonctions prédéfinies.

Mais il est également possible d'effectuer des graphiques plus sophistiqués permettant de représenter des données complexes telles que des courbes de surface ou de niveau, des volumes affichés avec un effet 3D, des courbes de densité, et bien d'autres choses encore. Il nous est également possible d'y ajouter des formules mathématiques.

Quant à la plupart des techniques avancées de statistique sont disponible au travers de module externe appelés package.

Parmi les techniques avancé de statistique qui s'intéressent dans notre thématique, le package FactoMinR et missMDA qui répondent aux objectifs fixés.

FactoMineR est un package R dédié à l'analyse exploratoire multidimensionnelle de données, il permet de mettre en oeuvre des méthodes d'analyses de données telles que l'analyse en composantes principales (ACP), l'analyse des correspondances (AC), l'analyse des correspondances multiples (ACM) ainsi que des analyses plus avancées. Il fournit un point de vue géométrique et de nombreuses sorties graphiques, il fournit de nombreuses aides à l'interprétation (description automatique des axes, nombreux indicateurs, ...), il permet l'ajout d'information supplémentaire telle que des individus et/ou des variables supplémentaires, il gère les données manquantes avec missMDA.

Ce dernier est complémentaire de FactoMinR, il permet de gérer les données manquantes pour les méthodes d'analyses factorielles (ACP, AFC, ACM, ...).

Il permet de faire de l'imputation simple qui consiste à remplacer les valeurs manquantes par des valeurs plausibles. Cela revient à compléter le jeu de données qui peut ensuite être analysé par n'importe quelle méthode d'analyse factorielle.

missMDA impute les valeurs manquantes de sorte que les valeurs imputées n'ont aucune influence sur les résultats de l'analyse factorielle (pas d'influence dans le sens où les valeurs imputées n'ont aucun poids, et donc les résultats de l'analyse factorielle sont obtenues uniquement avec les valeurs observées.

3.2 Application : Etude des indicateurs de développement des pays : Algérie, Tunisie, Maroc

Présentation des données

Notre analyse s'effectue sur trois tableaux, chaque tableau est composé de 14 variables extrait de l'ensemble des indicateurs de développement de la banque mondiale durant la période allant de 1960 à 2017 :

A : Espérance de vie à la naissance

B : Population urbaine

C : Population rurale

D : Subventions de coopération technique (BDP, dollars US courants)

E : Subventions, hors coopération technique (BDP, dollars US courants)

F : Taux de change officiel (unités de devises locales par dollar US, moyenne pour la période)

G : Ratio de la population inactive à la population active (% de la population en âge de travailler)

H : Aide publique au développement nette et aide publique reçues (dollar US courants)

I : Flux nets de l'aide bilatérale des donateurs du CAD, total (dollar US courants)

J : Production céréalière (tonnes métriques)

K : Machinerie agricole, tracteurs par 100 km carrés de terres arables

L : Masse monétaire (% du PIB)

M : Exportations de marchandises vers les pays à revenu élevé (% des exportations totales de marchandises)

N : Importations de marchandises des pays à revenu élevé (% des importations totales de marchandises)

3.2.1 Etude des indicateurs de développement du Maroc

Cas de l'Algérie

le tableau de donnée de l'Algérie

Le tableau 3.12 compose de 14 variables et 58 individus sont stockées dans des tableaux excel sous forme .csv sous le nom (ACPDZ). Pour analyser ce tableau il suffit d'utiliser la commande :

```
X<-read.table(file="ACPDZ.csv",header=TRUE,sep=";",dec=",")
```

Nous procédons à l'extraction des nouveau tableaux dont nous appliquons l'ACP :

```
Y<-X[,2:15]
```

Estimation des données manquantes pour l'Algérie

Nous avons rencontré une difficulté particulière relative aux manque de données lors du traitement de l'algérie. Il est donc impératif de procéder à l'estimation de ces dernières avant toute analyse. Ces estimations peuvent avoir un impact sur notre analyse mais l'objectif ici n'est pas de développer les méthodes d'estimation de données manquantes. Nous avons utiliser la library MissMDA qui utilise la technique d'interpolation.

```
nb <- estim-ncpPCA(Y,ncp.$\min$=0,ncp.$\max$=14,method.cv="kfold")
nb\ncp
```

```
Z<- imputePCA(Y,ncp=4)}
Z\completeObs
```

Après l'estimation des valeurs manquantes, nous aurons en notre possession un tableau "complet" de données 3.2.1 qui servira d'outil pour effectuer notre analyse.

La commande qui permet d'obtenir le tableau des valeurs propres est :

```
Z.acp\seig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	7.67	54.80	54.80
comp 2	3.16	22.58	77.38
comp 3	1.13	8.07	85.45
comp 4	0.66	4.69	90.14
comp 5	0.45	3.24	93.38
comp 6	0.32	2.30	95.68
comp 7	0.26	1.85	97.53
comp 8	0.17	1.20	98.72
comp 9	0.06	0.45	99.17
comp 10	0.05	0.39	99.56
comp 11	0.04	0.27	99.83
comp 12	0.01	0.08	99.91
comp 13	0.01	0.07	99.99
comp 14	0.00	0.01	100.00

Table 3.1 – Tableau des valeurs propres(Algérie)

Le premier axe concerne 55% de l'inertie du nuage, il existe une structuration des données qui va se manifester sur le premier axe.

le second axe conserve une part importante de l'inertie total, 23%.

La chute est importante dès le troisième axe qui conserve 8% de l'inertie total.

Nous décidons de ne retenir que les deux premiers axes car il est compréhensible par l'oeil (c'est un plan), il explique 78% de l'information préservée. **Matrice de corrélation**

```
R<-cor(Z$completeObs)
```

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	1.00	0.95	0.88	0.76	-0.27	0.84	-0.89	0.17	0.10	0.58	0.97	0.55	-0.45	-0.77
B	0.95	1.00	0.71	0.73	-0.16	0.94	-0.95	0.14	0.13	0.66	0.92	0.50	-0.56	-0.87
C	0.88	0.71	1.00	0.64	-0.44	0.56	-0.62	0.16	0.01	0.34	0.86	0.51	-0.27	-0.47
D	0.76	0.73	0.64	1.00	-0.17	0.64	-0.76	0.26	0.20	0.60	0.73	0.40	-0.24	-0.59
E	-0.27	-0.16	-0.44	-0.17	1.00	-0.04	-0.05	0.63	0.71	0.07	-0.26	-0.67	0.23	0.26
F	0.84	0.94	0.56	0.64	-0.04	1.00	-0.95	0.17	0.22	0.62	0.80	0.29	-0.55	-0.79
G	-0.89	-0.95	-0.62	-0.76	-0.05	-0.95	1.00	-0.30	-0.33	-0.67	-0.85	-0.30	0.47	0.76
H	0.17	0.14	0.16	0.26	0.63	0.17	-0.30	1.00	0.90	0.18	0.19	-0.49	0.11	0.15
I	0.10	0.13	0.01	0.20	0.71	0.22	-0.33	0.90	1.00	0.21	0.11	-0.55	0.13	0.10
J	0.58	0.66	0.34	0.60	0.07	0.62	-0.67	0.18	0.21	1.00	0.57	0.30	-0.42	-0.63
K	0.97	0.92	0.86	0.73	-0.26	0.80	-0.85	0.19	0.11	0.57	1.00	0.54	-0.54	-0.74
L	0.55	0.50	0.51	0.40	-0.67	0.29	-0.30	-0.49	-0.55	0.30	0.54	1.00	-0.33	-0.68
M	-0.45	-0.56	-0.27	-0.24	0.23	-0.55	0.47	0.11	0.13	-0.42	-0.54	-0.33	1.00	0.60
N	-0.77	-0.87	-0.47	-0.59	0.26	-0.79	0.76	0.15	0.10	-0.63	-0.74	-0.68	0.60	1.00

Table 3.2 – Matrice de corrélation(Algérie)

Après avoir appliqué la commande PCA

```
Z.acp<- PCA(Z\$completeObs)
```

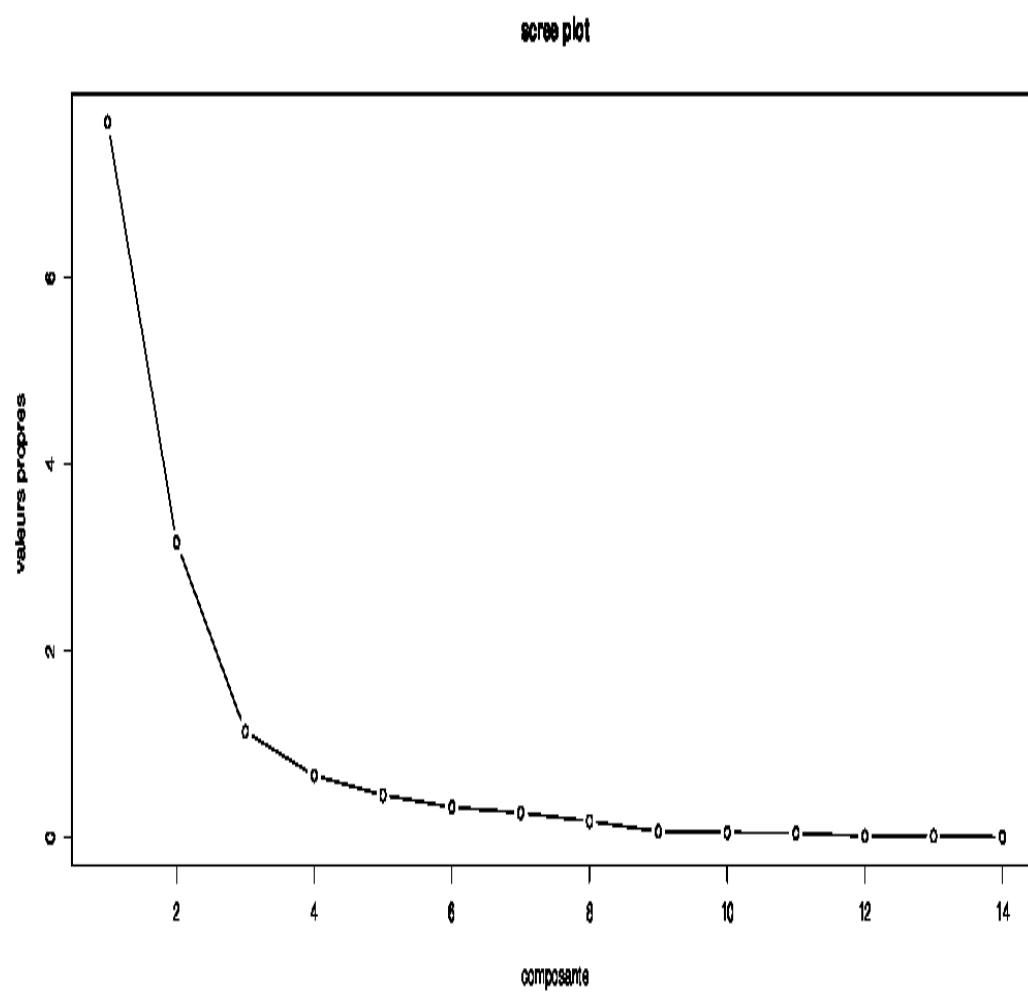


Figure 3.1 – Représentation des valeurs propres(Algérie)

Nous obtenons le cercle de corrélation et le graphe des individus.

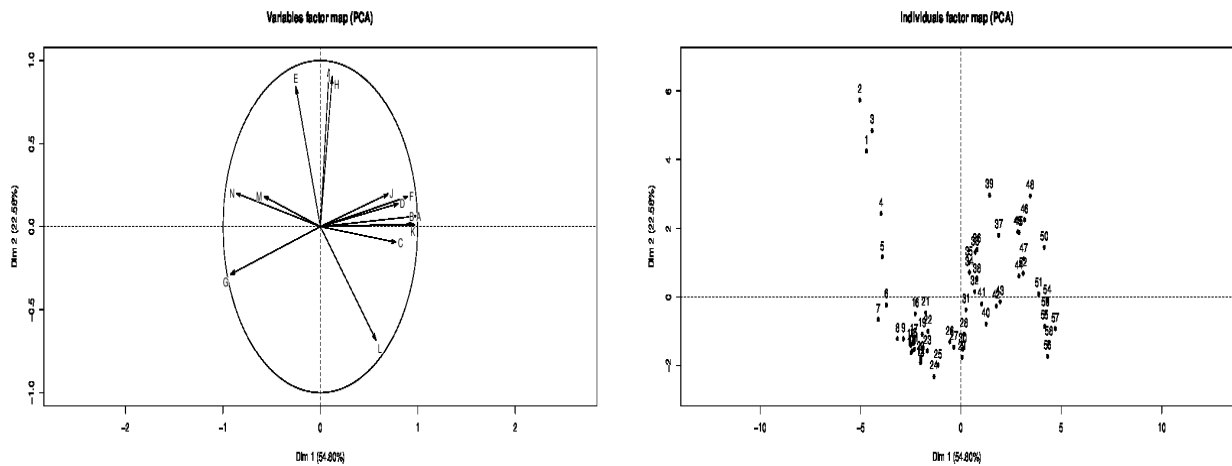


Figure 3.2 – Graphe des individus et cercle de corrélation (Maroc)

Contribution des individus et des variables

Les individus qui contribuent plus à la formation de premier axe de graphe des individus ceux dont les coordonnées sur cet axe sont les plus élevés en valeur absolue, autrement dit, se sont les individus qui prennent des valeurs supérieures à

$$\sqrt{\lambda_1} = 2.769$$

en d'autre terme :

$$|c_{i1}| > \sqrt{\lambda_1} = 2.769.$$

Ces individus extraient du tableau suivant :

```
> acp$scorces[,2]
 [1] 4.24660237 5.74163788 4.84242053 2.42632801 1.16029398 -0.24488561
 [7] -0.66508189 -1.22092224 -1.22776075 -1.41194938 -1.52873039 -1.86580101
[13] -1.92758831 -1.63319928 -1.37127472 -0.49524728 -1.26288396 -1.58196967
[19] -1.09563233 -1.76601380 -0.46575329 -1.00253334 -1.57971423 -2.33072382
[25] -1.99236171 -1.31253658 -1.46731370 -1.09023748 -1.76404508 -1.52100805
[31] -0.37820578 0.15114592 1.29991884 0.71462911 0.99654296 1.37531735
[37] 1.79749892 0.49190433 2.95958666 -0.79322888 -0.20885641 -0.27223215
[43] -0.14462422 0.60434605 1.87733573 2.24455196 1.10077946 2.94330114
[49] 1.89669531 1.44102883 0.08220794 0.68256114 -0.46913609 -0.12243505
[55] -0.86485031 -1.73969232 -0.93033238 -1.32787293
```

Quant aux variables qui contribuent à la formation de premier axe de cercle de corrélation sont les variables qui prennent des valeurs supérieures à

$$|\mu_{jk}| > 1/\sqrt{p} = 1/\sqrt{14} = 0.267$$

en d'autre terme :

$$|\mu_{jk}| > 1/\sqrt{p} = 0.267$$

```
> loadings(acp)[,1]
      A      B      C      D      E      F
0.34942515 0.35392428 0.27983782 0.28771061 -0.09099626 0.32315433
      G      H      I      J      K      L
-0.33314593 0.04440120 0.03259679 0.25152258 0.34358555 0.20695636
      M      N
-0.20832433 -0.31189613
```

Le tableau suivant nous résume les variables et les individus qui contribuent à la construction de premier axe.

Individu		Individu			Variable	Variable	
-		+			-	+	
1 ; 2	3 ; 4	44 ; 45	46 ; 47	G,N		A,B	
5 ; 6	7 ; 8	48 ; 49	50 ; 51				C,D,F
9	11 ; 12	52 ; 53	54 ; 55 ; 56 ; 57 ; 58				K

La même démarche qu'on suit pour la construction de deuxième axe de graphe des individus, les valeurs supérieures à $\sqrt{\lambda_2}$ sont celle qui ont contribué à la formation de cet axe.

Pour les variables, nous sélectionons toujours celles qui prennent les valeurs supérieurs à $1/\sqrt{p}$.

