كليـــــــة العلوم
السملالية - مراكش
FACULTÉ DES SCIENCES
SEMLALIA - MARRAKECH

Department of Mathematics

End of studies project

Principal Component Analysis Using R Software

Presented by
ATERHI Mouad

Jury members

Pr. Abdelaziz Nasroallah   Pr. Abdallah Mkhadri   Pr. Lalla Aicha Allamy

Academic Year 2019/2020

Contents

Acknowledgements

At the end of this work, I would like to express my heartfelt thanks to all those who contributed to the smooth running of this graduation project.

Thus, I express my deep gratitude and my heartfelt thanks to my mentor **L. A. ALLAMY**, professor at the Faculty of Sciences Semlalia in Marrakesh, for the valuable advice she has given me, for his follow-up and for his positive involvement during the entire period of my preparation. She led my project with a lot of patience and dedicated a lot of time to my work by always being available for us to talk, which encouraged me enormously.

I would also like to express my heartfelt thanks to my mother and sister for their support during all my years of study.

I would also like to thank the jury members **A. NASROALLAH** and **A. MKHADRI** who made me the honor of judging my work.

I thank the professors **M. HOUIMDI** and **M. H. LALAOUI** who introduced us to the scientific word processing software LATEX, which was very benign and facilitated our work.

To the entire faculty and administrative staff of the Faculty of Sciences Semlalia in Marrakesh, I extend special thanks for the quality of the teaching that has been given to me.

Introduction

In the XXth century and more precisely in the sixties, the world experienced an explosion of information and development of the mathematical foundations of data analysis. This is related to the progress of computing and the extension of the very extensive applications of the machine in the computations of difficult mathematical operations.

In particular, this strong evolution in computer science has paved the way for a qualitative leap in the field of descriptive statistics.

However, in a statistical study it is important to write and analyse a set of observations or data, paying attention to the graphic representation and the interpretation of the results, in order to make their comprehension simpler.

In univariate (or bivariate) descriptive statistics, the treatment of such a data set is simple to work with. On the other hand, in multivariate descriptive statistics, the graphical representation is much more difficult.

To do this, we use the method of **Principal Component Analysis** or **P.C.A**.

This method makes it possible to analyse and visualize the important information contained in a data table. This table contains individuals written by several quantitative variables.

The aim of this method is to construct a space with an reduced dimension (two or three), allowing to visualize graphically the data contained in our table, while keeping as much information as possible.

In the first chapter of this paper, we recall some basic concepts of descriptive statistics and algebra.

In the second chapter, we give the procedure for the readjustment of the **Principal Component Analysis** method.

Finally, we dedicate the last chapter to visualize and interpret the results of a data set submitted to a Principal Component Analysis, through a numerical application using the R software.

Chapter 1

Generalities

1.1 Concepts of descriptive statistics

Consider a sample of n observations $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ of a (X, Y) couple of quantitative variables.

We give some definitions concerning dispersion measures between X and Y.

Definition 1.1.1

The **covariance** of X and Y is given by:

$$Cov(X,Y) = \sum_{i=1}^{n} p_i(x_i - \bar{x})(y_i - \bar{y})$$

where $\bar{x}$ and $\bar{y}$ are the empirical averages of X and Y respectively and p_i is the weight of the i^{th} observation.

Note 1.1.1

In General, we work with $p_i = \frac{1}{n}, i = 1, 2, \cdots, n$.

Definition 1.1.2

The **correlation coefficient** of X and Y is given by:

$$r_{X,Y} = \frac{Cov(X,Y)}{s_X s_Y}$$

where s_X and s_Y are the standard deviations of X and Y respectively.

Notes 1.1.1

- $Cov(X,X) = s_X^2$: is the variance of the variable X.
- $Cov(X,Y) = Cov(Y,X)$ and $-1 \leq r_{X,Y} \leq 1$.

When we have a $p \geq 3$ number of variables $X_1, X_2, \ldots, X_p$, we introduce two interesting matrices.

Definition 1.1.3

The **variance-covariance matrix V** and the **correlation matrix R** are given by:

$$V = \begin{pmatrix} s_{X_1}^2 & Cov(X_1,X_2) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & s_{X_2}^2 & \cdots & Cov(X_2,X_n) \\ \vdots & \ddots & \ddots & \vdots \\ Cov(X_n,X_1) & \cdots & \cdots & s_{X_n}^2 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & r_{X_1,X_2} & \cdots & r_{X_1,X_n} \\ r_{X_2,X_1} & 1 & \cdots & r_{X_2,X_n} \\ \vdots & \ddots & \ddots & \vdots \\ r_{X_n,X_1} & \cdots & \cdots & 1 \end{pmatrix}$$

Notes 1.1.2

For any $i, j = 1, \ldots, n$;:

1. If r_{X_i,X_j} =0, then X_i and X_j are **uncorrelated linearly**.
2. If $r_{X_i,X_j} = \pm 1$, then X_i and X_j are **correlated linearly**.
3. If $Cov(X_i,X_j) \geq 0$, then X_i and X_j are **positively correlated** and they evaluate in the same direction.
4. If $Cov(X_i,X_j) \leq 0$, then X_i and X_j are **negatively correlated and assess in the opposite direction**.

1.2 Notions of linear algebra

In Principal Component Analysis, linear algebra plays a very important role in the mathematical explanation of the phenomena of a statistical study. For this, we recall some basic notions that we need.

1.2.1 Diagonalizable matrices

In this part, we define the eigenvalue and the characteristic polynomial of a square matrix, hence we recall that of diagonalizable matrices.

Definition 1.2.1
Let $A \in M_n(\mathbb{R})$ the set of real n size matrices.
A real λ is said an **eigenvalue of** A if there is $x \in \mathbb{R}^n$ nonzero such as:

$$Ax = \lambda x \tag{1.1}$$

The x vector is said to be an **eigenvector of** A **associated with the eigenvalue** *lambda*.

Equation (1.1) is equivalent to: $(\lambda I_n - A)x = 0$
This is equivalent to determining the roots of the polynomial characteristic of A, given by:

$$\chi_A(X) = \det(XI_n - A)$$

We thus define the eigenspace of a square matrix associated with a proper value.

Definition 1.2.2
Let $A \in M_n(\mathbb{R})$ and $\lambda \in \mathbb{R}$ be a clean value of A. Then:

$$E_\lambda(A) = E_\lambda = Ker(A - \lambda I_n) = \{x \in \mathbb{R}^n \ \ Ax = \lambda x\}$$

E_λ is **the eigenspace of** A **associated with** *lambda*.

A diagonalizable matrix can be defined as follows.

Definition 1.2.3
Let $A \in M_n(\mathbb{R})$. The A matrix is **diagonalizable** if and only if there is $P \in GL_n(\mathbb{R})$, the set of real inversible matrices, such as $P^{-1}AP$ is **diagonal**.

Note 1.2.1

Any symmetrical matrix is diagonalizable.

1.2.2 Euclidean spaces

In this paragraph, we give the definition of a scalar product from which we recall that of a standard and the distance between two vectors.

Definition 1.2.4

A scalar product on a $\mathbb{R}$-vector space E, is an application of $E \times E$ to $\mathbb{R}$, noted $< .,. >$, with the following conditions: For all $x, y, z \in E$ and all $\alpha, \beta \in \mathbb{R}$

1. $< x, x >= 0 \Longleftrightarrow x = 0$
2. $< x, y >=< y, x >$
3. $< \alpha x + \beta y, z >= \alpha < x, z > +\beta < y, z >$
4. $< z, \alpha x + \beta y >= \alpha < z, x > +\beta < z, y >$

E with a scalar product is called an **euclidean space**.

Note 1.2.2

In $\mathbb{R}^n$, the scalar product of $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ is given by:

$$< x, y >= \sum_{k=1}^{n} x_k y_k$$

A scalar product induces a **norm**, as an application of E to $\mathbb{R}_+$, noted $||.||$, defined by: $|x|| = \sqrt{< x, x >}$. In addition, a norm checks the following properties:

1. $\forall x \in E,\ ||x|| = 0 \Longleftrightarrow x = 0$
2. $\forall \lambda \in \mathbb{R},\ \forall x \in E,\ ||\lambda x|| = |\lambda|||x||$
3. $\forall x, y \in E,\ ||x + y|| \leq ||x|| + ||y||$ (Triangle inequality)

Thus, we determine the **distance** between two vectors from the definition of a norm.

Definition 1.2.5

The **distance** between two vectors x and y of E is defined by:

$$d(x, y) = ||y - x||$$

Notes 1.2.1

- The distance of a vector x to the origine 0 is therefore $||x||$.
- In $\mathbb{R}^n$, we work with **euclidean distance** definied by:

$$d(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2}$$

1.2.3 Orthogonality

In this last paragraph, we recall the definition of an orthogonal subspace and subsequently the definition of an orthogonal family.

Definition 1.2.6

Let F a vector sub-space of E. The **orthogonal of** F, noted $F^\perp$, is the vector sub-space of E defined by:

$$F^\perp = \{y \in E \,/ < x, y >= 0 \,, \forall x \in F\}$$

Therefore, the **orthogonal projection** on F, of a $x \in E$ vector is the only $y \in F$ vector such as: $x - y \in F^\perp$. What is equivalent to write:

$$< x - y, y >= 0$$

Note 1.2.3

The **distance of** x **to** F is the distance of x to its orthogonal projection over F.

Definition 1.2.7

Two vectors x and y of E are **orthogonal** if their scalar product equals zero:

$$< x, y >= 0$$

Note 1.2.4

A family $(x_i)_{i\in I}$ of E is called orthogonal if the x_i vectors are orthogonal two by two.

Chapter 2

Principal Component Analysis

2.1 Data sets

The Principal Component Analysis (PCA) is concerned with rectangular tables of data. This is X data set of quantitative data, with individuals in rows and variables in columns.

	V^1	$\cdots$	V^j	$\cdots$	V^p
I_1	x_1^1	$\cdots$	x_1^j	$\cdots$	x_1^p
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_i	$\cdots$	$\cdots$	x_i^j	$\cdots$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
I_n	x_n^1	$\cdots$	x_n^j	$\cdots$	x_n^p

The corresponding matrix $X \in M_{n\times p}(\mathbb{R})$ is written:

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

where $I_i = x_i = (x_i^1, x_i^2, \cdots, x_i^p) \in \mathbb{R}^p$ is the vector containing the data of the i^{th} individual, $V^j = x^j = (x_1^j, x_2^j, \cdots, x_n^j)^T \in \mathbb{R}^n$ is the vector containing thedata of the j^{th} variable and x_i^j is the corresponding data to the i^{th} individual and the j^{th} variable.

To facilitate the understanding of our work in PCA, we treat in this chapter the example of the notes of 6 students (individuals) in 4 modules (variables).

The table of initial data for the notes example is given by TABLE 2.1.

	Algèbre	Analyse	Programmation	Module option
Marouane	15.25	13.80	12.0	11.5
Ziad	13.50	12.00	10.0	14.0
Yasmine	17.00	18.00	7.0	12.8
Issam	16.50	15.00	14.0	15.5
Hafsa	10.00	8.75	10.5	11.0
Oussama	13.00	9.00	10.0	9.5

Table 2.1: Table of initial data

The objective of PCA is to analyze the information contained in the initial table. This is like analyzing the structure of the cloud of individuals in the space $\mathbb{R}^p$ and the structure of the cloud of variables in the space $\mathbb{R}^n$.