```
> acp$scores[,2]
[1] 4.24660237 5.74163788 4.84242053 2.42632801 1.16029398 -0.24488561
[7] -0.66508189 -1.22092224 -1.22776075 -1.41194938 -1.52873039 -1.86580101
[13] -1.92758831 -1.63319928 -1.37127472 -0.49524728 -1.26288396 -1.58196967
[19] -1.09563233 -1.76601380 -0.46575329 -1.00253334 -1.57971423 -2.33072382
[25] -1.99236171 -1.31253658 -1.46731370 -1.09023748 -1.76404508 -1.52100805
[31] -0.37820578 0.15114592 1.29991884 0.71462911 0.99654296 1.37531735
[37] 1.79749892 0.49190433 2.95958666 -0.79322888 -0.20885641 -0.27223215
[43] -0.14462422 0.60434605 1.87733573 2.24455196 1.10077946 2.94330114
[49] 1.89669531 1.44102883 0.08220794 0.68256114 -0.46913609 -0.12243505
[55] -0.86485031 -1.73969232 -0.93033238 -1.32787293

> loadings(acp)[,2]
      A      B      C      D      E      F
0.006345617 0.035284683 -0.053026754 0.077290617 0.474497930 0.101153677
      G      H      I      J      K      L
-0.161571099 0.509666852 0.533495529 0.110570523 0.008641212 -0.385521614
      M      N
0.101865772 0.112967405
```

Le tableau suivant nous résume les variables et les individus qui contribuent à la construction de second axe.

Individu		Individu			Variable	Variable
-		+			-	+
12 ; 13	24	1 ; 2 ; 3 ; 4				E,H
25		37 ; 39	45 ; 46			I
		48 ; 49				L

Interprétation

Les variables A,B,C,D,F et K possèdent une corrélation positive deux à deux surtout plus fort entre A,B,C. Ces variables sont corrélées négativement avec les deux variables G et N pendant les deux périodes (1960-1971) et (2003-2017), la première période oppose les variables A,B,C,D,F et K, elles contribuent négativement à leurs corrélations par contre elles contribuent positivement à celle des deux variables G et N.

Contrairement à la deuxième période qui contribue positivement à la corrélation des variables citées au début, et négativement à celles de G et N.

Les variables E,H et I sont corrélés deux à deux et ne possèdent pas une corrélation avec celles que nous avons déjà citées pendant les années situées dans le deuxième tableau des contribution. Quant à la qualité de représentation, nous trouvons que sauf J et M qui sont assez mal représentées et elles n'ont pas participé à la construction des axes principaux.

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	9.94	70.98	70.98
comp 2	1.27	9.06	80.05
comp 3	0.75	5.36	85.40
comp 4	0.68	4.87	90.27
comp 5	0.52	3.71	93.99
comp 6	0.39	2.80	96.79
comp 7	0.25	1.81	98.60
comp 8	0.09	0.63	99.23
comp 9	0.05	0.34	99.57
comp 10	0.04	0.26	99.83
comp 11	0.01	0.10	99.93
comp 12	0.01	0.04	99.97
comp 13	0.00	0.02	99.99
comp 14	0.00	0.01	100.00

Table 3.3 – Tableau des valeurs propres (Maroc)

Cas de Maroc

Pour rappel, l'objectif ici est de traiter les indicateurs de développement du Maroc. Pour précision, nous déterminons les composantes principales, c'est-à-dire une combinaison linéaire de variables considérées permettant de réduire les dimensions sous contraintes de sauvegarder le maximum d'information.

A travers le tableau 3.3, il ressort clairement qu'en retenant les deux premières composantes principales c_1 et c_2 , 80.05% de l'information est préservée. Ce qui est jugé très satisfaisant. Ceci est aussi validé par la figure 3.3 des valeurs propre où le coude apparaît à partir de la seconde valeur propre.

Remarque 3.1 *On peut se contenter de la première composante principale car elle sauvegarde 70% de l'information. En ce qui nous concerne, nous estimons que vaut mieux retenir les deux première composantes principales et éteindre à 80% l'information préservée. En outre, comme toute science décisionnelle le dernier mot revient au décideur.*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	1.00	0.99	0.96	0.85	0.73	0.83	-0.98	0.79	0.87	0.55	0.98	0.91	-0.12	-0.66
B	0.99	1.00	0.92	0.82	0.73	0.80	-0.98	0.81	0.90	0.55	0.97	0.94	-0.09	-0.67
C	0.96	0.92	1.00	0.85	0.66	0.82	-0.91	0.70	0.73	0.52	0.96	0.77	-0.19	-0.61
D	0.85	0.82	0.85	1.00	0.72	0.69	-0.85	0.57	0.63	0.40	0.82	0.76	0.07	-0.58
E	0.73	0.73	0.66	0.72	1.00	0.54	-0.73	0.81	0.76	0.41	0.74	0.71	-0.17	-0.53
F	0.83	0.80	0.82	0.69	0.54	1.00	-0.81	0.49	0.60	0.40	0.80	0.61	0.00	-0.46
G	-0.98	-0.98	-0.91	-0.85	-0.73	-0.81	1.00	-0.78	-0.86	-0.50	-0.96	-0.92	0.06	0.61
H	0.79	0.81	0.70	0.57	0.81	0.49	-0.78	1.00	0.92	0.47	0.83	0.79	-0.34	-0.51
I	0.87	0.90	0.73	0.63	0.76	0.60	-0.86	0.92	1.00	0.51	0.86	0.90	-0.18	-0.62
J	0.55	0.55	0.52	0.40	0.41	0.40	-0.50	0.47	0.51	1.00	0.53	0.50	-0.19	-0.50
K	0.98	0.97	0.96	0.82	0.74	0.80	-0.96	0.83	0.86	0.53	1.00	0.86	-0.25	-0.63
L	0.91	0.94	0.77	0.76	0.71	0.61	-0.92	0.79	0.90	0.50	0.86	1.00	0.04	-0.62
M	-0.12	-0.09	-0.19	0.07	-0.17	0.00	0.06	-0.34	-0.18	-0.19	-0.25	0.04	1.00	0.24
N	-0.66	-0.67	-0.61	-0.58	-0.53	-0.46	0.61	-0.51	-0.62	-0.50	-0.63	-0.62	0.24	1.00

Table 3.4 – Tableau de corrélation des variables (Maroc)

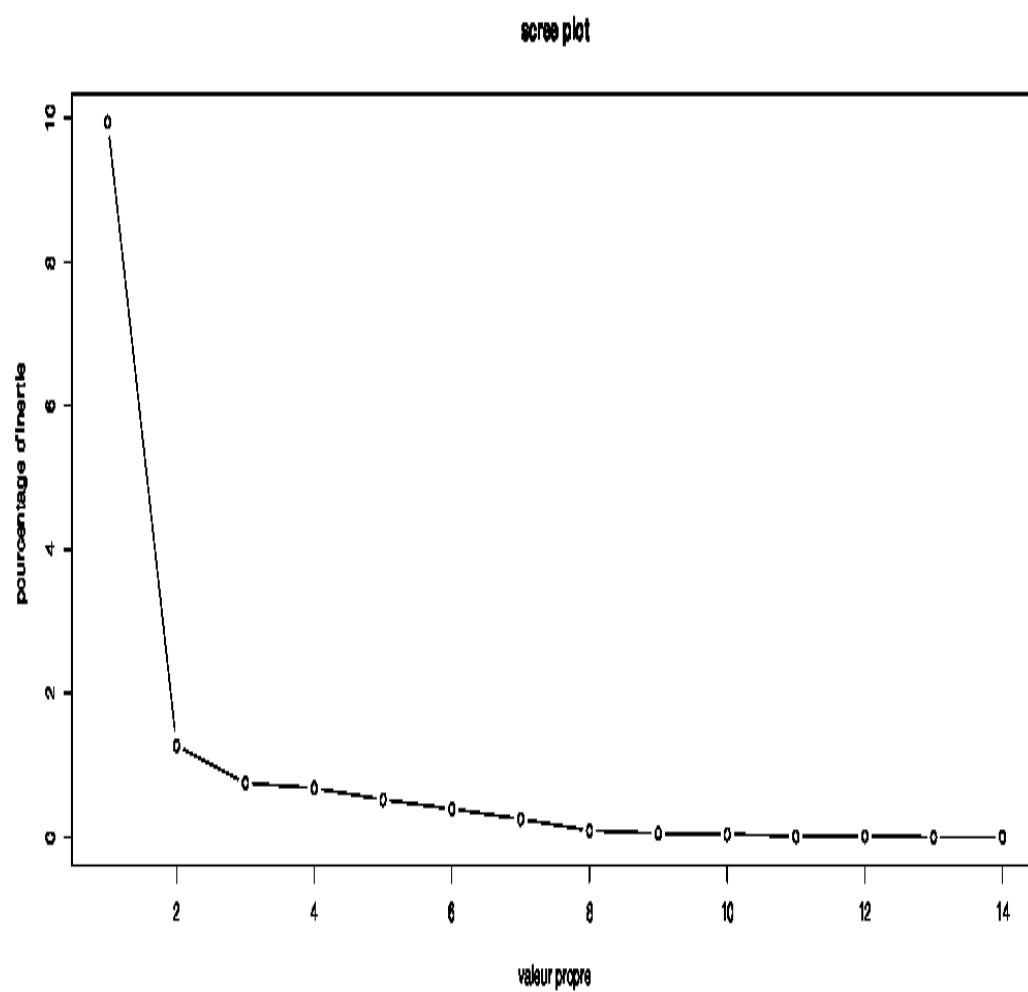


Figure 3.3 – Représentation des valeurs propres (Maroc)

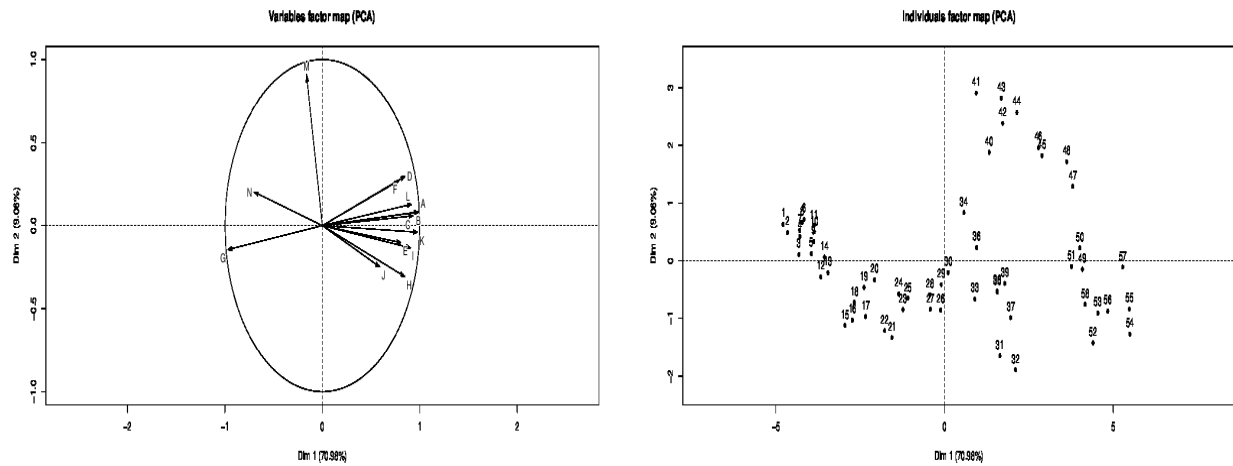


Figure 3.4 – Graphes des corrélations et des individus (Maroc)

Contribution des variables et des individus

```
> acp$scores[,1]
[1] -4.78301327 -4.65209767 -4.31177130 -4.28713164 -3.94589253 -4.16543114
[7] -4.29251179 -4.21763579 -3.87399348 -3.85468013 -3.86859907 -3.66712021
[13] -3.45202800 -3.55524897 -2.94932610 -2.72912242 -2.33943475 -2.67138090
[19] -2.38286678 -2.07300604 -1.55740139 -1.77517184 -1.23059565 -1.35981670
[25] -1.08338494 -0.11507486 -0.42005556 -0.43203039 -0.09669915 0.10884333
[31] 1.64108430 2.10757091 0.89617542 0.57456748 1.56482284 0.95432278
[37] 1.96405732 1.56046056 1.78742970 1.32567961 0.94025959 1.72447114
[43] 1.68357585 2.14579045 2.89062066 2.78357722 3.80087475 3.62166875
[49] 4.08760566 4.01140879 3.76224138 4.40375935 4.54425303 5.49255887
[55] 5.47844270 4.83803376 5.28170150 4.16666478
```

```
> loadings(acp)[,1]
      A      B      C      D      E      F
0.31373778 0.31294595 0.29488251 0.26885898 0.25538821 0.24970863
      G      H      I      J      K      L
-0.30779426 0.26826247 0.28722294 0.18694160 0.31010484 0.29060091
      M      N
-0.05171167 -0.22221797
```

Le tableau suivant nous résume les variables et les individus qui contribuent à la construction de premier axe.

Individi		Individu			Variable	Variable
-	+	-	+	-	+	
1 ; 2	3 ; 4	47 ; 48	49 ; 50	G		A, B
5 ; 6	7 ; 8	51 ; 52	53 ; 54			H, I, C, D
9 ; 10	11 ; 12	55 ; 56	57 ; 58			K, L

Pour le second axe :

```
> acp$scores[,2]
[1] 0.62759237 0.48672343 0.10647851 0.41719883 0.12225594 0.71323089
[7] 0.52648846 0.66372166 0.33238357 0.50084098 0.60408186 -0.28367618
[13] -0.20749818 0.06097330 -1.12633729 -1.03390042 -0.97058845 -0.72153186
[19] -0.46462637 -0.33302581 -1.33599836 -1.21316215 -0.85013400 -0.57676142
[25] -0.65206280 -0.85795229 -0.84448051 -0.58977127 -0.41404451 -0.20470026
[31] -1.64927324 -1.89381469 -0.66649730 0.83465750 -0.54377868 0.22585300
[37] -0.98862432 -0.52640117 -0.39398870 1.87843345 2.90749723 2.38320031
[43] 2.81950648 2.57262159 1.82120310 1.95768744 1.28723488 1.71680698
[49] -0.15030774 0.22018991 -0.09984888 -1.42606330 -0.90941745 -1.27328439
[55] -0.84073999 -0.87573923 -0.11230905 -0.75652145
```