The analysis of such a table is carried out after a pre-processing of the data, we can:

- **Center** the variables (have the variables (columns) of **mean zero**).
- **Center and reduce** the variables (have the variables of **mean zero** and o **variance equals to 1**).

Indeed, centering the variables allows to build a new data set Y (centered data set) whose matrix form is:

$$Y = X - 1_n\bar{x}$$

where $1_n = (1, 1, \cdots, 1)^T \in \mathbb{R}^n$, $\bar{x} = (\bar{x^1}, \bar{x^2} \cdots, \bar{x^p}) \in \mathbb{R}^p$ called **the centre of gravity** and $\bar{x^j} = \dfrac{1}{n}\sum_{k=1}^{n} x_k^j$ is the empirical average [1] of the j^{th} variable.

The empirical averages $\bar{x^1}$, $\bar{x^2}$, $\bar{x^3}$ and $\bar{x^4}$, of students notes in each module (variable), Algèbre, Analyse, Programmation and Module option respectively, are given in the following table (TABLE 2.2).

[1] We assume that the variables have the same weight $p_i = \frac{1}{n}$. (See Chaptre 1)

Algèbre	Analyse	Programmation	Module option
14.20833	12.75833	10.58333	12.38333

Table 2.2: Averages of the variables in Table X

From the two tables TABLE 2.1 and TABLE 2.2, we deduct table Y below of the centered data.

	Algèbre	Analyse	Programmation	Module option
Marouane	1.0416667	1.0416667	1.41666667	-0.8833333
Ziad	-0.7083333	-0.7583333	-0.58333333	1.6166667
Yasmine	2.7916667	5.2416667	-3.58333333	0.4166667
Issam	2.2916667	2.2416667	3.41666667	3.1166667
Hafsa	-4.2083333	-4.0083333	-0.08333333	-1.3833333
Oussama	-1.2083333	-3.7583333	-0.58333333	-2.8833333

Table 2.3: Table Y below of the centered data

Note that the product of the Y matrix transpose by the Y matrix equals the p sized variance-covariance matrix with a coefficient plus:

$$V = \frac{1}{n} Y^T Y$$

Thus, from TABLE 2.3, we obtain the variance-covariance matrix corresponding to our example, given by TABLE 2.4 below.

	[,1]	[,2]	[,3]	[,4]
[1,]	5.6336806	7.133681	0.1284722	2.590972
[2,]	7.1336806	10.725347	-1.1131944	3.900972
[3,]	0.1284722	-1.113194	4.5347222	1.459722
[4,]	2.5909722	3.900972	1.4597222	3.918056

Table 2.4: Variance-covariance matrix

Similarly, we construct the Z table of centered-reduced data by dividing the columns of the Y table, of centered data, by the standard deviation of each of the variables in the X table of initial data.

The matrix form of the Z data set construction is:

$$Z = (X - 1_n \bar{x}) D_{\frac{1}{s}} = Y D_{\frac{1}{s}}$$

where $D_{\frac{1}{s}} = diag\left(\frac{1}{s_1}, \cdots, \frac{1}{s_p}\right)$ with $s_j = \left(\frac{1}{n}\sum_{k=1}^{n}(x_k^j - \bar{x^j})^2\right)^{1/2}$ is the standard deviation of the j^{th} variable.

For our example, the 4 standard deviations of variables (modules) are given in TABLE 2.5.

Algèbre	Analyse	Programmation	Module option
2.600080	3.587536	2.332738	2.168333

Table 2.5: Standard deviations of variables in table X

Nous déduisons le tableau Z, des données centrées-réduites présenté dans TABLE 2.6. We deduct table Z, of centered-scaled data presented in TABLE 2.6.

	Algèbre	Analyse	Programmation	Module option
Marouane	0.4006287	0.2903571	0.6072978	-0.4073791
Ziad	-0.2724275	-0.2113800	-0.2500638	0.7455805
Yasmine	1.0736849	1.4610770	-1.5361062	0.1921599
Issam	0.8813831	0.6248485	1.4646594	1.4373563
Hafsa	-1.6185399	-1.1172942	-0.0357234	-0.6379710
Oussama	-0.4647293	-1.0476084	-0.2500638	-1.3297467

Table 2.6: Table Z of centered-scaled data

Note also that the product of the Z matrix transpose by the Z matrix gives us the p size correlation matrix with one coefficient plus:

$$R = \frac{1}{n} Z^T Z \tag{2.1}$$

Thus, we can obtain the correlation matrix through the variance-covariance matrix V by the following formula:

$$R = D_{\frac{1}{s}} V D_{\frac{1}{s}}$$

The correlation matrix in the notes example is a direct application of the formula (2.1) working with the Z matrix corresponding to the Z data set (TABLE 2.6). It is given by TABLE 2.7 .

	[,1]	[,2]	[,3]	[,4]
[1,]	1.00000000	0.9177235	0.02541779	0.5514820
[2,]	0.91772353	1.0000000	-0.15962099	0.6017719
[3,]	0.02541779	-0.1596210	1.00000000	0.3463057
[4,]	0.55148201	0.6017719	0.34630565	1.0000000

Table 2.7: Correlation matrix

A table submitted to a PCA is always centered, however, the choice to reduce the variables is determined by the nature of the relationship between the variables. If the variables are **homogens** (same unit of measure, same meaning...), then we can choose to center without reducing. On the other hand, if the variables are **heterogeneous** (for example, different units), then reducing the variables allows us to compare the values taken by these variables, which leads to give them the same importance.

Note 2.1.1

We are now working with the table Z of centered-reduced data and with the correlation matrix R.

2.2 Study of individuals

In the Z data set, each individual is represented by a point of $\mathbb{R}^p$ space, said **individuals space**.

2.2.1 Notion of similarity

The space of individuals has the structure of an Euclidean space, which allows us to define a distance between individuals.

Two individuals z_i and z_j are similar if they take values close in the p variables space. So we can define the distance $d(i,j)$ between two individuals z_i and z_j:

$$d^2(i,j) = d^2(z_i, z_j) = ||z_i - z_j||^2_M, \quad i,j = 1, 2, \cdots, n$$

With $z_i = (z_i^1, z_i^2, \cdots, z_i^p)$ is the i^{th} individual vector, while $z_i^k = \frac{x_i^k - \bar{x}^k}{s_k}$ and $M \in M_p(\mathbb{R})$ is a symmetrical matrix defined positively, specifying the selected distance, called **metric**.

Working with the Z, our metric will be the identity matrix I_p. In this case, the distance used will be **Euclidean distance**:

$$d(z_i, z_j) = \left(\sum_{k=1}^{p} (z_i^k - z_j^k)^2 \right)^{1/2}$$

2.2.2 Inertia

The PCA aims to find a space of reduced dimension that best summarizes the information contained in the data. In other words, to provide a simple image of the point cloud that does not distort the distances between individuals too much.

In large spaces, we work with a dispersion measure called **inertia**, given by:

$$\begin{aligned} I_t &= \frac{1}{n} \sum_{i=1}^{n} d^2(z_i, O) \quad \text{(because variables mean is zero)} \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} (z_i^j)^2 \end{aligned}$$

I_t is the total inertia of the point cloud. The greater the total inertia, the more dispersed the cloud is. On the contrary, the smaller the total inertia, the more concentrated the cloud is around the center of gravity O.

Note that:

$$I_t = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_i^j - \bar{x^j}}{s_j} \right)^2 = \sum_{j=1}^{p} \frac{1}{n} \sum_{i=1}^{n} 1 = \sum_{j=1}^{p} 1 = p = Tr(R)$$

The total inertia of the cloud is nothing more than the trace of the correlation matrix R.

To build our space of reduced size (at most 2 dimensions), we must look for the axes that generate it.

For this, we will need to use the notion of a space and its orthogonal.

So let's represent, in FIGURE 2.1, a F axis passing through O (the center of gravity), of a vector u and call $F^{\perp}$ its orthogonal.

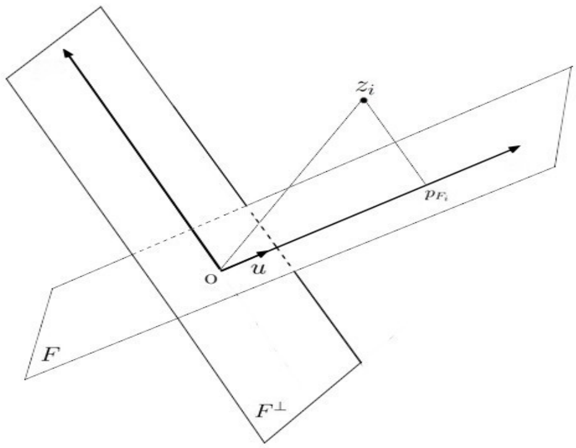

Figure 2.1: Orthogonal projection on the F axis of an individual z_i in $\mathbb{R}^p$

According to the *Pythagoras* theorem $d^2(z_i, O) = d^2(O, p_{F_i}) + d^2(z_i, p_{F_i})$

Which gives us:

$$I_t = I_F + I_{F^{\perp}}$$

The most faithful representation to look for is to have a good separation of points that allows us to see individuals better. This increases the dispersion or variability of the points.

In other words, we aim to maximize the inertia I_F carried by the F axis and minimize the inertia $I_{F^{\perp}}$ carried by the $F^{\perp}$ axis:

$$I_t = \underbrace{I_F}_{maximize} + \underbrace{I_{F^{\perp}}}_{minimize}$$

So:

$$\begin{aligned} I_F = \frac{1}{n}\sum_{i=1}^{n} d^2(O, p_{F_i}) &= \frac{1}{n}\sum_{i=1}^{n} < z_i, u >^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}((z_i)^T u)^T((z_i)^T u) \\ &= \frac{1}{n} u^T \left(\sum_{i=1}^{n}(z_i)^T z_i\right) u \end{aligned}$$

Then:

$$I_F = \frac{1}{n} u^T Z^T Z u$$

Next:

$$\boxed{I_F = u^T R u}$$

where $R = \dfrac{1}{n} Z^T Z$ is the correlation matrix.

2.3 Study of variables

The representation of variables differs from that of individuals. In fact, given the previous section, individuals are represented by points in the $\mathbb{R}^p$space. Here, variables are represented by vectors in space $\mathbb{R}^n$ said **variables space**.

In the study of variables, we are interested in angles rather than distances.

Let $z^k = (z_1^k, z_2^k, \cdots, z_n^k)$ and $z^l = (z_1^l, z_2^l, \cdots, z_n^l)$ two centered and redued variables taken from table Z and let $\theta_{k,l}$ the angle between z^k and z^l.

$$\cos(\theta_{k,l}) = \frac{< z^k, z^l >}{||z^k||||z^l||} = \frac{\sum_{i=1}^{n} z_i^k z_i^l}{\sqrt{< z^k, z^k >}\sqrt{< z^l, z^l >}} = \frac{n\, r_{k,l}}{n} = r_{k,l}$$

where $r_{k,l}$ is the correlation coefficient of two variables z^k et z^l.

We will see in the following the importance of the above result in the construction of the variable cloud.

2.4 Principal Component Analysis

2.4.1 Principal Components

In order to preserve the information contained in the data table, we construct new variables $C^i (i = 1, \cdots, q)$, called **principal components**, which are **linear combinations** of the initial variables. They are written in the following form:

$$
\begin{aligned}
C^1 &= a_1^1 x^1 + a_2^1 x^2 + \cdots + a_p^1 x^p \\
C^2 &= a_1^2 x^1 + a_2^2 x^2 + \cdots + a_p^2 x^p \\
&\vdots \\
C^q &= a_1^q x^1 + a_2^q x^2 + \cdots + a_p^q x^p
\end{aligned}
$$

with $1 \leq q < p$ where p is the number of variables.

The principal components must verify the following criteria:

- The principal components are two to two **uncorrelated**. This is to say, $r_{C^i,C^j} = 0$, for all $i \neq j$.

- The first principal component C^1 must contain the maximum amount of information, and therefore the maximum variability of the individuals.

2.4.2 Main axes

In this part, we define the main axes and give the approach to follow for their construction.

> **Definition 2.4.1**
> The **main axes** are the axes that generate the new space of reduced dimension whose inertia explained by these axes, is maximum.

According to paragraph 2.2.2, the inertia explained by a F axis is given by:

$$I_F = u^T R u$$

where u is the guiding vector for F and $R = \frac{1}{n} Z^T Z$ is the correlation matrix for $x^1, \cdots, x^p$ variables.

The graphical representation of individuals consists of finding the u orthonormed director vector of the F axis, which maximizes the amount $uTRu$.

This solves the following optimization problem:

$$\begin{cases} \max_u \ u^T R u \\ \\ u^T u = 1 \end{cases}$$

The method of *Lagrange multipliers* can then be used.

The **Lagrangian** function of the optimization problem under the constraint $u^T u - 1 = 0$ is:

$$\mathcal{L}(u, \lambda) = u^T R u - \lambda(u^T u - 1)$$

λ being the Lagrange multiplier.

Look for the u vector of $\mathbb{R}^p$ such as:

$$(1) \begin{cases} \dfrac{\partial \mathcal{L}}{\partial u}(u, \lambda) = 0 \\ \\ u^T u - 1 = 0 \end{cases}$$

Recall that for any square matrix $A \in M_p(\mathbb{R})$ and for all vector $x \in \mathbb{R}^p$:

$$\frac{\partial}{\partial x}(x^T A x) = (A + A^T)x$$

Indeed, for any matrix $A \in M_p(\mathbb{R})$, the f_A function defined by:

$$\begin{aligned} f_A : \mathbb{R}^p &\longrightarrow \mathbb{R} \\ x &\longmapsto f_A(x) = < x, Ax > = x^T A x \end{aligned}$$

is differentiable in $\mathbb{R}^p$ and its differential is:

$$df_A(x) = \frac{\partial f_A}{\partial x}(x) = (A + A^T)x, \ \forall x \in \mathbb{R}^p$$

Proof
(See **Appendix**)

Note 2.4.1
If A is symmetrical, then $A^T = A$, so: $\dfrac{\partial}{\partial x}(x^T A x) = 2Ax$

Then (1) $\iff \begin{cases} 2Ru - 2\lambda u = 0 \\ \\ u^T u = 1 \end{cases} \iff (2) \begin{cases} Ru = \lambda u \\ \\ u^T u = 1 \end{cases} \iff \begin{cases} u \text{ is the eigenvector of} \\ \text{R matrix, associated to the} \\ \text{eigenvalue } \lambda \\ \\ u^T u = 1 \end{cases}$

Multiplying the first equation of (2) by u^T, gives:

$$\begin{cases} u^T Ru = \lambda \\ \\ u^T u = 1 \end{cases}$$

Then the maximum value of $uTRu$ is the largest eigenvalue of the R matrix, and the maximizing vector is none other than the eigenvector associated with this largest eigenvalue.

The correlation matrix being symmetrical, so diagonalizable, then we can write:

$$R = PDP^{-1}$$

where $D = diag(\lambda_1, \lambda_2, \cdots, \lambda_p)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ are the eigenvalues of R matrix and P is the passage matrix.

The eigen space E_{λ_i} of the matrix R, associated with the eigenvalue λ_i allows us to find the eigenvector v_i associated with that eigenvalue. And therefore, we build the passage matrix P through the eigenvectors $v_1, v_2, \cdots, v_p$ associated to the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_p$ respectively, as the column vectors of P.

So the axis carrying maximum inertia, noted F_1, has as guiding vector $u = v_1$ associated to the highest eigenvalue λ_1.

Similarly, we build the second main axis, noted F_2, that has as guiding vector v_2 **orthogonal to** v_1, associated with the second eigenvalue $lambda_2$ and carrying inertia $I_{F_2} = lambda_2$.

Finally, the $k < p$ dimension space is generated by the k first main axes whose vector directors are the eigenvectors associated with the k largest eigenvalues of the correlation matrix R.

The rate of information preserved by the main axes $F_1, \cdots, F_k$ is given by:

$$\frac{1}{I_t}\sum_{i=1}^{k} I_{F_i} = \frac{1}{p}\sum_{i=1}^{k} \lambda_i$$

Let us apply these results to our example.

From the correlation matrix R given by TABLE 2.7, we calculate the associated 4 eigenvalues and present the results in descending order. Thus, we determine the percentage of information preserved by each main axis and subsequently calculate the cumulative percentage.

The results of our work are presented in TABLE 2.8.

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.40106164	60.026541	60.02654
comp 2	1.18646238	29.661560	89.68810
comp 3	0.36624332	9.156083	98.84418
comp 4	0.04623266	1.155816	100.00000

Table 2.8: Eigenvalues of the correlation matrix

In addition, the coordinate c_i^k, $(i = 1, \cdots, n)$ of i^{th} individual according to the k^{th} main axis of the point cloud, is in the following form:

$$c_i^k = \alpha_{i,1}v_{1,k} + \alpha_{i,2}v_{2,k} + \cdots + \alpha_{i,p}v_{p,k}$$

where $v_{j,k}$, $(j = 1, \cdots, p)$ is the j^{th} coordinate of eigenvector v_k.

We can summarize this with the following formula:

$$\boxed{C^k = Zv_k}$$

where $C^k = (c_1^k, \cdots, c_n^k)^T \in \mathbb{R}^n$ is the k^{th} principal component in which the variance is $s^2_{C^k} = \lambda_k$.

In our notes example, students coordinates ,under the 4 main axes, are grouped under the C^1, C^2, C^3 and C^4 principal components, with variance of $lambda_1, lambda_2, lambda_3$ and $lambda_4$, respectively, represented in TABLE 2.8. These components are given by TABLE 2.9.

2.4.3 Choice of principal components

The choice problem is to determine the number of components q to use in order to interpret them. In the statistical literature we find several rules of choice. We quote some of them:

	Dim.1	Dim.2	Dim.3	Dim.4
Marouane	0.27331843	0.2795422	0.86286512	0.19522214
Ziad	0.08130184	0.1566973	-0.90075945	-0.20555528
Yasmine	1.67157695	-2.0017794	-0.11656720	0.07767241
Issam	1.91718088	1.6660419	0.02930247	-0.04531945
Hafsa	-2.17192575	0.2875431	-0.49387225	0.29615616
Oussama	-1.77145234	-0.3880450	0.61903132	-0.31817598

Table 2.9: Principal components

- A first empirical rule proposed in 1960 by *Kaiser* indicates that we use only the main components for which the variance (the associated eigenvalue) is greater than the mean variance:

$$\frac{1}{p}\sum_{i=1}^{p}\lambda_i$$

  For centered-reduced data, the average of the eigenvalues is 1.

- Another empirical rule introduced in 1966 by *Cattell*, called **scree test**, proposes to study the graph of eigenvalues of the R matrix according to their rank, called **scree plot**. The idea is to retain the components whose corresponding eigenvalues are above the right passing through the first "elbow" signaling the first change in the structure of the graph.

The selection criteria for the principal components are numerous. We therefore choose the **scree test** and we get the *scree plot* given by FIGURE 2.2.

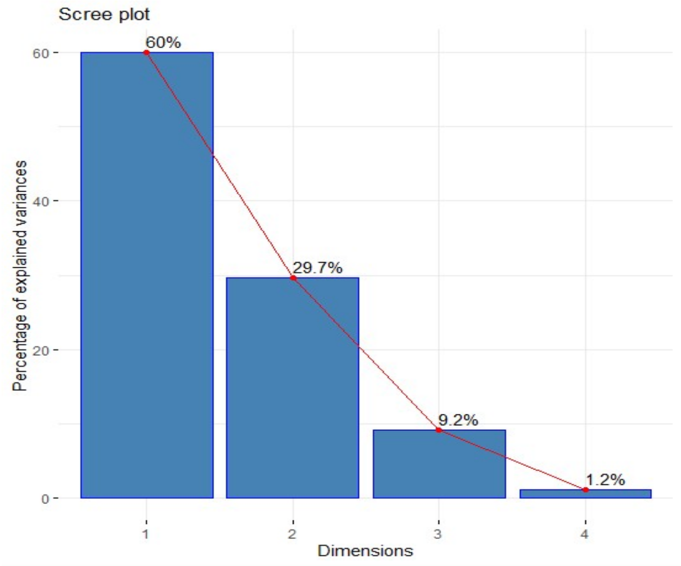

Figure 2.2: Scree plot

The *scree plot* allows us to identify the first elbow changing the structure of the graph and which is the third eigenvalue. So we choose the first two axes as the main axes expressing 89.7% of the information.

2.4.4 Representation of individuals

The k^{th} principal component $C^k = (c_1^k, c_2^k, \cdots, c_n^k)^T \in \mathbb{R}^n$ provides coordinates for n individuals on the main axis F_k.

If we want a flat representation of individuals, the best one will be the one achieved with the first two main axes, F_1 carrying 60% of the information and F_2 carrying 29.7%. We obtain the plane representation of the cloud given by FIGURE 2.3.

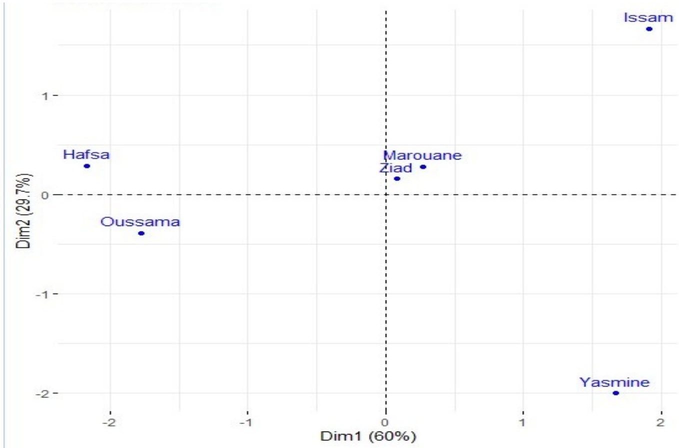

Figure 2.3: Cloud of individuals

2.4.5 Representation of variables

A variable zj is represented relative to the F_k axis in a circle of center O (the center of gravity) and radius unit, called **correlation circle**.