```
> loadings(cop)[,2]
      A      B      C      D      E      F
0.07280316 0.07307260 0.06187122 0.26276078 0.09111591 0.221063952
      G      H      I      J      K      L
0.12807320 0.27205109 0.11920208 0.22071931 0.03522280 0.11105652
      M      N
0.80722123 0.17917161
```

Individi		Individu			Variable	Variable
-		+		-	+	
15 ; 16	21 ;22	40 ;41	42 ;43	H	M	
31 ; 32	52 ;54	44 ;45	46 ;47			
		48				

Interprétation

Les indicateurs du développements A,B,H,I,C,D,K et L possèdent des corrélations (fortes et assez fortes) deux à deux et ils opposent l'indicateur G tel que dans la position droite à l'axe des coordonnées contribuent positivement à cette corrélation et ce contrairement aux individus situés à gauche (1960-1971).

L'indicateur A,B,C et K sont fortement corrélés et le mérite est du essentiellement aux individus situé au quatrième quadrant.

La variable M ne possède aucune corrélation avec toutes les variables que nous avons déjà cité.

Quant à la qualité de représentation, la plupart des variables sont bien représentées à l'exception des variables E,F,J et N qui sont assez mal représentées et elles n'interviennent pas dans la construction des axes principaux.

3.2.2 Cas de Tunisie

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	10.11	72.19	72.19
comp 2	1.32	9.46	81.65
comp 3	0.86	6.15	87.81
comp 4	0.71	5.07	92.88
comp 5	0.42	2.97	95.85
comp 6	0.26	1.88	97.73
comp 7	0.16	1.12	98.85
comp 8	0.08	0.59	99.44
comp 9	0.03	0.19	99.63
comp 10	0.02	0.16	99.79
comp 11	0.02	0.11	99.89
comp 12	0.01	0.07	99.97
comp 13	0.00	0.02	99.99
comp 14	0.00	0.01	100.00

Table 3.5 – Tableau des valeurs propres(Tunisie)

Selon le tableau 3.5 des valeurs propres, le premier axe concerne 72% de l'inertie totale. Il existe une structuration des données qui se manifeste sur le premier axe.

Le second axe conserve 9% de l'inertie totale. Donc nous retenons les deux premières composantes principales. Ce qui nous permet de préserver 81% de l'information initiale.

La figure 3.5 des valeurs propres nous indique que le coude apparait à partir de la deuxième valeur propre.

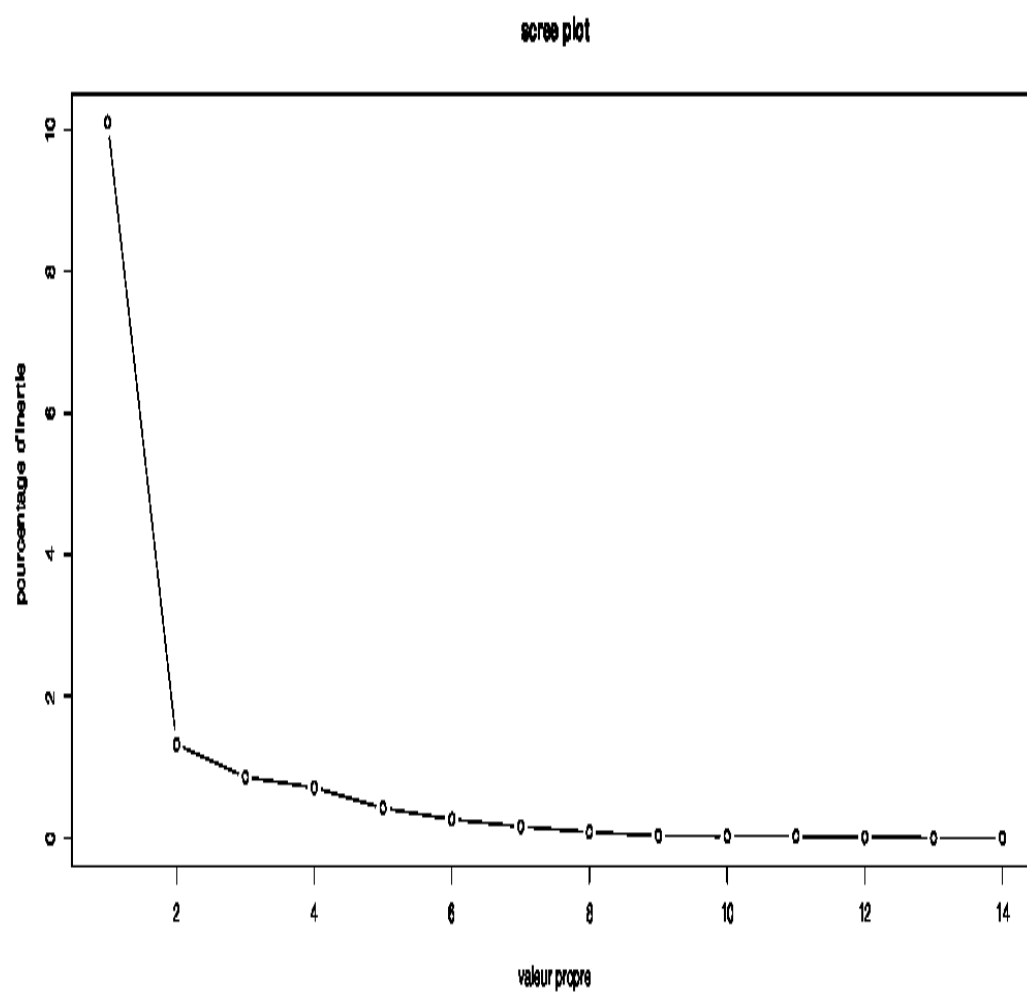


Figure 3.5 – Représentation des valeurs propres (Tunisie)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	1.00	0.95	0.99	0.86	0.57	0.81	-0.91	0.65	0.67	0.54	0.90	0.88	0.40	-0.55
B	0.95	1.00	0.93	0.87	0.69	0.93	-0.98	0.74	0.79	0.56	0.89	0.94	0.48	-0.70
C	0.99	0.93	1.00	0.83	0.56	0.81	-0.89	0.62	0.64	0.54	0.87	0.87	0.33	-0.54
D	0.86	0.87	0.83	1.00	0.74	0.71	-0.88	0.64	0.65	0.53	0.92	0.81	0.41	-0.61
E	0.57	0.69	0.56	0.74	1.00	0.65	-0.71	0.82	0.82	0.51	0.51	0.73	0.32	-0.71
F	0.81	0.93	0.81	0.71	0.65	1.00	-0.91	0.75	0.81	0.44	0.75	0.91	0.41	-0.77
G	-0.91	-0.98	-0.89	-0.88	-0.71	-0.91	1.00	-0.74	-0.79	-0.56	-0.89	-0.91	-0.55	0.67
H	0.65	0.74	0.62	0.64	0.82	0.75	-0.74	1.00	0.98	0.51	0.48	0.88	0.35	-0.80
I	0.67	0.79	0.64	0.65	0.82	0.81	-0.79	0.98	1.00	0.53	0.52	0.89	0.40	-0.79
J	0.54	0.56	0.54	0.53	0.51	0.44	-0.56	0.51	0.53	1.00	0.46	0.49	0.29	-0.35
K	0.90	0.89	0.87	0.92	0.51	0.75	-0.89	0.48	0.52	0.46	1.00	0.77	0.44	-0.50
L	0.88	0.94	0.87	0.81	0.73	0.91	-0.91	0.88	0.89	0.49	0.77	1.00	0.39	-0.82
M	0.40	0.48	0.33	0.41	0.32	0.41	-0.55	0.35	0.40	0.29	0.44	0.39	1.00	-0.14
N	-0.55	-0.70	-0.54	-0.61	-0.71	-0.77	0.67	-0.80	-0.79	-0.35	-0.50	-0.82	-0.14	1.00

Table 3.6 – Tableau des corrélations(Tunisie)

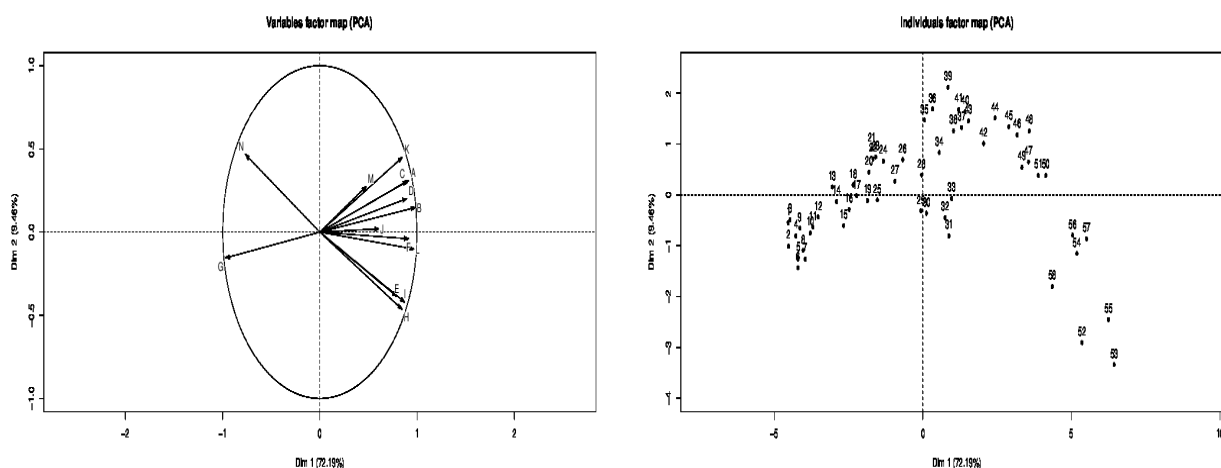


Figure 3.6 – Cercle de corrélation et graphe des individus (Tunisie)

Contribution des individus et des variables

A travers les commandes utilisées avec le logiciel R , nous obtenons les tableaux ci-dessus :

```
> acp$scores[,1]
[1] -4.53109480 -4.51872762 -4.47158792 -4.27315629 -4.18950942 -4.19982881
[7] -3.95859021 -4.03080838 -4.13661479 -3.78653310 -3.69678601 -3.52022837
[13] -3.04901853 -2.90785111 -2.66715466 -2.48149314 -2.22583944 -2.34350199
[19] -1.86086532 -1.81478153 -1.71941786 -1.67228624 -1.58268035 -1.32825511
[25] -1.53018873 -0.67373628 -0.94061710 -0.03990976 -0.06572228 0.12350171
[31] 0.88201283 0.74821960 0.97083084 0.55186316 0.05641188 0.32597422
[37] 1.30404521 1.03377548 0.84866411 1.42696030 1.20715653 2.04582974
[43] 1.53533064 2.42969078 2.89576371 3.17454758 3.55817703 3.58241433
[49] 3.33669888 4.14109232 3.88965132 5.35686008 6.44100115 5.18547506
[55] 6.24884214 5.04402505 5.51161482 4.36035464
```

```
> loadings(acp)[,1]
      A      B      C      D      E      F      G
0.2882798 0.3079820 0.2820646 0.2810154 0.2481666 0.2871001 -0.3050890
      H      I      J      K      L      M      N
0.2657093 0.2741601 0.1902910 0.2669479 0.3039980 0.1498822 -0.2414600
```

A partir de ces derniers tableaux, nous obtenons des tableaux ci-dessus qui résument les variables et les individus qui contribuent à la construction de premier axe.

Individi		Individu			Variable	Variable
-		+			+	
1 ; 2	3 ; 4	47 ; 48	49 ; 50	G		A,B
5 ; 6	7 ; 8	51 ; 52	53 ; 54			C,D,F
9 ; 10	11 ; 12	55 ; 56	57 ; 58			I,L

Pour la construction de deuxième axe, nous utilisons le tableau suivant :

```
> acp$scores[,2]
[1] -0.54636406 -1.00864503 -0.49088094 -0.80556512 -1.23710474 -1.43344755
[7] -1.26432239 -1.09221382 -0.65569467 -0.75029786 -0.63526573 -0.43440325
[13] 0.15545012 -0.13090556 -0.60152050 -0.28609270 -0.01665544 0.19459968
[19] -0.11035983 0.44602180 0.90274636 0.71185676 0.75009839 0.66173351
[25] -0.09822788 0.69419160 0.26861240 0.39257326 -0.31571528 -0.36076473
[31] -0.80666941 -0.44949022 -0.07473076 0.83322986 1.47316597 1.68998470
[37] 1.32669350 1.26110322 2.11768095 1.63366494 1.68188911 1.00972844
[43] 1.45463462 1.51295821 1.34183089 1.18062410 0.65000831 1.25609457
[49] 0.54254412 0.38247245 0.38116359 -2.90402444 -3.33599883 -1.15173096
[55] -2.45357717 -0.79108749 -0.86413984 -1.80145920
```

```
> loadings(acp)[,2]
      A      B      C      D      E      F
0.27009437 0.12771492 0.26535316 0.17458624 -0.33500309 -0.03477024
      G      H      I      J      K      L
-0.13491135 -0.40304757 -0.36546576 0.01701537 0.39022425 -0.08849385
      M      N
0.23832388 0.40620399
```

Nous obtenons :

Individi		Individu			Variable	Variable
-		+			+	
5 ; 6	35 ; 36	37,38	39 ; 40	E,H,I		A
7 ; 52	53 ; 54	41 ; 43	44 ; 45			K
55 ; 58			46			N

Interprétation

Les indicateurs A,B,C,D,F,I et L sont corrélés deux à deux et anticorrélés avec l'indicateur G dans les deux périodes qui s'étale de 1960 à 1971 et de 2006 à 2017. La première période qui se situe au côté gauche des graphes des individus contribue négativement aux corrélations des variables A,B,C,D,F et L. Contrairement à la deuxième période. Aussi la variable A est corrélée avec les deux variables K et N qui possèdent de leur tour une corrélation et elles opposent les variables E,H et I pendant les années situées dans le tableau des contributions. Concernant la qualité de représentation, la plupart des variables sont bien représentées et participées à la construction des axes principaux, à l'exception des deux variables J et M.

3.2.3 Etude comparative

Les investigations que nous avons menées lors de cette étude sur les facteurs du développement des trois pays Algérie, Maroc et Tunisie, nous permettent de conclure :

- D'un point de vue statistique, nous avons réduit les dimension de 14 à 2. Ce qui est acceptable sur le plan théorique.
- Les inducteurs : Taux de vie à la naissance, population urbaine, population rurale, subvention de coopération technique, le ratio de la population inactive à la population active, aide publique au développement nette et aide publique reçue, flux nets de l'aide bilatérale des donateurs du CAD, machinerie agricole et la masse monétaire interviennent pratiquement dans la construction des axes des trois pays particulièrement les cinq premiers indicateurs cités qui ont contribué dans la construction de premier axe des trois pays.
- Contrairement aux indicateurs sus-cités, les indicateurs interviennent dans la construction des axes principaux pour certains pays et pas d'autres, particulièrement l'indicateur : Production céréalière qui n'a pas participer dans aucune construction , et ce, en raison des stratégies adoptés par ces derniers.
- Quant aux individus, nous constatons que les années 90-98 ont un impact presque négligeable, particulièrement sur l'Algérie. Ceci est expliqué très probablement à l'instabilité politique

3.3 Classement des pays africains en fonction de leurs développement économique

Présentation des données

Dans cette partie, nous nous intéressons au classement des pays africains en fonction des indicateurs du développement économique. Techniquement, nous réalisons une analyse discriminante à base de la fonction score afin de pouvoir déterminer les facteurs (axes) permettant de séparer le mieux possible (discriminer) les pays africains (individus).

Nous ciblons pour la construction de cette analyse deux types de pays : les pays en voie de développement et les pays non développés.

Notre but est de chercher à diviser l'échantillon en deux groupes. On a choisis les indicateurs importants qui indiquent le développement d'un pays.

Les variables retenues pour notre étude sont fournis par base de données de la Banque Mondiale.

PIB : Produit intérieur brute(\$ US courants).

Mass M : masse monétaire (%).

Inflation : inflation, déflateur du PIB (% annuel).

PT : Population total.

FMI : utilisation des crédits du FMI (dette en cours et décaissée, \$ US courant).

PA : population active.

IPC : indice des prix à la consommation (2010=100).

DEND : dette extérieur non décaissée, total(dette non décaissée,\$ US courant).

Introduire les données sous R

```

Y<-read.table(file="IDPAZ.csv", header=TRUE,sep=";", dec=",")
> nrow(Y)
[1] 40
> ncol(Y)
[1] 9
>names(Y)
[1] "classement" "PIB" "Masse.M" "Inflation" "P.T"
[6] "FMI""PA""IPC" "DEND"

```

Le tableau (3.14) que nous avons en notre disposition contient 40 lignes et 9 colonnes, la première colonne correspond à la classification qu'on a effectué au début, les autres colonnes sont les 8 indicateurs de développement.

Dans le but de tester l'indépendance entre les différentes variables à introduire dans le modèle, nous effectuons une analyse de corrélation des 8 indicateurs.

```

>cor(Y[,2:9])
> pairs(Y[,2:9])

```

	PIB	Masse.M	Inflation	P.T	FMI	PA	IPC	DEND
PIB	1.00	0.26	0.25	0.62	0.67	0.49	-0.10	0.41
Masse.M	0.26	1.00	-0.18	0.10	0.33	0.00	-0.08	0.15
Inflation	0.25	-0.18	1.00	0.51	0.37	0.47	-0.13	0.43
P.T	0.62	0.10	0.51	1.00	0.58	0.93	-0.14	0.51
FMI	0.67	0.33	0.37	0.58	1.00	0.45	-0.08	0.73
PA	0.49	0.00	0.47	0.93	0.45	1.00	-0.15	0.51
IPC	-0.10	-0.08	-0.13	-0.14	-0.08	-0.15	1.00	-0.07
DEND	0.41	0.15	0.43	0.51	0.73	0.51	-0.07	1.00

Table 3.7 – Tableau de corrélation

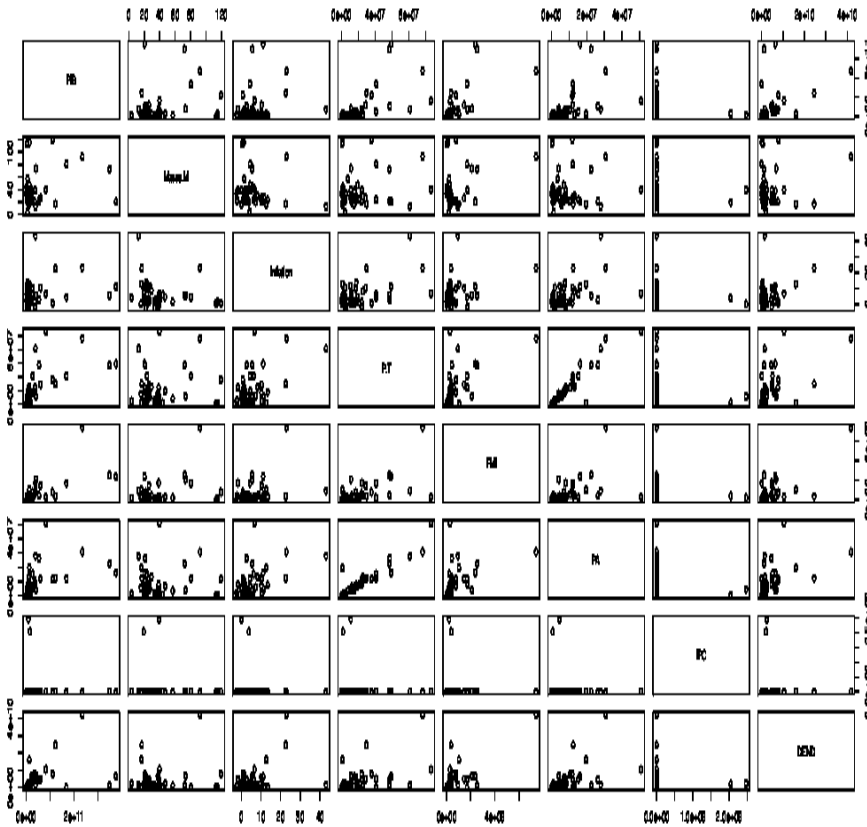


Figure 3.7 – Représentation de la corrélation entre les variables

Le tableau de corrélation 3.7 et la figure 3.7 mettent en évidence la non corrélation de deux variables. Quant aux autres, il reste non évident d'affirmer ou non la corrélation en raison des valeurs situées entre 0.1 et 0.8 , donc la sélection de variable à pouvoir discriminatoire est non trivial. C'est pourquoi, nous allons subir la commande ci-dessus, pour pouvoir sélectionner les variables qui auront un pouvoir discriminatoire.

Pour ce faire, nous allons subdiviser la population mère en deux échantillons, dont les proportions sont 70% et 30%.

Le premier échantillon servira à la construction de fonction score et à effectuer l'analyse discriminante. Le second arbitera la qualité de cette étude. Cette démarche est communément appelée validation croisé.

```
>d = sort(sample(nrow(Y), nrow(Y) * 0.7))
>appren<-Y[d,]
>nrow(appren)
[1] 28}
>test<$-Y[-d,]
>nrow(test)
[1] 12
```

Afin de réaliser notre étude, le préalable est la construction de la fonction score qui repose sur la sélection des variables discriminantes et la création de la fonction de l'analyse discriminante.

Le traitement de notre base des données par logiciel R, nous a permis d'élaborer le résultat suivant :

```
selection<-greedy.wilks(classement~ .,data=appren, niveau=0.1)
selection$formula classement ~ DEND + IPC
```

Nous appliquons donc une analyse discriminante linéaire sur les deux facteurs résultants de la sélection des variables(DEND et IPC) qui discrimine au mieux les deux groupes.

```
>afd=lda(selection$formula,data=appren)
> afd
```

Coefficients of linear discriminants

```
LD1
DEND -2.628576E - 10
IPC -7.069218E3
```

Donc notre fonction score peut s'écrire ainsi :

$$S_{ADL} = -2.628576e - 10 * DEND + -7.069218e - 3 * IPC$$

A partir de la fonction discriminante, on peut calculer les scores Z pour chacun des pays et voir comment cette fonction les classe dans les deux groupes. Les résultats sont résumés dans la matrice de confusion ci-dessus.

La lecture de cette matrice de confusion est la suivante :

```
>pred.lda<-$-predict(afd,newdata=appren)
>mc.a<-$-table(pred.lda$class,appren$classement)
>print(mc.a)
```

	0	1	total
0	12	1	13
1	1	14	15

sur la première ligne, nous voyons que parmi les 13 pays faisant parti des pays non développé, 12 pays sont classés correctement et 1 pays est mal classé.

Sur la deuxième ligne, nous voyons que parmi les 15 pays faisant parti des pays développé, 14 pays sont classés correctement et 1 pays est mal classé.

Afin de valider les résultat de l'étape précédente, il convient de tester ces résultats.

```
>pred.lda<-predict(afd,newdata=test)
>mc.t<-table(pred.lda$class,test$classement)
>print(mc.t)
```

	0	1	total
0	7	0	7
1	0	5	5

```
>print(1-sum(diag(mc.t))/sum(mc.t))
```

0%

La fonction score fournie par le premier échantillon est validé par l'échantillon test. En effet, la proportion de mal classé est de $\frac{2}{28}$ et dans le second échantillon, elle est nulle.

Visualisation et validation du modèle

Cette étape est très importante car le modèle peut donner des faux résultats, dans ce stade nous utilisons la courbe "ROC" dans le but de visualiser et valider le modèle.

La courbe ROC mesure la performance d'un classificateur binaire, présente un outil graphique qui permet d'évaluer et de comparer le modèle.

La courbe sous R

Le package qui permet la visualisation de la courbe est le package ROCR.

Passons par la fonction prédiction.

```
z<- -2.628576e-10*(Y\$$DEND)
q<- -7.069218e-3*(Y\$$IPC)
>a=z+q
roc<-data.frame(score=a,label=Y\$$classement)
pred<-prediction(roc\$$score,roc\$$label)
```

Ensuite faire appel à la fonction performance