This representation is achieved by the angle created between zj and the principal component C^k, i.e. by the correlation coefficient of zj and Ck.

In our example, the coordinates of the variables (modules) are given by TABLE 2.10.

	Dim.1	Dim.2	Dim.3	Dim.4
Algèbre	0.9320508	-0.1502867	0.30271743	-0.13060401
Analyse	0.9408206	-0.2933991	0.05799471	0.15940572
Programmation	0.1068104	0.9608477	0.25163189	0.04521850
Module option	0.7973651	0.3931476	-0.45598628	-0.04147714

Table 2.10: Coordinates of variables

We then obtain the representation of the variables in the correlation circle (FIGURE 2.4).

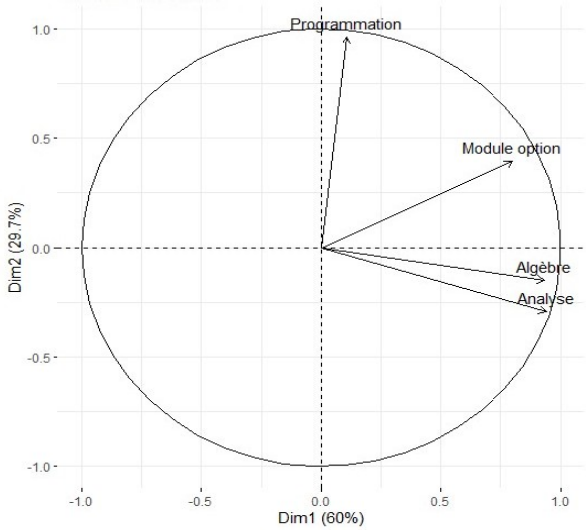

Figure 2.4: Correlation circle

2.5 Results interpretation

In this paragraph, we give an interpretation of the results obtained concerning the cloud representation of individuals on the one hand and the representation of variables in the correlation circle on the other hand.

2.5.1 Individuals interpretation

FIGURE 2.3 represents 6 students in the plan generated by the first two main axes retaining 89% of the information in the initial table.

The students are very distinguished. Indeed, we notice that they constitute 3 groups: the first group is composed of Hafsa and Oussama, the second group is composed of Marouane and Ziad and finally the third group is composed of Issam and Yasmine.

The question is, why are they grouped in this way?

First of all, we define some tools that help in the interpretation of the cloud of individuals and that therefore allow to answer our question.

Definition 2.5.1

The representation quality of an individual z_i according to the k^{th} main axis is defined by the cosine squared from the axis to the vector from the center of gravity O, to the point representing the i^{th} individual. It is given by:

$$\cos^2_k(z_i) = \frac{(c_i^k)^2}{\sum_{l=1}^{p}(c_i^l)^2}$$

where c_i^k is the coordinate of i^{th} individual according to the k^{th} main axe, for $i = 1, \cdots, n$ and $k = 1, \cdots, p$.

Notes 2.5.1

- The main axes being orthogonal, the quality of representation can be added.

  So the representation quality of the i^{th} individual in the plane generated by the first two main axes is given by:

$$\cos^2_{1,2}(z_i) = \frac{(c_i^1)^2 + (c_i^2)^2}{\sum_{l=1}^{p}(c_i^l)^2}$$

- The closer the $cos^2_k(z_i)$ amount is to 1, the better the representational quality of the individual z_i is.

Let us look at this on our example of notes. The quality of representation or cosine squared of students according to each axis is given by TABLE 2.11.

	Dim.1	Dim.2	Dim.3	Dim.4
Marouane	0.079853970	0.08353209	0.7958743366	0.0407396057
Ziad	0.007470732	0.02775144	0.9170227357	0.0477550941
Yasmine	0.409647530	0.58747589	0.0019920964	0.0008844863
Issam	0.569487763	0.43006098	0.0001330352	0.0003182200
Hafsa	0.919264463	0.01611226	0.0475313093	0.0170919675
Oussama	0.831697619	0.03990899	0.1015621199	0.0268312736

Table 2.11: Cosine squared of individuals

We can view the data from TABLE 2.11 in FIGURE 2.5 below.

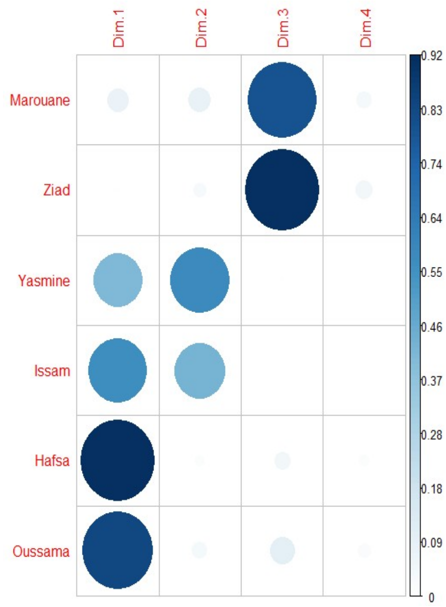

Figure 2.5: Visualisation of cosine squared for individuals

FIGURE 2.5 shows that the students best represented in the first axis are Hafsa and Oussama. Their representational qualities are close to 1 and they are 0.92 and 0.83 respectively. However, they are poorly represented along the other axes.

On the other hand, Yasmine and Issam have an acceptable representation according to the first two main axes but they are poorly represented according to the last two axes. Thus the two students Ziad and Marouane have a very good representation according to the third axis and a bad representation according to the other axes.

The quality of representation is not enough for the interpretation of the cloud of individuals. To do that, we use an additional tool, contribution.

Definition 2.5.2

The contribution of individuals is another criterion for the interpretation of the point cloud. It expresses the percentage of the contribution of an individual z_i in building the k^{th} main axis. It is given by:

$$CTRB_k(z_i) = \frac{(c_i^k)^2}{\sum_{j=1}^{n}(c_j^k)^2} \times 100$$

where c_i^k is the coordinate of i^{th} individual according to the k^{th} main axe, for $i = 1, \cdots, n$ and $k = 1, \cdots, p$.

The contribution of the students in the construction of the 4 axes is shown in the following table (TABLE 2.12).

	Dim.1	Dim.2	Dim.3	Dim.4
Marouane	0.5185412	1.0977146	33.88167429	13.7390927
Ziad	0.0458824	0.3449196	36.92297519	15.2320081
Yasmine	19.3953752	56.2894788	0.61834635	2.1748706
Issam	25.5135927	38.9911998	0.03907396	0.7404048
Hafsa	32.7442757	1.1614505	11.09962465	31.6185132
Oussama	21.7823327	2.1152367	17.43830557	36.4951107

Table 2.12: Contribution of individuals for all axes

Note that all students contribute to the construction of the first axis with a percentage higher than the average $\frac{1}{6}$ which is worth almost 17%, except Marouane and Ziad who have a small contribution. On the other hand, the second axis is built with a large contribution from Yasmine and Issam (56% and 39% respectively). As these two students contrast in the cloud of individuals (FIGURE 2.3) with respect to the second axis. This opposition steps from the fact that this axis takes into account their notes in a specific module or modules. This means that one of the two students had a good grade in one module while the other had a bad grade.

Returning to the table of initial data (TABLE 2.1), we notice that Yasmine has a bad note in the Programming module unlike Issam. This means that the second axis opposes individuals according to their note in the Programming module.

Moreover, by projecting the points according to the first two axes, the 4 students: Hafsa, Oussama, Marouane and Ziad position themselves close to the average (center of gravity O) according to the second axis. On the other hand, they are represented differently according to the first axis.

Indeed, Marouane and Ziad are represented close to the average of the first axis while Hafsa and Oussama are far from the average.

This suggests that the first axis is constructed from the variables Analysis, Algebra and Module option.

2.5.2 Variables interpretation

The interpretation of variables is done in the same way as for individuals.

It is carried out using the two criteria: the quality of representation and the contribution of the variables in the construction of each main axis.

Definitions of these tools are given below in order to interpret the correlation circle.

Definition 2.5.3

The representation quality of a z^j variable according to k^{th} main axis is given by:

$$\cos^2(\theta_{k,j}) = \frac{(r_{C^k,z^j})^2}{\sum_{l=1}^{p}(r_{C^l,z^j})^2} = (r_{C^k,z^j})^2$$

where $\theta_{k,j}$ is the angle between variable z^j and the k^{th} main axis, and r_{C^k,z^j} is the correlation coefficient of k^{th} principal component C^k and the variable z^j, for $k, j = 1, \cdots, p$.

The previous formula allows us to calculate the representation quality of each module in the notes example, following the 4 main axes.

We obtain the results given by TABLE 2.13.

	Dim.1	Dim.2	Dim.3	Dim.4
Algèbre	0.86871866	0.02258608	0.091637844	0.017057407
Analyse	0.88514339	0.08608304	0.003363386	0.025410183
Programmation	0.01140847	0.92322821	0.063318607	0.002044713
Module option	0.63579112	0.15456505	0.207923483	0.001720353

Table 2.13: Cosine squared of variables

We can visualize these results in the correlation circle shown in FIGURE 2.6.

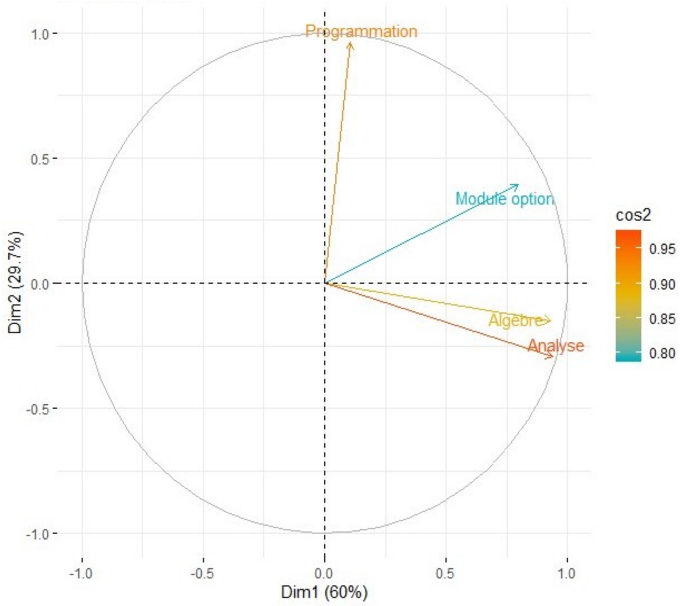

Figure 2.6: Cosine squared of variables

FIGURE 2.6 represents the variables in different places according to the quality of their representation in the first principal plane, summing their quality of representation according to the first and second principal axis.

We note that the Programming module is poorly represented in the first axis and is well represented in the second. On the other hand, the Algebra and Analysis modules are well represented on the first axis and are poorly represented on the second axis.

Variables, like individuals, contribute to the formation of the main axes. We therefore define this contribution of the variables in order to examine the variables that are most involved in the formation of the axes.