```
pref<-performance(pred,"tpr","fpr")
```

enfin tracer la courbe

```
plot(pref)
```

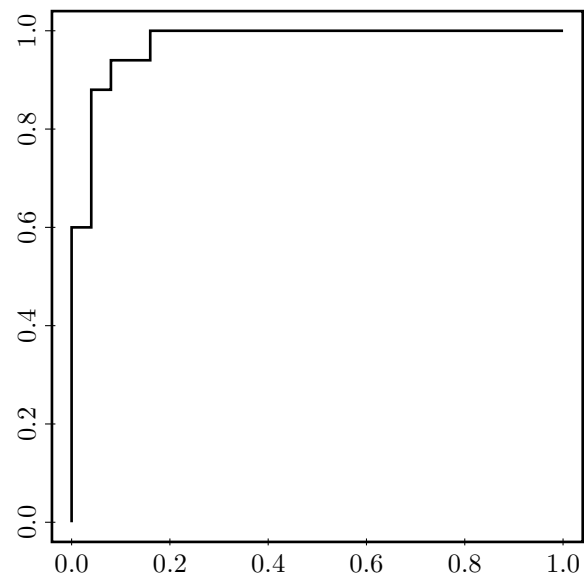


Figure 3.8 – La courbe ROC

La courbe ROC traduit graphiquement la performance de cette fonction. En effet, l'éloignement de la courbe par la première bisectrice est assez considérable pour valider.

Le traitement de données multidimensionnelles est un problème concret et toujours embarrassant lorsqu'il s'agit de données réelles. En effet, dans les applications, on est très souvent heurté à des problèmes telles que l'excès d'information, les données manquantes, les valeurs atypiques pour lesquelles on peut être conduit facilement vers de mauvaises et/ou de falesieuses interprétations.

Ce mémoire est une rétrospective de certaine techniques de traitement de données multidimensionnelles. Se sont des techniques avancées implémentables sur plusieurs logiciels statistiques tel que le R, le SPSS, Eviews ... etc. Toutefois, l'interprétation des résultats chiffrés est assez complexe et nécessite l'usage des plusieurs caractéristiques statistiques à la fois.

Lors de la réalisation de notre mémoire, nous avons été heurté au problème de données manquantes, mais cela n'a pas réduit l'intérêt de l'étude car le nombre de données manquantes est trop faible.

A base des outils statistique relevant de l'ACP et de l'AD à base de la fonction score, nous avons affectué une étude sur les indicateurs du développements des trois pays Algérie, Tunisie et Maroc. Aussi, nous avons effectué une classification d'une quarentaine de pays africains par rapport à la caractéristique du développement.

Le travail réalisé nous a permis de pouvoir mettre en pratique nos acquis thériques. En effet, au delà de l'effort informatique, particulièrement de programmation, nous avons eu en notre possession des données réelles assez complexe en raison des dimensions très élevées.

Enfin, ce travail ouvre le chemin à plusieurs perspectives et d'investigation falesieuses. Il serait intéressant de reprendre l'étude et d'amliorer les interprétations en les accompagnant par des conclusions économiques plus éclairantes. Comme extensions théorique, il serait intéressant d'utiliser des techniques d'estimation de données manquantes plus puissantes et surtout de construire un modèle de séries chronologiques en exploitant tous nos résultats.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	48.46	3619287	8709245	660000	45040000	5.06	91.16	59800000.00	59940000			27.84	77.79	82.17
2	48.88	3793971	8916576	1460000	81540000	5.06	93.96	111330000.00	110570000	1537830	12.14	29.47	77.29	84.04
3	49.31	3973623	9121195	1400000	47900000	5.06	97.12	79790000.00	79090000	3511979	12.29	28.98	77.07	78.78
4	49.74	4157495	9320737	1730000	67320000	5.06	100.20	94510000.00	93480000	3856071	12.33	27.69	79.61	79.14
5	50.17	4344630	9512512	2240000	57510000	5.06	102.55	101380000.00	99840000	3403305	12.54	26.71	79.32	71.83
6	50.59	4533658	9695386	2480000	71650000	5.06	103.82	106610000.00	104130000	3486365	12.66	27.10	82.06	77.36
7	51.01	4724868	9868416	21680000	37990000	5.06	105.25	98210000.00	79960000	2179840	13.04	27.56	78.89	79.36
8	51.41	4918216	10032587	19970000	26940000	5.06	105.78	62210000.00	62550000	3523200	13.62	28.42	80.26	80.05
9	51.81	5114092	10188855	19150000	25230000	5.06	105.58	92690000.00	90290000	6352730	14.00	29.83	79.83	79.50
10	52.20	5313202	10338722	28090000	19360000	5.06	105.02	77630000.00	73260000	4089600	14.47	29.63	79.99	77.05
11	52.57	5516323	10483685	29000000	24250000	5.06	104.24	85560000.00	74090000	4251560	15.12	29.51	80.10	79.39
12	52.94	5723481	10623717	31050000	29980000	5.05	103.43	132520000.00	117640000	5352461	14.53	30.31	74.25	80.51
13	53.30	5955942	10739061	32960000	22780000	4.59	102.27	101450000.00	89180000	5167140	17.70	33.77	75.64	76.95
14	53.68	6197201	10851964	48710000	22740000	4.11	100.81	94160000.00	81400000	3228970	19.52	35.86	75.96	78.14
15	54.07	6449328	10967636	48410000	35800000	4.37	99.08	114750000.00	85280000	4809380	21.67	35.44	69.81	72.66
16	54.49	6714487	11089211	74090000	26040000	4.05	97.16	270110000.00	171170000	3728620	24.07	39.20	70.05	72.79
17	54.98	6993840	11216914	75020000	74340000	4.42	95.17	251510000.00	147410000	5713380	27.05	39.80	71.48	72.01
18	55.52	7286872	11350105	76950000	99030000	4.50	93.09	605690000.00	158670000	2888140	26.92	40.56	69.89	81.50
19	56.14	7594142	11487576	94140000	34940000	4.17	91.01	428860000.00	180460000	4744560	28.25	43.05	72.82	78.02
20	56.82	7915446	11627901	107620000	156620000	3.90	89.02	537380000.00	168930000	4103160	31.83	43.57	73.99	77.57
21	57.56	8250179	11769668	122550000	136240000	3.94	87.19	1160840000.00	199410000	4514720	32.78	35.03	69.42	76.19
22	58.34	8598463	11913138	105740000	109830000	5.17	85.99	1044100000.00	237770000	2130300	38.64	37.92	65.56	83.48
23	59.14	8960310	12056508	86530000	240090000	6.02	84.86	850540000.00	248850000	4916509	40.02	36.37	70.05	79.93
24	59.94	9331960	12196542	93120000	57290000	7.11	83.81	446780000.00	230360000	3581070	40.38	39.46	68.55	78.24
25	60.72	9710432	12327178	90610000	43110000	8.81	82.79	339770000.00	251530000	3759130	39.95	38.49	66.53	77.53
26	61.47	10092012	12445364	94540000	466930000	10.06	81.75	865860000.00	347280000	5312160	40.62	37.71	67.16	80.43
27	62.19	10475430	12548505	116310000	56560000	9.10	81.11	427780000.00	303280000	7824550	41.22	37.53	67.25	77.30
28	62.88	10860432	12637334	138790000	82000000	8.36	80.24	425790000.00	360730000	4336810	41.02	40.17	67.79	74.88
29	63.54	11248397	12713423	149220000	135820000	8.21	79.24	483260000.00	407180000	7958960	41.94	39.90	70.86	77.19
30	64.16	11640605	12780586	152140000	150690000	8.49	78.25	499120000.00	407340000	7428590	43.40	42.31	72.08	75.73
31	64.73	12039263	12839873	181120000	672980000	8.24	77.33	1241090000.00	624160000	6275864	44.97	46.15	67.04	68.36
32	65.25	12444707	12892155	189240000	863660000	8.71	76.19	1532670000.00	659330000	8664685	47.03	47.70	66.15	73.95
33	65.72	12855512	12935982	197120000	190260000	8.54	75.29	953600000.00	783460000	2950232	45.89	50.90	67.91	72.98
34	66.13	13267999	12969418	202490000	363610000	9.30	74.42	780000000.00	629330000	2817864	45.56	53.72	74.75	81.77
35	66.50	13678862	12988186	200400000	333550000	9.20	73.30	683520000.00	517000000	9638630	44.94	53.18	71.88	69.49
36	66.85	13995729	13079503	263320000	107860000	8.54	71.83	500560000.00	454270000	1783230	45.96	55.96	71.30	66.74
37	67.18	14285555	13175048	246830000	357720000	8.72	70.47	649590000.00	608690000	10103304	45.07	52.84	68.94	67.90
38	67.53	14566859	13259042	214110000	289400000	9.53	68.72	471560000.00	414430000	4097671	48.14	61.93	67.36	66.42
39	67.89	14842165	13333098	226940000	355620000	9.60	66.78	531210000.00	486370000	6629516	47.85	60.99	68.88	70.30
40	68.28	15114839	13399959	229570000	321600000	9.80	64.95	681940000.00	637350000	3843292	49.02	66.14	80.96	79.73
41	68.72	15386945	13462676	192090000	219330000	10.63	63.38	447380000.00	410390000	1996649		70.91	85.90	77.43
42	69.21	15659555	13522277	210880000	185540000	11.30	61.77	493850000.00	474210000	4603350		77.54	84.40	71.77
43	69.74	15933727	13578641	282930000	172260000	11.02	60.53	364390000.00	344190000	5287815		81.96	84.96	75.67
44	70.30	16210331	13633606	349990000	276470000	9.57	59.52	584390000.00	483500000	7969118		82.55	85.40	78.15
45	70.87	16491168	13688117	377830000	491040000	8.87	58.54	780600000.00	615940000	8598635		84.32	82.98	74.25
46	71.46	16839695	13681375	381570000	424300000	8.87	57.49	732640000.00	581650000	4279354		92.02	81.43	71.13
47	72.03	17208426	13660920	401000000	480060000	8.80	56.56	1101530000.00	907810000	9236077		99.47	81.23	71.08
48	72.58	17585479	13640402	390480000	649070000	8.19	55.52	1219230000.00	937970000	2505022		109.02	81.39	71.04
49	73.10	17974187	13622668	324930000	787810000	7.75	54.48	1453360000.00	943870000	5330290		111.57	72.18	71.42
50	73.57	18378836	13611061	323130000	459050000	8.06	53.61	1045960000.00	987210000	10443719		114.36	75.38	71.22
51	74.00	18803424	13606215	291980000	416410000	8.42	52.99	985500000.00	819860000	7834479		113.67	71.32	70.20
52	74.38	19249027	13609796	171660000	632250000	8.09	52.35	1440240000.00	1251000000	8689334		115.76	67.64	69.43
53	74.72	19714269	13619520	279140000	448520000	8.63	51.97	1471140000.00	1374550000	5311219		117.02	67.14	68.69
54	75.03	20193387	13631382	260830000	705800000	8.41	51.78	2008650000.00	1695150000	9874543		113.91	70.76	69.36
55	75.31	20678703	13639379	27760000	768610000	8.41	51.70	2240150000.00	1586250000	6937687		117.38	72.27	69.75
56	75.57	21163552	13639770	232130000	417020000	9.76	51.64	1481430000.00	1033860000	11689100		116.20	70.21	68.65
57	75.82	21645836	13630950	253600000	842450000	9.81	51.81	1992340000.00	1569970000	3561589		118.63	73.45	71.20
58	76.06	22125659	13613921	4720000	9920000	9.69	51.89	1884890000.00	1533750000	9787246		119.35	73.72	72.06

Table 3.8 – Tableau de données de la Maroc

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	48.46	3619287.00	8709245.00	660000.00	45040000.00	5.06	91.16	59800000.00	59940000.00	474602.23	11.68	27.84	77.79	82.17
2	48.88	3793971.00	8916576.00	1460000.00	81540000.00	5.06	93.96	111330000.00	110570000.00	1537830.00	12.14	29.47	77.29	84.04
3	49.31	3973623.00	9121195.00	1400000.00	47900000.00	5.06	97.12	79790000.00	79090000.00	3511979.00	12.29	28.98	77.07	78.78
4	49.74	4157495.00	9320737.00	1730000.00	67320000.00	5.06	100.20	94510000.00	93480000.00	3856071.00	12.33	27.69	79.61	79.14
5	50.17	4344630.00	9512512.00	2240000.00	57510000.00	5.06	102.55	101380000.00	99840000.00	3403305.00	12.54	26.71	79.32	71.83
6	50.59	4533658.00	9695386.00	2480000.00	71650000.00	5.06	103.82	106610000.00	104130000.00	3486365.00	12.66	27.10	82.06	77.36
7	51.01	4724868.00	9868416.00	2168000.00	37990000.00	5.06	105.25	98210000.00	79960000.00	2179840.00	13.04	27.56	78.89	79.36
8	51.41	4918216.00	10032587.00	19970000.00	26940000.00	5.06	105.78	62210000.00	62550000.00	3523200.00	13.62	28.42	80.26	80.05
9	51.81	5114092.00	10188855.00	19150000.00	25230000.00	5.06	105.58	92690000.00	90290000.00	6352730.00	14.00	29.83	79.83	79.50
10	52.20	5313202.00	10338722.00	28090000.00	19360000.00	5.06	105.02	77630000.00	73260000.00	4089600.00	14.47	29.63	79.99	77.05
11	52.57	5516323.00	10483685.00	29000000.00	24250000.00	5.06	104.24	85560000.00	74090000.00	4251560.00	15.12	29.51	80.10	79.00
12	52.94	5723481.00	10623717.00	31050000.00	29980000.00	5.05	103.43	132520000.00	117640000.00	5352461.00	14.53	30.31	74.25	80.51
13	53.30	5955942.00	10739061.00	32960000.00	22780000.00	4.59	102.27	101450000.00	89180000.00	5167140.00	17.70	33.77	75.64	76.95
14	53.68	6197201.00	10851964.00	48710000.00	22740000.00	4.11	100.81	94160000.00	81400000.00	3228970.00	19.52	35.86	75.96	78.14
15	54.07	6449328.00	10967636.00	48410000.00	35800000.00	4.37	99.08	114750000.00	85280000.00	4809380.00	21.67	35.44	69.81	72.66
16	54.49	6714487.00	11089211.00	74090000.00	26040000.00	4.05	97.16	270110000.00	81400000.00	3728620.00	24.07	39.20	70.05	72.79
17	54.98	6993840.00	11216914.00	75020000.00	74340000.00	4.42	95.17	251510000.00	147410000.00	5713380.00	27.05	39.80	71.48	72.01
18	55.52	7286872.00	11350105.00	76950000.00	99030000.00	4.50	93.09	605690000.00	158670000.00	2888140.00	26.92	40.56	69.89	81.50
19	56.14	7594142.00	11487576.00	94140000.00	34940000.00	4.17	91.01	428860000.00	180460000.00	4744560.00	28.25	43.05	72.82	78.02
20	56.82	7915446.00	11627901.00	107620000.00	156620000.00	3.90	89.02	537380000.00	168930000.00	4103160.00	31.83	43.57	73.99	77.57
21	57.56	8250179.00	11769668.00	122550000.00	136240000.00	3.94	87.19	1160840000.00	199410000.00	4514720.00	32.78	35.03	69.42	76.19
22	58.34	8598463.00	11913138.00	105740000.00	109830000.00	5.17	85.99	1044100000.00	237770000.00	2130300.00	38.64	37.92	65.56	83.48
23	59.14	8960310.00	12056508.00	86530000.00	240090000.00	6.02	84.86	850540000.00	248850000.00	4916509.00	40.02	36.37	70.05	79.93
24	59.94	9331960.00	12196542.00	93120000.00	57290000.00	7.11	83.81	446780000.00	230360000.00	3581070.00	40.38	39.46	68.55	78.24
25	60.72	9710432.00	12327178.00	90610000.00	43110000.00	8.81	82.79	339770000.00	251530000.00	3759130.00	39.95	38.49	66.53	77.53
26	61.47	10092012.00	12445364.00	94540000.00	466930000.00	10.06	81.75	865860000.00	347280000.00	5312160.00	40.62	37.71	67.16	80.43
27	62.19	10475430.00	12548505.00	116310000.00	56560000.00	9.10	81.11	427780000.00	303280000.00	7824550.00	41.22	37.53	67.25	77.30
28	62.88	10860432.00	12637334.00	138790000.00	82000000.00	8.36	80.24	425790000.00	360730000.00	4336810.00	41.02	40.17	67.79	74.88
29	63.54	11248397.00	12713423.00	149220000.00	135820000.00	8.21	79.24	483260000.00	407180000.00	7958960.00	41.94	39.90	70.86	77.19
30	64.16	11640605.00	12780586.00	152140000.00	150690000.00	8.49	78.25	499120000.00	407340000.00	7428590.00	43.40	42.31	72.08	75.73
31	64.73	12039263.00	12839873.00	181120000.00	672980000.00	8.24	77.33	1241090000.00	624160000.00	6275864.00	44.97	46.15	67.04	68.36
32	65.25	12444707.00	12892155.00	189240000.00	863660000.00	8.71	76.19	1532670000.00	659330000.00	8664685.00	47.03	47.70	66.15	73.95
33	65.72	12855512.00	12935982.00	197120000.00	190260000.00	9.54	75.29	953600000.00	783460000.00	2950232.00	45.89	50.90	67.91	72.98
34	66.13	13267999.00	12969418.00	202490000.00	363610000.00	8.30	74.42	780000000.00	629330000.00	2817864.00	45.56	53.72	74.75	81.77
35	66.50	13678862.00	12988186.00	200400000.00	333550000.00	9.20	73.30	683520000.00	517000000.00	9638630.00	44.94	53.18	71.88	69.49
36	66.85	13995729.00	13079503.00	263320000.00	107860000.00	8.54	71.83	500560000.00	454270000.00	1783230.00	45.96	55.96	71.30	66.74
37	67.18	14285555.00	13175048.00	246830000.00	357720000.00	8.72	70.47	649590000.00	608690000.00	10103304.00	45.07	52.84	68.94	67.90
38	67.53	14566859.00	13259042.00	214110000.00	289400000.00	9.53	68.72	471560000.00	414430000.00	4097671.00	48.14	61.93	67.36	66.42
39	67.89	14842165.00	13333098.00	226940000.00	355620000.00	9.60	66.78	531210000.00	486370000.00	6629516.00	47.85	60.99	68.88	70.30
40	68.28	15114839.00	13399959.00	229570000.00	321600000.00	9.60	64.95	681940000.00	637350000.00	3843292.00	49.02	66.14	80.96	79.73
41	68.72	15386945.00	13462676.00	192090000.00	219330000.00	10.63	63.38	447380000.00	410390000.00	1996649.00	45.14	70.91	85.90	77.43
42	69.21	15659555.00	13522277.00	210880000.00	185540000.00	11.30	61.77	493850000.00	474210000.00	4603350.00	46.64	77.54	84.40	71.77
43	69.74	15933727.00	13578641.00	282930000.00	172260000.00	11.02	60.53	364390000.00	344190000.00	5287815.00	47.99	81.96	84.96	75.67
44	70.30	16210331.00	13633606.00	349990000.00	276470000.00	9.57	59.52	584390000.00	483500000.00	7969118.00	49.57	82.55	85.40	78.15
45	70.87	16491168.00	13688117.00	377830000.00	491040000.00	8.87	58.54	780600000.00	615940000.00	8598635.00	50.73	84.32	82.98	74.25
46	71.46	16839695.00	13681375.00	381570000.00	424300000.00	8.87	57.49	732640000.00	581650000.00	4279354.00	51.89	92.02	81.43	71.13
47	72.03	17208426.00	13660920.00	401000000.00	480060000.00	8.80	56.56	1101530000.00	907810000.00	9236077.00	54.50	99.47	81.23	71.08
48	72.58	17585479.00	13640402.00	390480000.00	649070000.00	8.19	55.52	1219230000.00	937970000.00	2505022.00	54.92	109.02	81.39	71.04
49	73.10	17974187.00	13622668.00	324930000.00	787810000.00	7.75	54.48	1453360000.00	943870000.00	5330290.00	58.11	111.57	72.18	71.42
50	73.57	18378836.00	13611061.00	323130000.00	459050000.00	8.06	53.61	1045960000.00	987210000.00	10443719.00	56.65	114.36	75.38	71.22
51	74.00	18803424.00	13606215.00	291980000.00	416410000.00	8.42	52.99	985500000.00	819860000.00	7834479.00	57.99	113.67	71.32	70.20
52	74.38	19249027.00	13609796.00	171660000.00	632250000.00	8.09	52.35	1440240000.00	1251000000.00	8689334.00	59.50	115.76	67.64	69.43
53	74.72	19714269.00	13619520.00	279140000.00	448520000.00	8.63	51.97	1471140000.00	1374550000.00	5311219.00	63.64	117.02	67.14	68.69
54	75.03	20193387.00	13631382.00	260830000.00	705800000.00	8.41	51.78	2008650000.00	1695150000.00	9874543.00	63.59	113.91	70.76	69.36
55	75.31	20678703.00	13639379.00	277760000.00	768610000.00	8.41	51.70	2240150000.00	1586250000.00	6937687.00	64.44	117.38	72.27	69.75
56	75.57	21163552.00	13639770.00	232130000.00	417020000.00	9.76	51.64	1481430000.00	1033860000.00	11689100.00	62.12	116.20	70.21	68.65
57	75.82	21645836.00	13630950.00	253600000.00	842450000.00	9.81	51.81	1992340000.00	1569970000.00	3561589.00	64.21	118.63	73.45	71.20
58	76.06	22125659.00	13613921.00	4720000.00	9920000.00	9.69	51.89	1884890000.00	1533750000.00	9787246.00	62.41	119.35	73.72	72.06

Table 3.9 – Données après estimation des valeurs manquantes Maroc

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	42.02	1566559	2609707	630000	49200000	0.42	89.42	51830000	51200000				80.42	87.07
2	42.66	1607538	2628399	2050000	90150000	0.42	89.78	93820000	92170000	561618	33.23		79.78	87.37
3	43.35	1652015	2651116	1020000	66000000	0.42	90.95	70110000	69590000	814367	34.37		84.79	88.71
4	44.11	1699968	2677669	1620000	50840000	0.42	93.07	83560000	81840000	1193300	37.40		80.30	85.30
5	44.93	1751298	2707313	1810000	44660000	0.45	95.76	71070000	68460000	922000	39.08		76.32	79.40
6	45.82	1805590	2739749	2150000	55500000	0.52	98.32	90920000	86870000	1141500	44.03	30.68	69.14	81.23
7	46.78	1865978	2772297	20630000	23780000	0.52	99.58	82840000	62610000	813000	50.31	33.03	71.47	76.83
8	47.80	1944133	2793494	20850000	22840000	0.52	99.36	89520000	87160000	683000	57.02	33.69	68.56	80.45
9	48.88	2026350	2815817	19470000	19780000	0.52	98.25	71860000	69130000	812600	59.75	33.99	67.90	85.40
10	50.00	2111834	2838319	25060000	28130000	0.52	97.00	115730000	106200000	632400	62.77	33.94	70.35	83.19
11	51.15	2200362	2860035	26300000	42550000	0.52	95.98	112850000	99810000	704400	65.77	33.20	69.79	84.29
12	52.30	2291761	2880930	29050000	25150000	0.52	94.48	106430000	91550000	835800	68.75	34.49	70.72	83.74
13	53.45	2386374	2901169	32260000	20890000	0.48	93.25	113350000	102290000	1231300	72.76	33.22	81.18	84.14
14	54.59	2484247	2921108	36480000	31330000	0.42	92.20	134040000	117290000	1128200	71.43	37.02	77.69	83.02
15	55.70	2585973	2940791	35580000	62870000	0.44	91.08	174210000	131800000	1108600	71.26	35.30	75.01	79.61
16	56.80	2689900	2962576	53480000	54680000	0.40	89.83	214270000	114890000	1271100	70.97	38.35	75.67	82.89
17	57.87	2785901	2995895	50610000	49130000	0.43	88.67	217860000	150500000	1078300	70.40	39.95	82.02	83.24
18	58.93	2885281	3029725	51710000	31640000	0.43	87.50	247080000	164560000	880300	70.31	39.97	80.77	86.72
19	59.98	2989612	3065299	59630000	28090000	0.42	86.31	310500000	253090000	959000	73.84	42.26	79.88	84.46
20	61.01	3100869	3104343	65590000	30170000	0.41	85.12	209850000	155320000	979500	74.87	41.73	84.23	83.05
21	62.00	3220318	3147849	71860000	25530000	0.40	83.95	240480000	158730000	1196800	80.85	40.83	84.68	87.96
22	62.92	3348696	3196328	58160000	26470000	0.49	82.68	242860000	179980000	1262000	82.65	42.62	79.55	88.45
23	63.77	3485498	3248463	38020000	29490000	0.59	81.31	207330000	154440000	1285000	84.18	44.26	79.40	87.29
24	64.53	3628473	3301914	50440000	28190000	0.68	79.93	202760000	170840000	951000	82.80	43.67	80.27	83.94
25	65.21	3778593	3349348	55150000	34910000	0.78	78.66	176100000	153110000	1053420	83.33	43.11	82.36	81.24
26	65.83	3942025	3379851	54150000	29830000	0.83	77.57	160420000	130670000	2098000	85.02	45.06	77.68	80.87
27	66.42	4105358	3404398	66020000	69800000	0.79	76.49	220260000	163030000	782000	86.14	46.33	70.95	82.64
28	66.99	4268509	3423745	87760000	86500000	0.83	75.57	274260000	230450000	1919500	87.25	47.45	75.77	83.43
29	67.58	4432733	3438726	83380000	124140000	0.86	74.71	314830000	284420000	450000	88.56	51.74	72.08	79.87
30	68.18	4599578	3451354	82130000	88160000	0.95	73.76	279890000	212040000	655600	86.36	53.95	71.99	73.89
31	68.79	4770577	3462220	99910000	175650000	0.88	72.61	392540000	237640000	1654847	82.44	51.50	72.43	74.03
32	69.40	4945894	3471790	89740000	99680000	0.92	71.41	354460000	309610000	2575808	87.87	48.99	66.90	78.09
33	69.99	5124253	3478972	97760000	140400000	0.88	69.90	387990000	344100000	2219391	94.48	46.58	73.42	81.61
34	70.54	5302734	3482154	101870000	103190000	1.00	68.22	243200000	191890000	1939795	95.25	46.21	75.31	83.40
35	71.04	5470510	3486086	84150000	84250000	1.01	66.50	105790000	109550000	679200	111.13	46.30	77.70	82.22
36	71.50	5602725	3511250	121170000	59520000	0.95	64.85	75390000	86970000	647471	118.02	45.76	77.37	78.44
37	71.90	5726710	3529327	102160000	138670000	0.97	63.21	126860000	158580000	2896345	117.40	46.39	76.91	80.81
38	72.26	5842761	3541391	91760000	137510000	1.11	61.69	195220000	208280000	1084468	116.87	44.91	81.31	81.02
39	72.59	5951656	3547739	93370000	65630000	1.14	60.20	150960000	177130000	1697420	114.97	43.76	83.79	86.22
40	72.88	6054487	3549255	96440000	109180000	1.19	58.61	253430000	267810000	1838970	113.10	47.46	84.25	86.54
41	73.16	6152395	3546802	95650000	44650000	1.37	56.90	222840000	232950000	1121989	114.18	50.09	81.62	83.52
42	73.40	6245039	3540662	90060000	94640000	1.44	55.28	366220000	375500000	1391500	117.34	54.96	82.93	82.94
43	73.63	6333095	3531231	98020000	115000000	1.42	53.59	217350000	221810000	550526	116.54	54.68	82.96	82.36
44	73.84	6419541	3520137	129890000	100220000	1.29	51.91	301670000	302710000	2318390	119.53	53.27	84.86	82.25
45	74.03	6506432	3511169	141580000	143660000	1.25	50.35	333690000	321960000	2164100	125.34	52.48	88.72	78.78
46	74.20	6590556	3511926	145270000	140900000	1.30	48.98	368350000	364820000	2135246	136.16	54.11	86.00	77.57
47	74.35	6680916	3515220	149600000	190230000	1.33	47.71	436850000	435410000	1647565	140.73	55.34	85.42	75.23
48	74.48	6777068	3521019	163980000	174450000	1.28	46.61	326490000	323770000	2024804	148.73	57.59	85.51	74.82
49	74.59	6878521	3528815	156540000	122240000	1.23	45.70	378760000	333660000	2023940	142.64	58.87	79.94	67.28
50	74.69	6983973	3537861	162790000	144640000	1.35	45.00	493690000	457660000	2567679		62.58	79.91	72.76
51	74.79	7092259	3547672	162170000	141490000	1.43	44.53	550150000	447670000	1113434		65.08	81.07	73.55
52	74.90	7203511	3557956	102530000	332140000	1.41	44.35	925230000	934180000	2343806		69.73	83.30	70.52
53	75.02	7317800	3568868	164690000	436890000	1.56	44.42	1022130000	934540000	2305900		69.07	82.17	66.68
54	75.17	7434276	3580282	167670000	253800000	1.62	44.69	714660000	641480000	1328220		68.71	82.03	67.15
55	75.33	7552449	3591459	171190000	367620000	1.70	45.09	922700000	672560000	2349967		68.92	81.98	62.16
56	75.53	7672403	3601258	148200000	274890000	1.96	45.56	474540000	512500000	1339633		69.40	82.51	64.45
57	75.73	7793664	3609584	167090000	233580000	2.15	46.38	627440000	615020000	1324275		70.68	81.63	65.98
58	75.94	7915883	3616244	3700000	6600000	2.42	47.08	775770000	726790000	1642301		73.52	83.67	63.37

Table 3.10 – Tableau de données de la Tunisie

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	42.02	1566559.00	2609707.00	630000.00	49200000.00	0.42	89.42	51830000.00	51200000.00	1004876.12	45.27	26.16	80.42	87.07
2	42.66	1607538.00	2628399.00	2050000.00	90150000.00	0.42	89.78	93820000.00	92170000.00	561618.00	33.23	27.33	79.78	87.37
3	43.35	1652015.00	2651116.00	1020000.00	66000000.00	0.42	90.95	70110000.00	69590000.00	814367.00	34.37	26.52	84.79	88.71
4	44.11	1699968.00	2677669.00	1620000.00	50840000.00	0.42	93.07	83560000.00	81840000.00	1193300.00	37.40	28.18	80.30	85.30
5	44.93	1751298.00	2707313.00	1810000.00	44660000.00	0.45	95.76	71070000.00	68460000.00	922000.00	39.08	31.02	76.32	79.40
6	45.82	1805590.00	2739749.00	2150000.00	55500000.00	0.52	98.32	90920000.00	86870000.00	1141500.00	44.03	30.68	69.14	81.23
7	46.78	1865978.00	2772297.00	20630000.00	23780000.00	0.52	99.58	82840000.00	62610000.00	813000.00	50.31	33.03	71.47	76.83
8	47.80	1944133.00	2793494.00	20850000.00	22840000.00	0.52	99.36	89520000.00	87160000.00	683000.00	57.02	33.69	68.56	80.45
9	48.88	2026350.00	2815817.00	19470000.00	19780000.00	0.52	98.25	71860000.00	69130000.00	812600.00	59.75	33.99	67.90	85.40
10	50.00	2111834.00	2838319.00	25060000.00	28130000.00	0.52	97.00	115730000.00	106200000.00	632400.00	62.77	33.94	70.35	83.19
11	51.15	2200362.00	2860035.00	26300000.00	42550000.00	0.52	95.98	112850000.00	99810000.00	704400.00	65.77	33.20	69.79	84.29
12	52.30	2291761.00	2880930.00	29050000.00	25150000.00	0.52	94.48	106430000.00	91550000.00	835800.00	68.75	34.49	70.72	83.74
13	53.45	2386374.00	2901169.00	32260000.00	20890000.00	0.48	93.25	113350000.00	102290000.00	1231300.00	72.76	33.22	81.18	84.14
14	54.59	2484247.00	2921108.00	36480000.00	31330000.00	0.42	92.20	134040000.00	117290000.00	1128200.00	71.43	37.02	77.69	83.02
15	55.70	2585973.00	2940791.00	35580000.00	62870000.00	0.44	91.08	174210000.00	131800000.00	1108600.00	71.26	35.30	75.01	79.61
16	56.80	2689900.00	2962576.00	53480000.00	54680000.00	0.40	89.83	214270000.00	114890000.00	1271100.00	70.97	38.35	75.67	82.89