Definition 2.5.4

The contribution of a variable z^j to the k^{th} main axis is defined in the same way as for individuals. It is given by:

$$CTRB_k(z^j) = \frac{(r_{C^k,z^j})^2}{\sum_{l=1}^{p}(r_{C^k,z^l})^2} \times 100$$

The contribution of the 4 modules to the construction of the main axes is shown in the following table (TABLE 2.14).

	Dim.1	Dim.2	Dim.3	Dim.4
Algèbre	36.1806065	1.903649	25.0210280	36.894716
Analyse	36.8646675	7.255438	0.9183475	54.961547
Programmation	0.4751427	77.813526	17.2886721	4.422660
Module option	26.4795833	13.027387	56.7719525	3.721077

Table 2.14: Contribution of variables to each main axis (in %)

Following the first main axis, we take into account the modules Algebra, Analysis and Option Module which contribute the most and which are close to the edge of the correlation circle.

This is due to the percentage of the contribution of the three variables that exceeds the $\frac{1}{4}$ average and is worth 25%. On the other hand, the Programming module contributes the most to the construction of the second main axis with a percentage of 77.8%.

These results indicate that an individual will be affected by the variables for which he takes strong values. Conversely, it will be the opposite of the variables for which it takes low values.

Indeed, students are represented in the cloud of individuals following the direction of the variable (or module) where they got a good grade. On the contrary, students follow the opposite direction of a variable (or module), have had average or low scores.

In particular, Hafsa is represented in the opposite direction of the Analysis module. This comes from the fact that she had a bad grade in this module. Thus, since the Programming module contributes 77.8% in the construction of the second main axis and since Hafsa had a grade of 10.5, which is still the average of the grades taken by the students in this module, then it is represented close to the average O (by projecting along the second axis).

2.6 Representation of additional elements

When we do a Principal Component Analysis, it is practically common to consider additional variables or individuals (**illustratives**).

In this part, we deal first with the representation of additional quantitative variables, second with additional qualitative variables and finally with additional observations (or individuals).

2.6.1 Additional quantitative variables

Additional quantitative variables are represented in the correlation circle.

They do not contribute to the construction of the principle components.

Suppose we have an additional quantitative variable x^a and we want to represent it on the k^{th} main axis. To do this, simply project this x^a vector on the axis generated by v_k by calculating the correlation coefficient of x^a and the k^{th} principle component C^k.

Example 2.6.1

Consider in our example a new variable, given by attendance notes in practical programming sessions.

Student scores are given in the following table (TABLE 2.15).

	Algèbre	Analyse	Programmation	Module option	Présence
Marouane	15.25	13.80	12.0	11.5	15.5
Ziad	13.50	12.00	10.0	14.0	12.0
Yasmine	17.00	18.00	7.0	12.8	10.0
Issam	16.50	15.00	14.0	15.5	18.5
Hafsa	10.00	8.75	10.5	11.0	13.0
Oussama	13.00	9.00	10.0	9.5	12.0

Table 2.15: Additional variable added to the initial data set

In FIGURE 2.7, we represent the additional quantitative variable (Presence) in the correlation circle.

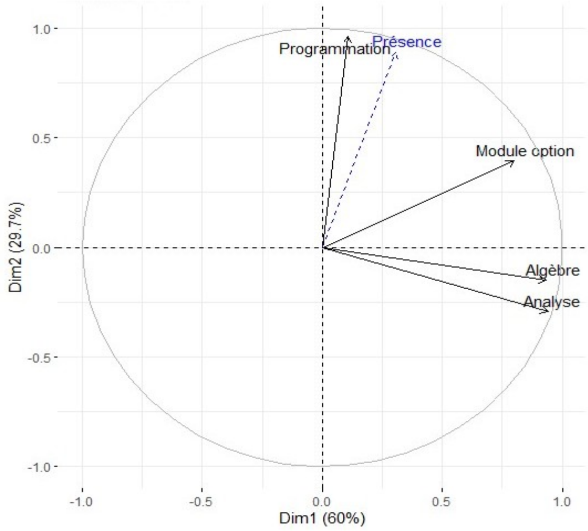

Figure 2.7: Correlation circle

The figure above allows us to see that the notes of the students in Programming are linked to their presence at the practical sessions.

In other words, the students who attend the practical sessions, all have good marks in Programming.

Conversely, students who did not successfully complete the Programming module missed practice sessions.

2.6.2 Additional qualitative variables

Qualitative variables have modalities taken by each individual.

For each modality, we calculate the barycenter of observations according to the k^{th} main axis, then we represent these modalities by points in the cloud of individuals.

Let xj, a an additional qualitative variable with modalities $x_{1,a}, x_{2,a}, \cdots, x_{r,a}$, for $j = 1, \cdots, p$.

Each individual x_i follows a modality $x_{l,a}$ of total n_l, for $i = 1, \cdots, n$ and $l = 1, \cdots, r$.

We then define the barycenter of an individual x_i that follows one of the modalities of the variable $x^{j,a}$.

Definition 2.6.1

The barycenter of an individual x_i that has a modality $x_{l,a}$, according to the k^{th} main axis is guven by:

$$\frac{1}{n_l}\sum_{i\in I} c_i^k \ , \forall l = 1, \cdots, r$$

where $I = \{i = 1, \cdots, n \ / \ x_i$ has the modality $x_{l,s}\}$ and c_i^k is the i^{th} individuals' coordinate according to the k^{th} main axis.

Example 2.6.2

In our example of grades, we can choose the additional qualitative variable as the gender of the students (Male or Female). The initial table becomes:

	Algèbre	Analyse	Programmation	Module option	Sexe
Marouane	15.25	13.80	12.0	11.5	M
Ziad	13.50	12.00	10.0	14.0	M
Yasmine	17.00	18.00	7.0	12.8	F
Issam	16.50	15.00	14.0	15.5	M
Hafsa	10.00	8.75	10.5	11.0	F
Oussama	13.00	9.00	10.0	9.5	M

Table 2.16: Initial data set with additional qualitative variable

The modality coordinates of the additional variable are given by TABLE 2.17.

	Dim.1	Dim.2	Dim.3	Dim.4
F	-0.2501744	-0.8571182	-0.3052197	0.18691428
M	0.1250872	0.4285591	0.1526099	-0.09345714

Table 2.17: Modality coordinates of the additional variable

As a result, modalities (M and F) are represented in the first main plane surface (FIGURE 2.8).

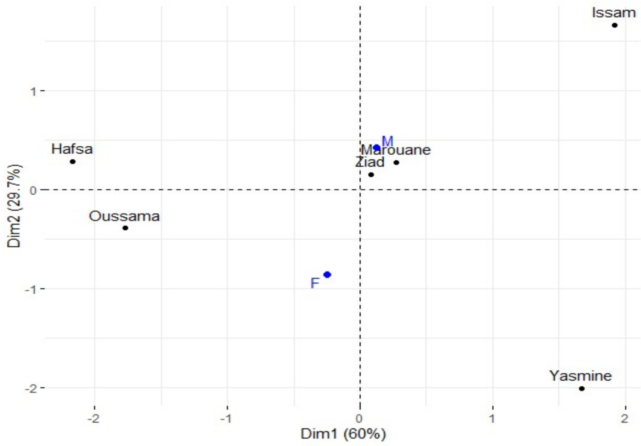

Figure 2.8: Individuals cloud and additional variable modalities

Based on the previous section (section 2.5) and FIGURE 2.8, we find that the majority of male students have near average scores in all modules.

In addition, Yasmine's grades affected the average of the students' grades. This explains the positioning of the F modality in the cloud of individuals following the point of the student Yasmine.

2.6.3 Additional observations

An additional individual or observation is an outlier or data that can distort the results of a statistical study. In another way, we can consider the additional individuals, as their name suggests, as new observations on which we want to apply a previously performed PCA.

Suppose we have an additional observation $x_{i,a}$ and we want to represent it according to the k^{th} main axis. For this we need only to project this point on the axis in question.

In other words, we calculate the scalar product of the individual $x_{i,a}$ with the v_k director vector of the k^{th} main axis.

Example 2.6.3
We consider two additional students Nabil and Khadija whose grades appear in the following table (TABLE 2.18).

	Algèbre	Analyse	Programmation	Module option
Nabil	6.5	16.0	8.00	13
Khadija	14.0	17.5	15.75	15

Table 2.18: Initial data set with additional individuals

We center and reduce the data in the above table and obtain the results given in TABLE 2.19. [2]

	Algèbre	Analyse	Programmation	Module option
Nabil	-2.96465109	0.9035923	-1.107424	0.2843982
Khadija	-0.08012446	1.3217066	2.214852	1.2067658

Table 2.19: Table of Centered-Reduced Data of Additional Individuals

Therefore, additional student coordinates according to the 4 main axes is given by TABLE 2.20.

	Dim.1	Dim.2	Dim.3	Dim.4
Nabil	-1.275767	-0.7762132	-2.2687833	2.391240
Khadija	1.673781	2.2394934	0.1076213	1.381936

Table 2.20: Coordinates of additional individuals according to the main axes

Finally, the representation of additional students in the cloud of individuals is given by FIGURE 2.9.

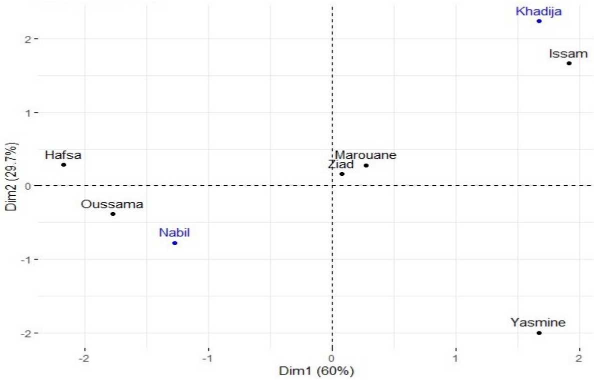

Figure 2.9: Cloud of individuals with additional students

We note that additional students are represented differently.

[2]The averages and standard deviations used in the pre-processing of centering and reduction are those in the initial table, given in section 2.1 by FIGURE 2.2 and FIGURE 2.5.

Indeed, Khadija has good grades in all modules, especially in Programming. This gives the impression that the student is represented close to the student Issam and closer to the second main axis, because of the grade of Issam in Progmmation and the important contribution of the module Programming in the construction of this axis.

On the contrary, Nabil is represented in the cloud of individuals close to Oussama by projecting them on the second main axis. This is because they did not get good marks in the Programming module. Thus, the projection of the two students according to the first main axis gives a difference between the two points. This is due to their notes in the modules Analysis, Algebra and Module option.

Chapter 3

Numerical application

In this last chapter, we apply the Principal Component Analysis method, which we presented previously, to a real example, using the R software.

The realization of this PCA on the real example, will be done by following the same steps given in the previous chapter.

First, we present the data set. Second, we build the principal components that we need for the construction of the main axes, in order to represent the cloud of individuals and the cloud of variables. We conclude with the interpretation of the results obtained.

Note 3.0.1
In this work, we will essentially use the two packages **"FactoMineR"** and **"factoextra"** in the R software.

3.1 Data sets

Over 28 years (from 1990 to 2017), we have the annual number of deaths in Morocco by reporting the different causes (diseases, road accidents, homicide, etc.).
(Hannah Ritchie and Max Roser, 2018)

The data are collected in TABLE 3.1, which crosses the 28 years (in rows) and the 25 causes of death (in columns).