17	57.87	2785901.00	2995895.00	50610000.00	49130000.00	0.43	88.67	217860000.00	150500000.00	1078300.00	70.40	39.95	82.02	83.24
18	58.93	2885281.00	3029725.00	51710000.00	31640000.00	0.43	87.50	247080000.00	164560000.00	880300.00	70.31	39.97	80.77	86.72
19	59.98	2989612.00	3065299.00	59630000.00	28090000.00	0.42	86.31	310500000.00	253090000.00	959000.00	73.84	42.26	79.88	84.46
20	61.01	3100869.00	3104343.00	65590000.00	30170000.00	0.41	85.12	209850000.00	155320000.00	979500.00	74.87	41.73	84.23	83.05
21	62.00	3220318.00	3147849.00	71860000.00	25530000.00	0.40	83.95	240480000.00	158730000.00	1196800.00	80.85	40.83	84.68	87.96
22	62.92	3348696.00	3196328.00	58160000.00	26470000.00	0.49	82.68	242860000.00	179980000.00	1262000.00	82.65	42.62	79.55	88.45
23	63.77	3485498.00	3248463.00	38020000.00	29490000.00	0.59	81.31	207330000.00	154440000.00	1285000.00	84.18	44.26	79.40	87.29
24	64.53	3628473.00	3301914.00	50440000.00	28190000.00	0.68	79.93	202760000.00	170840000.00	951000.00	82.80	43.67	80.27	83.94
25	65.21	3778593.00	3349348.00	55150000.00	34910000.00	0.78	78.66	176100000.00	153110000.00	1053420.00	83.33	43.11	62.36	81.24
26	65.83	3942025.00	3379851.00	54150000.00	29830000.00	0.83	77.57	160420000.00	130670000.00	2098000.00	85.02	45.06	77.68	80.87
27	66.42	4105358.00	3404398.00	66020000.00	69800000.00	0.79	76.49	220260000.00	163030000.00	782000.00	86.14	46.33	70.95	82.64
28	66.99	4268509.00	3423745.00	87760000.00	86500000.00	0.83	75.57	274260000.00	230450000.00	1919500.00	87.25	47.45	75.77	83.43
29	67.58	4432733.00	3438726.00	83380000.00	124140000.00	0.86	74.71	314830000.00	284420000.00	450000.00	88.56	51.74	72.08	79.87
30	68.18	4599578.00	3451354.00	82130000.00	88160000.00	0.95	73.76	279890000.00	212040000.00	655600.00	86.36	53.95	71.99	73.89
31	68.79	4770577.00	3462220.00	99910000.00	175650000.00	0.88	72.61	392540000.00	237640000.00	1654847.00	82.44	51.50	72.43	74.03
32	69.40	4945894.00	3471790.00	89740000.00	99680000.00	0.92	71.41	354460000.00	309610000.00	2575808.00	87.87	48.99	66.90	78.09
33	69.99	5124253.00	3478972.00	97760000.00	140400000.00	0.88	69.90	387990000.00	344100000.00	2219391.00	94.48	46.58	73.42	81.61
34	70.54	5302734.00	3482154.00	101870000.00	103190000.00	1.00	68.22	243200000.00	191890000.00	1939795.00	95.25	46.21	75.31	83.40
35	71.04	5470510.00	3486086.00	84150000.00	84250000.00	1.01	66.50	105790000.00	109550000.00	679200.00	111.13	46.30	77.70	82.22
36	71.50	5602725.00	3511250.00	121170000.00	59520000.00	0.95	64.85	75390000.00	86970000.00	647471.00	118.02	45.76	77.37	78.44
37	71.90	5726710.00	3529327.00	102160000.00	138670000.00	0.97	63.21	126860000.00	158580000.00	2896345.00	117.40	46.39	76.91	80.81
38	72.26	5842761.00	3541391.00	91760000.00	137510000.00	1.11	61.69	195220000.00	208280000.00	1084468.00	116.87	44.91	81.31	81.02
39	72.59	5951656.00	3547739.00	93370000.00	65630000.00	1.14	60.20	150960000.00	177130000.00	1697420.00	114.97	43.76	83.79	86.22
40	72.88	6054487.00	3549255.00	96440000.00	109180000.00	1.19	58.61	253430000.00	267810000.00	1838970.00	113.10	47.46	84.25	86.54
41	73.16	6152395.00	3546802.00	95650000.00	44650000.00	1.37	56.90	222840000.00	232950000.00	1121989.00	114.18	50.09	81.62	83.52
42	73.40	6245039.00	3540662.00	90060000.00	94640000.00	1.44	55.28	366220000.00	375500000.00	1391500.00	117.34	54.96	82.93	82.94
43	73.63	6333095.00	3531231.00	98020000.00	115000000.00	1.42	53.59	217350000.00	221810000.00	550526.00	116.54	54.68	82.96	82.36
44	73.84	6419541.00	3520137.00	129890000.00	100220000.00	1.29	51.91	301670000.00	302710000.00	2318390.00	119.53	53.27	84.86	82.25
45	74.03	6506432.00	3511169.00	141580000.00	143660000.00	1.25	50.35	333690000.00	321960000.00	2164100.00	125.34	52.48	88.72	78.78
46	74.20	6590556.00	3511926.00	145270000.00	140900000.00	1.30	48.98	368350000.00	364820000.00	2135246.00	136.16	54.11	86.00	77.57
47	74.35	6680916.00	3515220.00	149600000.00	190230000.00	1.33	47.71	436850000.00	435410000.00	1647565.00	140.73	55.34	85.42	75.23
48	74.48	6777068.00	3521019.00	163980000.00	174450000.00	1.28	46.61	326490000.00	323770000.00	2024804.00	148.73	57.59	85.51	74.82
49	74.59	6878521.00	3528815.00	156540000.00	122240000.00	1.23	45.70	378760000.00	333660000.00	1223940.00	142.64	58.87	79.94	67.28
50	74.69	6983973.00	3537861.00	162790000.00	144640000.00	1.35	45.00	493690000.00	457660000.00	2567679.00	141.56	62.58	79.91	72.76
51	74.79	7092259.00	3547672.00	162170000.00	141490000.00	1.43	44.53	550150000.00	447670000.00	1113434.00	143.35	65.08	81.07	73.55
52	74.90	7203511.00	3557956.00	102530000.00	332140000.00	1.41	44.35	925230000.00	934180000.00	2343806.00	84.10	69.73	83.30	70.52
53	75.02	7317800.00	3568868.00	164690000.00	438690000.00	1.56	44.42	1022130000.00	934540000.00	2305900.00	99.12	69.07	82.17	66.68
54	75.17	7434276.00	3580282.00	167670000.00	253800000.00	1.62	44.69	714660000.00	641480000.00	1328220.00	130.36	68.71	82.03	67.15
55	75.33	7552449.00	3591459.00	171190000.00	367620000.00	1.70	45.09	922700000.00	672560000.00	2349967.00	115.05	68.92	81.98	62.16
56	75.53	7672403.00	3601258.00	148200000.00	274890000.00	1.96	45.56	474540000.00	512500000.00	1339633.00	128.63	69.40	82.51	64.45
57	75.73	7793664.00	3609584.00	167090000.00	233580000.00	2.15	46.38	627440000.00	615020000.00	1324275.00	138.11	70.68	81.63	65.98
58	75.94	7915883.00	3616244.00	3700000.00	660000.00	2.42	47.08	775770000.00	726790000.00	1642301.00	86.37	73.52	83.67	63.37

Table 3.11 – Données après estimation des valeurs manquantes Tunisie

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	46.14	3394203	7730685	-50000	334700000	4.94	91.43	355050000	355100000				94.34	92.79
2	46.60	3626403	7778456	10000	400900000	4.94	93.59	437910000	437900000	937596	41.72		94.05	92.27
3	47.05	3882767	7807386	1320000	353310000	4.94	96.37	397630000	396310000	2369800	43.33			
4	47.51	4154288	7830848	5050000	248660000	4.94	99.27	281910000	283470000	2324215	44.35			
5	47.96	4443887	7852083	3020000	203020000	4.94	101.51	227840000	228130000	1487506	45.66	33.20		87.28
6	48.39	4753164	7873788	2220000	134910000	4.94	102.73	146230000	144230000	1735461	45.62	34.14		
7	48.81	5041536	7938731	45930000	78050000	4.94	103.87	126920000	129140000	776122	48.22	38.51	93.27	91.58
8	49.21	5208671	8145526	45620000	65510000	4.94	103.83	107640000	109860000	1641700	62.69	45.10	89.49	83.56
9	49.59	5383539	8360848	43190000	76130000	4.94	103.06	119610000	115480000	2128233	63.75	53.42	85.03	86.89
10	49.98	5563573	8580865	48440000	63120000	4.94	102.19	131230000	115940000	1852429	64.17	55.29	82.37	82.95
11	50.37	5747263	8802771	43730000	55620000	4.94	101.51	122210000	112860000	2058094	64.34	54.45	82.11	81.18
12	50.77	5933927	9026182	39890000	42080000	4.91	101.56	120990000	110390000	1735448	63.48	55.83	73.56	82.05
13	51.20	6124850	9252243	46830000	35960000	4.48	101.71	110300000	98570000	2362625	62.52	59.83	78.63	83.35
14	51.67	6321297	9483131	58120000	48770000	3.96	101.84	125090000	94290000	1595994	62.96	59.01	89.94	81.61
15	52.22	6525328	9721785	55260000	53110000	4.18	101.73	128330000	111370000	1480275	63.11	46.66	83.24	83.14
16	52.86	6738780	9970319	70290000	83400000	3.95	101.31	235270000	106890000	2680452	60.63	54.92	87.68	88.11
17	53.66	6961531	10228708	77000000	21060000	4.16	101.28	145370000	128820000	2313186	58.94	59.07	93.10	86.81
18	54.61	7240239	10449945	74690000	23350000	4.15	100.88	129150000	112730000	1142509	60.83	59.74	93.91	84.37
19	55.70	7611660	10600666	80690000	70500000	3.97	100.25	190290000	119470000	1538550	61.15	64.52	89.18	82.84
20	56.91	8004279	10756482	87830000	20720000	3.85	99.55	110840000	97660000	1620808	66.97	62.21	94.74	83.49
21	58.20	8420028	10917687	95900000	100630000	3.84	98.87	205160000	118280000	2419074	68.36	57.56	94.65	83.31
22	59.52	8858976	11084688	80550000	36670000	4.32	98.49	167070000	151130000	1832872	69.72	57.03	90.91	83.69
23	60.82	9321410	11254291	60840000	21570000	4.59	97.92	134700000	131690000	1524398	71.50	66.42	94.23	82.55
24	62.05	9805134	11423155	51510000	17690000	4.79	97.19	93280000	88520000	1290834	76.27	71.00	88.88	81.80
25	63.16	10307407	11586446	50010000	17100000	4.98	96.27	120980000	115000000	1461484	88.68	72.76	91.60	81.27
26	64.11	10824413	11741492	43830000	40580000	5.03	95.13	172150000	148940000	2918778	108.99	76.85	89.62	84.09
27	64.90	11355685	11885587	55040000	54800000	4.70	94.07	167330000	122630000	2403612	118.63	75.80	86.60	84.51
28	65.53	11892457	12025440	71770000	84480000	4.85	92.77	213450000	144900000	2066344	126.29	79.67	86.07	78.32
29	66.03	12421409	12170083	68620000	55800000	5.91	91.25	171980000	126200000	1038303	129.14	83.82	85.01	80.27
30	66.42	12956681	12300991	62500000	68460000	7.61	89.55	158210000	116690000	2006130	111.95	72.80	85.20	78.59
31	66.72	13496456	12415911	105700000	45310000	8.96	87.68	331650000	112230000	1627035	129.11	61.77	84.31	84.53
32	66.98	14039539	12514790	112570000	73200000	18.47	85.81	211170000	192680000	3809857	135.20	49.11	86.68	84.31
33	67.23	14584560	12596534	108800000	74280000	21.84	83.77	374790000	252060000	3330382	121.11	51.94	84.69	85.39
34	67.48	15126562	12659697	121760000	98820000	23.35	81.56	326310000	187870000	1453454	128.16	50.10	84.44	86.13
35	67.77	15661352	12700901	96070000	72740000	35.06	79.14	295490000	285110000	964744	121.13	45.32	85.78	82.57
36	68.11	16185540	12718758	97340000	73450000	47.66	76.52	295740000	293710000	2139705	121.30	37.17	84.99	82.67
37	68.50	16698037	12713378	122890000	90000000	54.75	73.96	306880000	288710000	4901678	122.35	33.01	79.33	81.88
38	68.93	17198083	12688756	109070000	33860000	57.71	71.21	247890000	226570000	870017	120.82	36.08	85.61	82.71
39	69.37	17689675	12646057	106300000	107500000	58.74	68.36	419870000	393260000	3026097	121.13	42.38	85.10	83.81
40	69.82	18188015	12577598	92680000	48290000	66.57	65.55	137980000	97380000	2021053	122.85	42.21	83.29	82.76
41	70.29	18684937	12498723	76250000	61180000	75.26	62.89	199960000	118230000	934656	124.92	37.83	84.38	82.86
42	70.78	19180228	12411925	77270000	42560000	77.22	60.09	198860000	161250000	2659595	128.15	56.85	84.08	83.20
43	71.28	19677273	12317773	96940000	93340000	79.68	57.57	186120000	173380000	1953325	129.46	62.72	84.39	79.69
44	71.78	20182205	12221309	146190000	72570000	77.39	55.32	239240000	225110000	4266387	130.34	62.82	86.41	75.76
45	72.28	20703617	12127479	189370000	101410000	72.06	53.30	316970000	306750000	4033242	130.53	59.27	89.59	76.54
46	72.77	21248009	12040428	197990000	125320000	73.28	51.57	347380000	323720000	3527824	133.31	53.83	84.32	74.44
47	73.22	21818169	11959746	199770000	149690000	72.65	50.29	240680000	192810000	4018105	137.03	57.28	87.67	74.62
48	73.65	22414414	11885662	228250000	162300000	69.29	49.36	395010000	377540000	3602256	138.65	64.09	85.96	75.19
49	74.03	23041887	11818828	177170000	142270000	64.58	48.76	324830000	333030000	1536002	139.58	62.99	86.38	72.32
50	74.37	23700349	11765411	191310000	118650000	72.65	48.49	302390000	283240000	5253472		73.16	84.67	69.02
51	74.68	24393852	11723785	186240000	74560000	74.39	48.54	201250000	195040000	4211370		69.05	82.16	71.68
52	74.94	25124194	11695364	74320000	217380000	72.94	49.12	193350000	186510000	4247550		68.06	81.51	70.89
53	75.19	25888503	11677344	168570000	81960000	77.54	49.78	147320000	144050000	5137470		67.95	80.66	67.76
54	75.42	26674438	11664124	168050000	99280000	79.37	50.60	203420000	149370000	4912551		71.73	80.30	68.03
55	75.64	27465760	11647553	158450000	80490000	80.58	51.60	160620000	141900000	3435535		79.31	77.85	68.52
56	75.86	28248180	11623348	126090000	66860000	100.69	52.72	87490000	80980000	3761224		82.00	79.18	65.60
57	76.08	29016679	11589373	144290000	69520000	109.44	53.91	157930000	147620000	3445158		78.84	77.90	62.98
58	76.29	29770548	11547594	560000	3240000	110.97	55.03	189190000	174150000	3478073		80.53	75.50	59.56

Table 3.12 – Tableau de données de l'Algérie

	indicateur	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1960.00	46.14	3394203.00	7730685.00	-50000.00	334700000.00	4.94	91.43	355050000.00	355100000.00	2121058.84	35.88	8.98	94.34	92.79
2	1961.00	46.60	3626403.00	7778456.00	10000.00	400900000.00	4.94	93.59	437910000.00	437900000.00	937596.00	41.72	-5.41	94.05	92.27
3	1962.00	47.05	3882767.00	7807386.00	1320000.00	353310000.00	4.94	96.37	397630000.00	396310000.00	2369800.00	43.33	2.91	89.58	92.04
4	1963.00	47.51	4154288.00	7830848.00	5050000.00	248660000.00	4.94	99.27	281910000.00	283470000.00	2324215.00	44.35	21.80	88.15	88.78
5	1964.00	47.96	4443887.00	7852083.00	3020000.00	203020000.00	4.94	101.51	227840000.00	228130000.00	1487506.00	45.66	33.20	86.95	87.28
6	1965.00	48.39	4753164.00	7873788.00	2220000.00	134910000.00	4.94	102.73	146230000.00	144230000.00	1735461.00	45.62	34.14	85.47	85.98
7	1966.00	48.81	5041536.00	7938731.00	45930000.00	78050000.00	4.94	103.87	126920000.00	129140000.00	776122.00	48.22	38.51	93.27	91.58
8	1967.00	49.21	5208671.00	8145526.00	45620000.00	65510000.00	4.94	103.83	107640000.00	109860000.00	1641700.00	62.69	45.10	89.49	83.56
9	1968.00	49.59	5383539.00	8360848.00	43190000.00	76130000.00	4.94	103.06	119610000.00	115480000.00	2128233.00	63.75	53.42	85.03	86.89
10	1969.00	49.98	5563573.00	8580865.00	48440000.00	63120000.00	4.94	102.19	131230000.00	115940000.00	1852429.00	64.17	55.29	82.37	82.95
11	1970.00	50.37	5747263.00	8802771.00	43730000.00	55620000.00	4.94	101.51	122210000.00	112860000.00	2058094.00	64.34	54.45	82.11	81.18
12	1971.00	50.77	5933927.00	9026182.00	39890000.00	42080000.00	4.91	101.56	120990000.00	110390000.00	1735448.00	63.48	55.83	73.56	82.05
13	1972.00	51.20	6124850.00	9252243.00	46830000.00	35960000.00	4.48	101.71	110300000.00	98570000.00	2362625.00	62.52	59.83	78.63	83.35
14	1973.00	51.67	6321297.00	9483131.00	58120000.00	48770000.00	3.96	101.84	125090000.00	94290000.00	1595994.00	62.96	59.01	89.94	81.61
15	1974.00	52.22	6525328.00	9721785.00	55260000.00	53110000.00	4.18	101.73	128330000.00	111370000.00	1480275.00	63.11	46.66	83.24	83.14
16	1975.00	52.86	6738780.00	9970319.00	70290000.00	83400000.00	3.95	101.31	126270000.00	106890000.00	2680452.00	60.63	54.92	87.68	88.11
17	1976.00	53.66	6961531.00	10228708.00	77000000.00	21060000.00	4.16	101.28	145370000.00	128820000.00	2313186.00	58.94	59.07	93.10	86.81
18	1977.00	54.61	7240239.00	10449945.00	74690000.00	23350000.00	4.15	100.88	129150000.00	112730000.00	1142509.00	60.83	59.74	93.91	84.37
19	1978.00	55.70	7611660.00	10600666.00	80690000.00	70500000.00	3.97	100.25	190290000.00	119470000.00	1538550.00	61.15	64.52	89.18	82.84
20	1979.00	56.91	8004279.00	10756482.00	87830000.00	20720000.00	3.85	99.55	110840000.00	97660000.00	1620808.00	66.97	62.21	94.74	83.49
21	1980.00	58.20	8420028.00	10917687.00	95900000.00	100630000.00	3.84	98.87	205160000.00	118280000.00	2419074.00	68.36	57.56	94.65	83.31
22	1981.00	59.52	8858976.00	11084688.00	80550000.00	36670000.00	4.32	98.49	167070000.00	151130000.00	1832872.00	69.72	57.03	90.91	83.69
23	1982.00	60.82	9321410.00	11254291.00	60840000.00	21570000.00	4.59	97.92	134700000.00	131690000.00	1524398.00	71.50	66.42	94.23	82.55
24	1983.00	62.05	9805134.00	11423155.00	51510000.00	17690000.00	4.79	97.19	93280000.00	88520000.00	1290834.00	76.27	71.00	88.88	81.80
25	1984.00	63.16	10307407.00	11586446.00	50010000.00	17100000.00	4.98	96.27	120980000.00	115000000.00	1461484.00	88.68	72.76	91.60	81.27
26	1985.00	64.11	10824413.00	11741492.00	43830000.00	40580000.00	5.03	95.13	172150000.00	148940000.00	2918778.00	108.99	76.85	89.62	84.09
27	1986.00	64.90	11355685.00	11885587.00	55040000.00	54800000.00	4.70	94.07	167330000.00	122630000.00	2403612.00	118.63	75.80	86.60	84.51
28	1987.00	65.53	11892457.00	12025440.00	71770000.00	84480000.00	4.85	92.77	213450000.00	144900000.00	2066344.00	126.29	79.67	86.07	78.32
29	1988.00	66.03	12421409.00	12170083.00	68620000.00	55800000.00	5.91	91.25	171980000.00	126200000.00	1038303.00	129.14	83.82	85.01	80.27
30	1989.00	66.42	12956681.00	12300991.00	62500000.00	68460000.00	7.61	89.55	158210000.00	116690000.00	2006130.00	111.95	72.80	85.20	78.59
31	1990.00	66.72	13496456.00	12415911.00	105700000.00	45310000.00	8.96	87.68	331650000.00	112230000.00	1627035.00	129.11	61.77	84.31	84.53
32	1991.00	66.98	14039539.00	12514790.00	112570000.00	73200000.00	18.47	85.81	211170000.00	192680000.00	3809857.00	135.20	49.11	86.68	84.31
33	1992.00	67.23	14584560.00	12596534.00	108800000.00	74280000.00	21.84	83.77	374790000.00	252060000.00	3330382.00	121.11	51.94	84.69	85.39
34	1993.00	67.48	15126562.00	12659697.00	121760000.00	98820000.00	23.35	81.56	326310000.00	187870000.00	1453454.00	128.16	50.10	84.44	86.13
35	1994.00	67.77	15661352.00	12700901.00	96070000.00	72740000.00	35.06	79.14	295490000.00	285110000.00	964744.00	121.13	45.32	85.78	82.57
36	1995.00	68.11	16185540.00	12718758.00	97340000.00	73450000.00	47.66	76.52	295740000.00	293710000.00	2139705.00	121.30	37.17	84.99	82.67
37	1996.00	68.50	16698037.00	12713378.00	122890000.00	90000000.00	54.75	73.96	306880000.00	288710000.00	4901678.00	122.35	33.01	79.33	81.88
38	1997.00	68.93	17198083.00	12688756.00	109070000.00	33860000.00	57.71	71.21	247890000.00	226570000.00	870017.00	120.82	36.08	85.61	82.71
39	1998.00	69.37	17689675.00	12646057.00	106300000.00	107500000.00	58.74	68.36	419870000.00	393260000.00	3026097.00	121.13	42.38	85.10	83.81
40	1999.00	69.82	18188015.00	12577598.00	92680000.00	48290000.00	66.57	65.55	37980000.00	97380000.00	2021053.00	122.85	42.21	83.29	82.76
41	2000.00	70.29	18684937.00	12498723.00	76250000.00	61180000.00	75.26	62.89	199960000.00	118230000.00	934656.00	124.92	37.83	84.38	82.86
42	2001.00	70.78	19180228.00	12411925.00	77270000.00	42560000.00	77.22	60.09	198860000.00	161250000.00	2659595.00	128.15	56.85	84.08	83.20
43	2002.00	71.28	19677273.00	12317773.00	96940000.00	93340000.00	79.68	57.57	186120000.00	173380000.00	1953325.00	129.46	62.72	84.39	79.69
44	2003.00	71.78	20182205.00	12221309.00	146190000.00	72570000.00	77.39	55.32	239240000.00	225110000.00	4266387.00	130.34	62.82	86.41	75.76
45	2004.00	72.28	20703617.00	12127479.00	189370000.00	101410000.00	72.06	53.30	316970000.00	306750000.00	4033242.00	130.53	59.27	89.59	76.54
46	2005.00	72.77	21248009.00	12040428.00	197990000.00	125320000.00	73.28	51.57	347380000.00	323720000.00	3527824.00	133.31	53.83	84.32	74.44
47	2006.00	73.22	21818169.00	11959746.00	199770000.00	149690000.00	72.65	50.29	240680000.00	192810000.00	4018105.00	137.03	57.28	87.67	74.62
48	2007.00	73.65	22414414.00	11885662.00	228250000.00	162300000.00	69.29	49.36	395010000.00	377540000.00	3602256.00	138.65	64.09	85.96	75.19
49	2008.00	74.03	23041887.00	11818828.00	177170000.00	142270000.00	64.58	48.76	324830000.00	333030000.00	1536002.00	139.58	62.99	86.38	72.32
50	2009.00	74.37	23700349.00	11765411.00	191310000.00	118650000.00	72.65	48.49	302390000.00	283240000.00	5253472.00	142.08	73.16	84.67	69.02
51	2010.00	74.68	24393852.00	11723785.00	186240000.00	74560000.00	74.39	48.54	201250000.00	195040000.00	4211370.00	142.92	69.05	82.16	71.68
52	2011.00	74.94	25124194.00	11695364.00	74320000.00	217380000.00	72.94	49.12	193350000.00	186510000.00	4247550.00	133.13	68.06	81.51	70.89
53	2012.00	75.19	25888503.00	11677344.00	168570000.00	81960000.00	77.54	49.78	147320000.00	144050000.00	5137470.00	142.10	67.95	80.66	67.76
54	2013.00	75.42	26674438.00	11664124.00	168050000.00	99280000.00	79.37	50.60	203420000.00	149370000.00	4912551.00	144.76	71.73	80.30	68.03
55	2014.00	75.64	27465760.00	11647553.00	158450000.00	80490000.00	80.58	51.60	160620000.00	141900000.00	3435535.00	148.76	79.31	77.85	68.52
56	2015.00	75.86	28248180.00	11623348.00	126090000.00	66860000.00	100.69	52.72	87490000.00	80980000.00	3761224.00	147.40	82.00	79.18	65.60
57	2016.00	76.08	29016679.00	11589373.00	144290000.00	69520000.00	109.44	53.91	157930000.00	147620000.00	3445158.00	154.16	78.84	77.90	62.98
58	2017.00	76.29	29770548.00	11547594.00	560000.00	3240000.00	110.97	55.03	189190000.00	174150000.00	3478073.00	155.36	80.53	75.50	59.56

Table 3.13 – Données après estimation des valeurs manquantes de l'Algérie

classement		PIB	Masse.M	Inflation	P.T	FMI	PA	IPC	DEND	Pays
1	p	122000000000.00	16.73	22.61	29784193	388801462.90	12315755	280.76	24589134341.00	Angola
2	p	3172416146.00	27.34	11.47	10827024	188076863.40	4704153	188.68	354420183.30	Burundi
3	p	8573159697.00	39.23	0.03	11175692	226384650.10	4557465	2467846859.00	2467846859.00	Bénine
4	n	12322864245.00	46.94	4.52	19193382	284059238.30	7025864	108.12	2202894109.00	BFA
5	n	17406565823.00	40.24	2.61	2205128	81790877.30	1088986	141.14	299307670.90	Botswana
6	n	37353276059.00	38.71	-1.74	24294750	1737919425.00	8020583	112.45	4774525557.00	CIV
7	p	34922782311.00	20.52	1.48	24053727	602413053.00	10524055	114.58	7649373896.00	CAMROUNE
8	n	1844674435.00	113.03	0.45	944097	43954479.00	409428	116.22	629239934.00	Djibouti
9	n	2167501640.00	23.83	10.41	4596028	235403692.20	1883073	190.65	203155141.20	CAF
10	d	8701334800.00	33.25	-2.45	5110702	120028881.40	2045481	120.66	1024585195.00	congo
11	d	38019265626.00	12.67	43.07	81398764	953621644.60	27921049	135.26	1504725736.00	Rdc
12	d	168000000000.00	80.53	4.43	41318142	1706376276.00	12087847	142.38	89274001.70	Algérie
13	d	235000000000.00	92.30	22.93	96442593	7400858183.00	30802887	231.11	41822502782.00	Egypte
14	n	81716326731.00	39.74	6.72	106400024	301195845.50	50756189	249.09	10510820259.00	Ethiopie
15	d	14892609693.00	19.36	3.81	2064823	412399367.20	686504	2044616607.00	2044616607.00	Gabon
16	d	58996776238.00	26.10	10.35	28833629	1465019703.00	11969878	232.26	4991245394.00	Ghana
17	d	9915311049.00	25.37	3.12	12067539	426088145.30	4501365	218.71	690691768.70	Guinée
18	n	1489464788.00	38.68	4.90	2100568	93943014.70	682636	151.38	329102848.50	Gambie
19	n	1346841897.00	43.90	5.94	1828146	51604928.00	779272	111.65	178594735.10	Guinée B
20	p	12289913729.00	16.38	12.67	1262001	1046744358.00	19620177	172.43	16067953306.00	Kenya
21	n	3285455000.00	22.07	-2.18	4731906	401806656.70	1545265	180.58	598436029.30	Lebiria
22	d	2578265356.00	35.21	2.94	2091412	108050488.40	902949	142.67	289078431.60	Lesotho
23	d	110000000000.00	119.35	0.79	35739580	799540289.50	11793016	108.77	7863559102.00	Maroc
24	d	11465850504.00	29.38	8.28	25570540	439937850.80	13030680	161.17	1590183626.00	Madagascar
25	d	15334336144.00	26.74	1.93	18541980	334118980.70	6885003	110.23	2235866044.00	Mali
26	d	5024708656.00	3.47	4.20	4420184	199723433.10	1228317	128.06	2141067113.00	Mouritanie
27	d	13259351418.00	115.01	1.30	1264613	137864738.60	602840	125.36	210192415.50	Mourice
28	p	6303292264.00	23.00	13.46	17670260	318897834.00	7641860	340.24	490952484.20	Malawi
29	n	8119710126.00	24.34	0.93	21602472	286917475.70	8474249	108.88	2248049059.00	Niger
30	n	376000000000.00	20.38	11.12	59039086	2385958891.00	234	214.23	6429417337.00	Nigeria
31	d	9135454442.00	19.97	7.34	12208407	290711369.30	6133785	146.63	1327422885.00	Rowanda
32	d	21081669870.00	37.55	1.56	15419381	282670707.40	4084914	106.87	6440425719.00	Sénégal
33	p	375040173.80	33.08	1.98	207089	16807632.20	70560	171.59	62297600.80	STP
34	n	39952095561.00	73.52	5.57	11532127	2105458564.00	4085723	138.44	6895585055.00	Tunisie
35	n	53320625959.00	21.11	2.70	57310019	458750664.00	26307346	175.04	4624297597.00	Tanzanie
36	n	25995031850.00	23.57	6.28	41162465	246406070.40	15269508	166.78	5129288697.00	Ouganda
37	p	25868142073.00	21.96	10.20	17094130	793620163.50	7099751	180.95	5871251675.00	Zebabwie
38	P	9975692095.00	15.63	0.03	15016773	254233595.00	5658069	116.14	669314563.50	Tchad
39	n	4765866980.00	56.91	1.45	7698475	226017223.50	3498921	111.50	1021470684.00	Togo
40	n	349000000000.00	72.21	5.54	57000451	2542670622.00	22503030	146.05	1215409585.00	AFS

Table 3.14 – Tableau de données des pays africaines

BIBLIOGRAPHIE

Celeux, G. *Analyse discriminante*, Dans L'analyse des données, G.Govaert éditeur, Hermes Science, paris, France,2003.

F.Husson, S.Le, J.Pages. 2009 *Analyse de données avec R* , ed Broché. G. Celeux : *Analyse des données*, chapitre Analyse discriminante. Hermes Sciences publications, 2003

G.Celeux,J.P.Nakache, *Analyse discriminante sur variable qualitative*, Polytechnica, 1994.

Guide de la Banque mondiale Copyright I 2005 by The International Bank for Reconstruction and Development/ The World Bank 1818 H Street, NW, Washington, DC 20433, USA

HAMDAD Leila.*Analyse de données Polycopie 1* préparée par ,Version 2015/2016

Hassen MATHLOUTHI,*Cours de méthodes de scoring* , Année universitaire 2013-2014 L.Lebart,

A.Morineau et M.Pion. 1995. *Statistique exploratoire multidimensionnelle*, Dunod. Saporta G., *Probabilités, Analyse des Données et Statistique* , Tehcnip, 1990

<https://statistique-et-logiciel-r.com/>

[https://fr.wikipedia.org/wiki/R-\(langage\)](https://fr.wikipedia.org/wiki/R-(langage))

<http://factominer.free.fr/missMDA/index-fr.html>

<http://factominer.free.fr/index-fr.html>

<http://www.worldbank.org/afr/ik>>. Scoring Marie Chavent

<http://www.math.u-bordeaux.fr/machaven/> 2014-2015