	Méningite	M.CardVasc	Démence	M.rénales	M.respiratoires	M.foie	M.digestives	Hépatite	Cancers	M.Parkinson	Incendie	Noyade	Homicide
1990	2613.1718	68113.48	3409.848	2197.028	5062.737	2047.441	4220.234	419.3037	11665.68	477.1120	1312.4129	1262.3640	162.9377
1991	2472.6239	69982.79	3568.961	2248.075	5150.595	2083.032	4260.302	413.7064	12023.59	495.2448	1291.1449	1223.1060	166.8475
1992	2357.2556	72069.75	3728.750	2308.440	5250.235	2122.535	4314.267	409.8701	12393.88	514.1185	1276.9145	1185.8210	173.6411
1993	2270.4595	74488.41	3883.469	2383.024	5376.122	2166.869	4373.922	409.8000	12813.21	533.1681	1275.4919	1166.4320	186.0940
1994	2194.1139	76062.30	4024.732	2431.381	5447.028	2205.767	4417.180	406.4645	13148.10	550.4030	1267.2684	1139.8844	199.8480
1995	2149.2736	78357.05	4204.793	2499.681	5587.333	2252.526	4479.152	406.1926	13541.11	575.4659	1267.8399	1134.0867	222.4213
1996	2069.2916	80373.59	4413.382	2565.158	5729.854	2298.301	4532.668	405.0873	13929.63	605.4306	1259.3373	1120.4567	252.8850
1997	1994.5001	82917.10	4613.138	2644.070	5899.275	2352.289	4601.581	407.6282	14415.09	634.7471	1279.4252	1090.1026	290.5563
1998	1946.2904	83779.00	4737.499	2672.937	5958.158	2384.683	4625.274	405.9216	14720.13	651.7071	1228.0708	1056.7310	332.1024
1999	1817.5908	85487.05	4908.197	2745.740	6056.514	2419.706	4668.085	408.7345	15153.41	676.1720	1203.5353	1020.4925	376.8501
2000	1739.3265	85073.27	4981.039	2763.914	5998.420	2410.790	4633.504	405.6150	15332.01	687.8187	1158.5837	973.6299	399.2823
2001	1665.9295	85635.37	5073.662	2824.210	6003.596	2418.924	4635.284	405.6140	15672.61	704.7916	1125.7363	934.8643	423.9390
2002	1611.4680	86236.11	5230.037	2898.838	6013.198	2442.258	4677.436	403.1456	16015.40	728.2437	1142.2994	899.7393	428.8540
2003	1546.2682	87864.02	5433.038	2995.548	6125.297	2478.416	4741.529	403.5337	16490.62	762.7233	1063.8669	865.6297	433.7005
2004	1463.1429	89228.55	5615.953	3082.939	6210.606	2506.066	4777.577	402.0412	16905.17	791.9819	1033.1322	825.8705	433.6052
2005	1387.9424	91086.89	5832.290	3186.474	6337.725	2541.227	4827.700	401.7915	17369.35	827.7822	1006.4589	790.7448	431.2093
2006	1308.0923	93463.48	6064.130	3305.473	6503.069	2576.744	4871.404	402.1084	17872.36	869.7586	981.1870	750.9388	427.1719
2007	1083.8619	95397.27	6314.258	3424.632	6601.059	2628.613	4947.123	403.2743	18430.20	903.5865	962.7053	724.3995	428.0004
2008	1032.5038	96996.39	6519.595	3536.142	6675.597	2671.008	5002.028	404.1532	18948.81	933.2104	949.5312	713.2213	424.7323
2009	1017.4787	98296.68	6736.117	3647.611	6720.645	2731.217	5088.194	405.9495	19508.53	959.6418	948.7942	726.0476	424.0180
2010	981.0829	99901.92	6994.116	3764.490	6791.072	2787.941	5167.707	406.6074	20074.48	991.8480	936.9483	719.2943	422.8038
2011	952.7313	101789.14	7254.256	3881.497	6895.518	2848.283	5250.368	407.3246	20677.36	1027.2725	924.7383	711.1858	421.3180
2012	882.5301	103116.28	7567.467	3980.958	6937.446	2921.824	5360.013	406.6320	21275.13	1053.7961	906.8102	695.5654	420.9434
2013	828.2160	105336.94	7911.301	4097.866	7071.558	2990.116	5459.569	407.2595	21909.34	1094.2639	889.6246	672.9687	418.6134
2014	817.3917	107655.93	8259.669	4224.341	7221.337	3072.862	5584.867	410.1256	22592.68	1135.6850	890.0438	677.3846	418.4486
2015	780.8404	109867.91	8612.358	4340.255	7357.274	3143.615	5688.385	411.1349	23231.12	1175.9927	877.1500	660.1984	416.3797
2016	747.9474	112449.60	8981.755	4466.399	7523.635	3220.645	5804.883	413.0087	23895.93	1219.7158	864.6482	644.9237	413.4907
2017	715.6991	115124.03	9342.936	4543.652	7679.800	3298.015	5931.750	416.1090	24504.90	1260.7696	850.7343	626.2136	409.9523

	VIH-SIDA	Drogue	Tuberculose	Blessures_routes	Alcoolisme	Cata.nature	M.diarrhéiques	Chaleur	Déf.nutri	Suicide	Diabète	Empoisonnement
1990	102.6213	156.4818	8179.234	8630.540	67.52394	0.0000	10568.070	67.41869	284.4891	1743.509	3216.645	424.4564
1991	126.6168	168.8346	7991.330	8607.926	70.73605	0.0000	9449.497	67.62764	269.3666	1793.666	3333.962	420.6562
1992	154.5161	181.6036	7899.289	8577.386	73.96767	0.0000	8500.347	68.20392	253.4120	1848.150	3466.788	418.3754
1993	186.3237	196.9084	7892.195	8632.545	77.60691	0.0000	7716.551	68.94641	239.5743	1922.655	3613.961	420.2818
1994	221.7210	209.1156	7743.423	8550.356	80.16097	0.0000	7007.127	68.08474	231.4778	1965.348	3717.234	420.4150
1995	260.6550	224.4191	7671.296	8575.198	83.17964	791.0000	6418.918	67.68120	224.3007	2022.683	3847.195	424.6940
1996	302.3606	241.6860	7498.452	8592.586	86.29873	39.0000	5851.457	67.18313	223.8907	2114.020	3983.918	426.0528
1997	347.5995	261.1678	7405.914	8598.165	89.88805	60.0000	5350.968	66.50966	217.6318	2215.880	4159.285	427.4263
1998	397.4086	278.3022	7196.543	8553.039	92.79060	0.0000	4888.443	65.80822	207.6319	2277.104	4243.464	425.9507
1999	452.5084	301.6064	7072.663	8524.796	93.65098	0.0000	4489.608	66.15376	204.8648	2378.528	4429.548	430.2098
2000	513.4181	321.9221	6819.227	8388.157	92.24055	6.0000	4158.920	66.30804	198.1142	2443.065	4545.136	433.6762
2001	579.4767	345.5173	6651.502	8327.226	92.55057	15.0000	3824.367	66.57107	190.2621	2494.147	4706.453	436.4429
2002	650.7402	360.1614	6481.024	8160.729	92.21823	80.0000	3557.687	67.15373	185.1399	2514.957	4874.058	435.2367
2003	702.0753	375.8362	6390.871	8042.533	93.05399	35.0000	3294.348	67.45341	179.1093	2532.140	5076.415	432.8566
2004	724.0640	391.6431	6269.669	7898.893	94.33384	628.0001	3023.000	67.44988	169.3440	2540.375	5279.320	428.1391
2005	753.5192	408.4522	6176.879	7775.263	96.21336	3.0000	2780.694	67.30368	160.8578	2548.867	5501.167	423.9636
2006	798.4886	426.2963	6096.508	7609.751	98.31496	17.0000	2494.518	66.65203	153.9387	2558.760	5731.467	418.1029
2007	840.0950	447.0165	5995.504	7541.420	101.09089	0.0000	2229.200	66.71306	144.0683	2594.012	5954.560	414.3592
2008	861.8822	467.8056	5880.673	7513.786	103.81112	39.0000	2033.323	66.43622	140.4856	2623.342	6154.941	413.0904
2009	874.3410	488.0326	5747.502	7554.570	106.57783	30.0000	1928.496	66.53624	138.3309	2649.062	6347.521	417.6161
2010	891.1155	509.0350	5608.582	7537.396	109.49952	42.0000	1773.492	66.16846	134.1939	2672.516	6550.522	416.0665
2011	926.1205	530.0691	5480.819	7519.154	112.46183	0.0000	1635.634	65.80196	130.3721	2653.812	6758.355	414.0591
2012	907.5586	549.5831	5314.797	7478.995	115.18911	0.0000	1446.387	65.22509	122.7152	2638.890	6948.005	408.7572
2013	848.3426	571.2070	5202.565	7412.117	118.12568	0.0000	1294.098	64.53547	117.1692	2620.200	7171.180	403.2194
2014	778.0617	594.0638	5120.271	7417.521	121.16714	60.0000	1240.052	64.37809	116.9905	2604.689	7402.115	406.5954
2015	720.7532	617.0505	5022.072	7364.261	124.10875	0.0000	1140.491	63.67210	112.8122	2585.155	7619.746	403.2006
2016	653.6933	641.3559	4943.794	7320.465	127.08556	0.0000	1058.353	63.01441	111.2332	2566.498	7864.307	399.7751
2017	602.4716	663.8751	4882.782	7263.557	129.70783	0.0000	1022.393	62.57203	109.5787	2574.463	8062.255	395.0883

Table 3.1: Initial data set

Note 3.1.1

The table below gives the meaning of some causes of death presented in TABLE 3.1.

Causes	Signification
M.CardVasc	Cardiovascular diseases
M.rénales	Kidney diseases
M.respiratoires	Respiratory diseases
M.foie	Liver diseases
M.digestives	Digestion diseases
M.Parkinson	Parkinson's disease
Blessures_ routes	Traffic
Cata.nature	Natural disasters
M.diarrhéiques	Diarrhoea
Chaleur	Heat (cold or hot)
Déf.nutri	Nutritional deficiency

For data manipulation, we choose to center and reduce the variables, in order to construct the Z table of centered-reduced data.

Note 3.1.2

In this example, we study variables of the same nature, which allows us to limit ourselves to centering the variables without reducing them.

We use the **scale()** function of the R software to get the Z table of centered-reduced data. In addition, the **scale()** function returns the averages and standard deviations of each variable.

```
> Z <- scale(X)
```

The table of variable averages is given by TABLE 3.2.

```
   Méningite      M.CardVasc     Démence  M.rénales M.respiratoires      M.foie    M.digestives        Hépatite     Cancers
  1515.96515     90576.79568  5864.88376 3202.17050      6292.31083  2572.20410      4890.78523       407.43347 17303.92238
 M.Parkinson        Incendie      Noyade   Homicide        VIH-SIDA      Drogue     Tuberculose Blessures_routes  Alcoolisme
   815.80183      1077.65835   893.29633  355.73736       577.80529   390.32315      6451.26342       7998.86903    97.98408
 Cata.nature  M.diarrhéiques     Chaleur  Déf.nutri         Suicide     Diabète Empoisonnement
    65.89286      3934.87286    66.34151  177.54842      2382.01766  5377.12577       419.24155
```

Table 3.2: Table of averages

The variables standard deviations are given by TABLE 3.3.

Méningite	M.CardVasc	Démence	M.rénales	M.respiratoires	M.foie	M.digestives	Hépatite	Cancers
596.855531	13155.640036	1728.477584	731.873278	730.784382	352.889983	480.359564	4.269576	3844.419086
M.Parkinson	Incendie	Noyade	Homicide	VIH-SIDA	Drogue	Tuberculose	Blessures_routes	Alcoolisme
237.900577	163.260660	208.652547	100.565594	272.023711	155.745373	1056.383783	522.739121	17.016560
Cata.nature	M.diarrhéiques	Chaleur	Déf.nutri	Suicide	Diabète	Empoisonnement		
184.549781	2738.042855	1.544794	52.062186	299.450321	1506.977539	10.948526		

Table 3.3: Table of standard deviations

Note 3.1.3
We make a PCA using the R correlation matrix given by the matrix form: $R = \frac{1}{28} Z^T Z$

We can visualize the correlation matrix using the **corrplot()** function of the **"corrplot"** package of the R software. We then obtain the following result given by FIGURE 3.1.

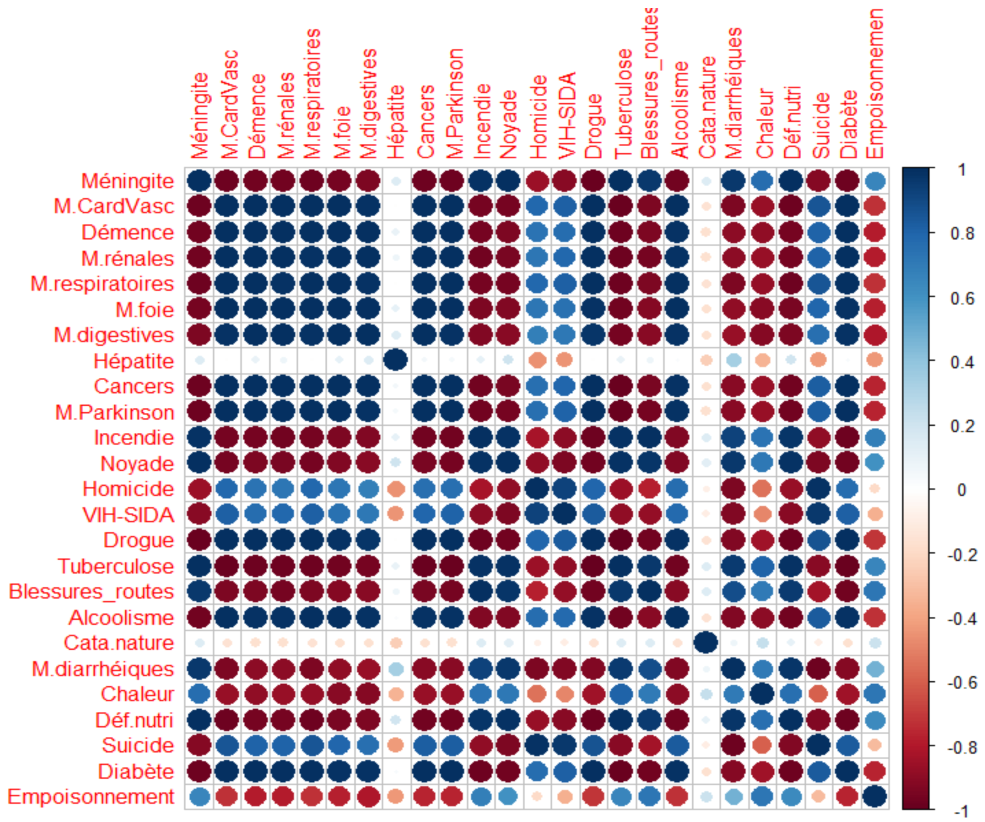

Figure 3.1: Visualisation of correlation matrix

3.2 Main axes

The construction of the main axes begins with the diagonalzsation of the R correlation matrix and the obtaining of the corresponding eigenvalues and eigenvectors, allowing the construction of the principal components.

3.2.1 Eigenvalues and eigenvectors of the correlation matrix

The correlation matrix R, given by FIGURE 3.1, being symmetrical, is then diagonalizable. So we can write:

$$R = PDP^{-1}$$

where D is the diagonal matrix of eigenvalues $\lambda_1, \cdots, \lambda_{25}$ of the R matrix, and P is the passage matrix containing eigenvectors associated with eigenvalues of the correlation matrix.

Using the **eigen()** function in R software, we can extract eigenvalues and eigenvectors of the R matrix.

We list eigenvalues and eigenvectors using the following commands respectively:

```
> R_Valeurs_propres <- eigen(R)$values
```

```
> R_Vecteurs_propres <- eigen(R)$vectors
```

3.2.2 Principal components

The principal components $C^k (k = 1, 2, \cdots, 25)$ are given by the formula:

$$C^k = Zv_k$$

where Z is the centered-reduced data matrix and v_k is the k^{th} directional vector, associated with the $lambda_k$ eigenvalue of the correlation matrix R, for $k = 1, 2, \cdots, 25$.

The calculation of the 25 principal components is done in the R software using the **PCA()** function of **FactoMineR** package.

To access the principal components, we write:

```
> Comp <- PCA(X)$ind$coord
```

The princial components, containing the coordinates of the individuals according to the first 5 main axes, are given by TABLE 3.4.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
1990	-7.70416614	3.011287817	-0.264483429	-0.010106246	1.45526347
1991	-7.03341371	2.149351168	-0.240651245	-0.471165705	0.43103834
1992	-6.44655989	1.435200237	-0.229362708	-0.895930254	-0.23541329
1993	-6.00847364	1.091861781	-0.260383963	-0.893854862	-0.21961781
1994	-5.37313254	0.611795782	-0.278985636	-0.709978567	-0.80059760
1995	-5.02334693	-0.478008891	3.880757959	0.101828730	-0.23913664
1996	-4.24639688	0.004391253	-0.187214491	0.193440169	-0.91021170
1997	-3.55091455	0.196212312	-0.125740524	0.851596754	-0.50974206
1998	-2.82047900	-0.091021609	-0.499818701	0.978991476	-0.73720248
1999	-2.28223295	-0.186758427	-0.619713319	1.295485305	0.07534748
2000	-1.89643185	-1.078828139	-0.715569640	1.110321625	0.04266982
2001	-1.46934174	-1.467041224	-0.758227295	1.096281707	0.38238554
2002	-1.10184598	-2.053335422	-0.421763318	0.623541407	0.21550313
2003	-0.47162402	-1.983791703	-0.627625229	0.221285883	0.38718366
2004	0.01730553	-2.689509523	2.604893466	0.008860709	0.52432130
2005	0.77850080	-1.891842746	-0.649059894	-0.515813529	-0.04236846
2006	1.56063167	-1.553293997	-0.473442735	-0.711687648	-0.14778299
2007	2.26606076	-1.286858497	-0.497131280	-0.925997072	-0.05595008
2008	2.76200742	-1.084999551	-0.237853828	-0.838145967	0.02843429
2009	3.05041219	-0.920190556	-0.284990689	-0.481347682	0.36619555
2010	3.58707299	-0.668316289	-0.156454360	-0.367433109	0.35949126
2011	4.14486947	-0.304665659	-0.299443278	-0.325467241	0.30637726
2012	4.78275080	-0.004648690	-0.169642939	-0.404863001	-0.08576272
2013	5.45740769	0.585424885	-0.006931799	-0.385113208	-0.33745060
2014	5.82504959	1.091285610	0.396508039	0.123206165	0.04053587
2015	6.45239319	1.740225612	0.219252802	0.265662724	-0.14356410
2016	7.06946351	2.476697838	0.375037205	0.475045105	-0.16950561
2017	7.67443422	3.349376628	0.528040828	0.591356334	0.01955919

Table 3.4: Coordinates of individuals according to the first 5 main axes

3.2.3 Choice of principal components

In section 2.4.3, we discussed two criteria for selecting the principal components.

The *Kaiser* criterion allows us to select the first two principal components, whose variances are worth $s^2_{C^1} = \lambda_1 = 20,85$ and $s^2_{C^2} = \lambda_2 = 2,44$ respectively, which are above the average value of 1.

On the other hand, the *Cattell* criterion (or the scree test) allows us to choose only the first principal component. This is due to the fact that the structure of the eigenvalue graph (scree plot) changes starting from the second eigenvalue. This forces us to choose the first eigenvalue above the horizontal line passing through the first elbow. (See FIGURE 3.2)

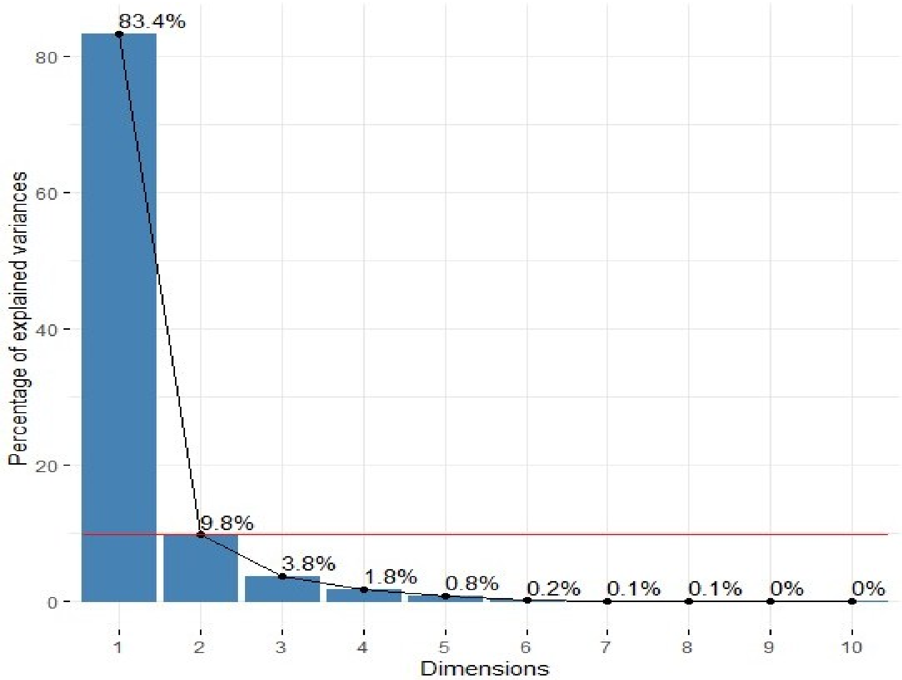

Figure 3.2: Scree plot

3.3 Representation of individuals

The representation of individuals according to the criterion of *Cattell* is an axial representation, choosing the first principal component. This representation is given by the first main axis explaining 83.4% of the information.

However, the criterion of *Kaiser* allows to obtain a flat representation, choosing the first two principal components, explaining 93.2% of the information.

Gr ace to **fviz_ pca_ ind()**, we represent the individuals in the main plane surface, using the command below.

```
> fviz_pca_ind(PCA(X),repel=TRUE)
```

We obtain the cloud of individuals given by FIGURE 3.3.

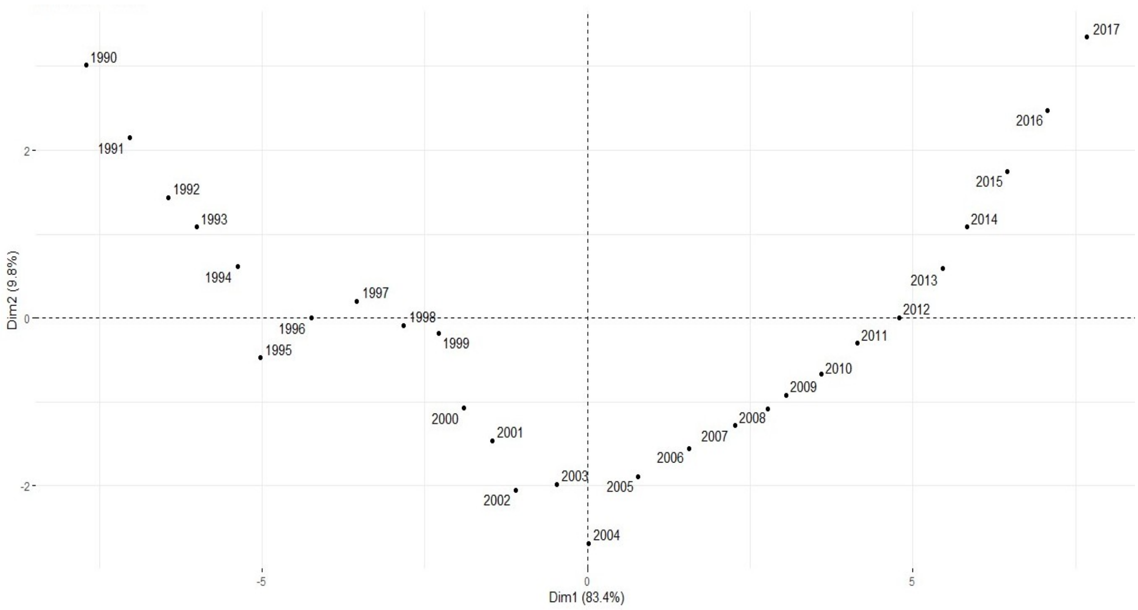

Figure 3.3: Cloud of individuals

3.4 Representation of variables

Causes of death (variables) are represented in the correlation circle.

The coordinates of the variables are given, using the function **get_ pca_ var()**, by the command:

```
> get_pca_var(PCA(X))$coord
```

The **fviz_ pca_ var()** function of the **factoextra** package allows us to construct the correlation circle of the 25 variables, given by FIGURE 3.4.

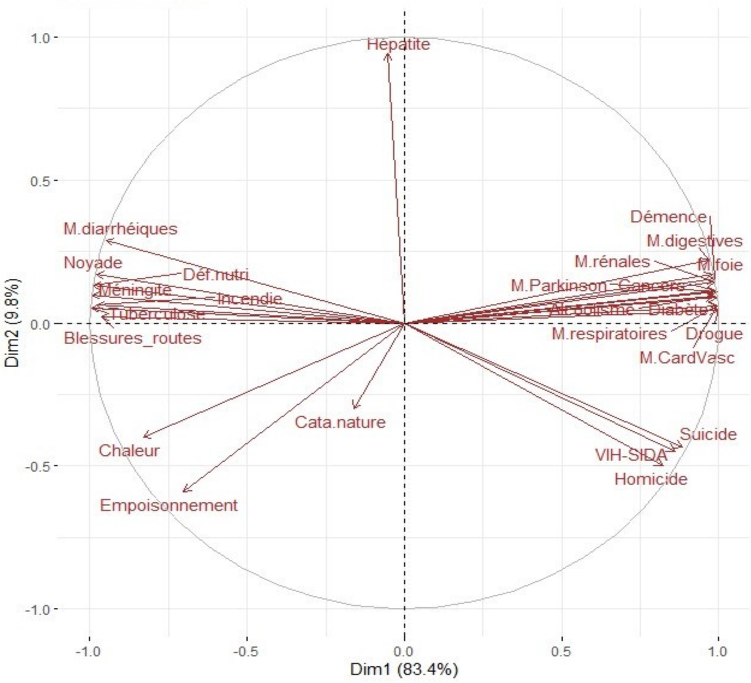

Figure 3.4: Correlation circle

3.5 Results interpretation

The main axes provide an approximate image of the point cloud. It is therefore necessary to calculate the quality of representation of individuals as well as that of variables.

In addition, it is important to calculate the contribution of individuals and variables in the construction of the main axes.

3.5.1 Individuals interpretation

According to the formulas given in section 2.5.1 of the previous chapter, we calculate the quality of representation of the 28 years according to the first 5 main axes as well as their contribution in the construction of these axes.

These qualities of representation are calculated by the software R using the command:

```
> PCA(X)$ind$cos2
```

We use the **corrplot()** function to visualize the qualities of representation, and we obtain the following result given by FIGURE 3.5.

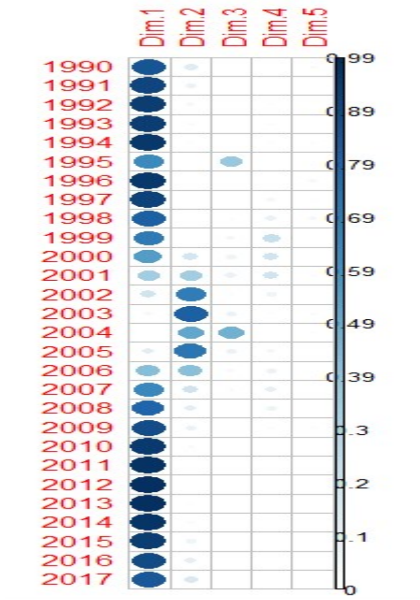

Figure 3.5: Visualization of the qualities of representation for individuals

The first 11 years and the last 11 years are well represented according to the first axis, while the years from 2001 to 2006 are poorly represented according to the same axis. In contrast, they are well represented in the second main axis compared to other years.

The other tool to be used in the interpretation of individuals is the contribution of years in the construction of the main axes.

The contributions of individuals are given in the R software by the command:

```
> PCA(X)$ind$contrib
```

The visualization of individuals' contributions is given by FIGURE 3.6.

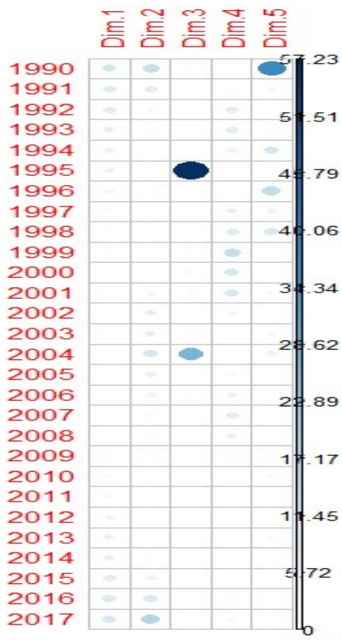

Figure 3.6: Visualization of individuals' contributions

It is clear that year 1995 contributed the most to the construction of the third main axis, with a contribution exceeding 50%, followed by 2004 with a percentage of more than 20%.

On the other hand, years with good representation on the first or second main axis (according to FIGURE 3.5) contribute with a percentage of less than 10% in the construction of the third axis.

3.5.2 Variables interpretation

For the interpretation of variables we use the same approach as that used for the interpretation of individuals.

We calculate, from the formulas given in section 2.5.2, the quality of the representation of the variables or the cosine squared of the angle between the variables and the main axes.

Cosine squared variables are given by the following command:

```
> PCA(X)$var$cos2
```

Les résultats sont présentés par FIGURE 3.7.

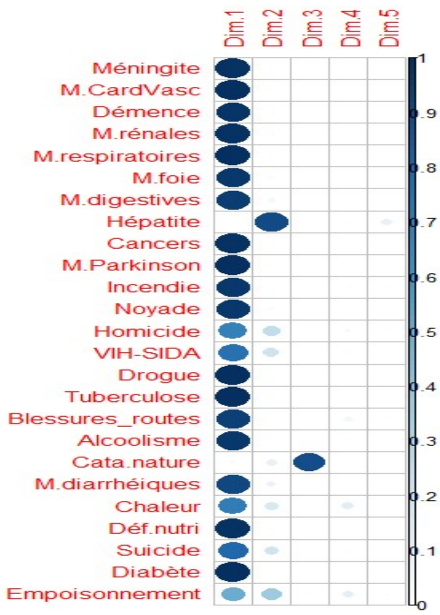

Figure 3.7: Visualization of cosine squared for variables

We note that all variables are well represented on the first main axis except for the two variables "Hépatite" (Hepatitis) and "Cata.nature" (Natural disasters), which are well represented on the second and third main axis, respectively.

In same way, we obtain the contributions of the variables using the following command:

```
> PCA(X)$var$contrib
```

The figure below allows us to visualize the different contributions.

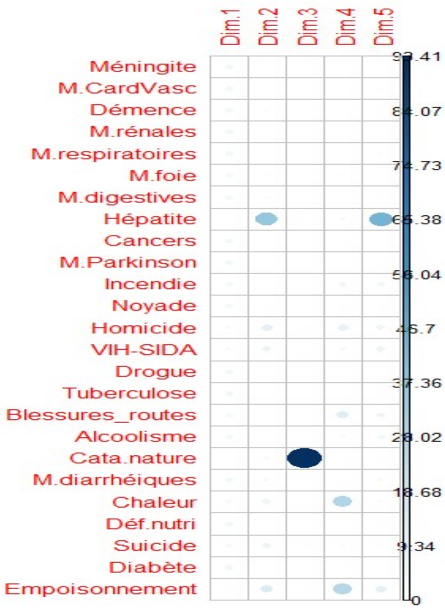

Figure 3.8: Visualization of variables contributions

FIGURE 3.8 gives us an average contribution of the variable "Hépatite" (Hepatitis) to create the second main axis. Thus, we note that all other variables have a small contribution in the construction of the first main axis.

In the creation of the third axis, we obtain a large contribution from the variable "Cata.nature" (Natural disasters). This allows us to say that the third main axis is related to natural disasters.

Thus, the first main axis is composed of most of the causes of death (variables) except some, which are contributing the most in the creation of second and third main axis ("Hépatite" and "Cata.nature" respectively).

In the next section, we explain the results obtained concerning the quality of representation and contribution of individuals, as well as those of variables.

3.5.3 Broad interpretation

The point cloud of individuals and the correlation circle give the representation of individuals and variables respectively in a separate way.

Fortunately, the R software gives us the ability to represent individuals and variables in a single graph, called **Biplot**. To achieve this double representation, we use the **fviz_ pca_ biplot()** function as follows:

```
> fviz_pca_biplot(PCA(X), repel=TRUE)
```

We obtain the double representation of individuals and variables in the first main plane, given by FIGURE 3.9.

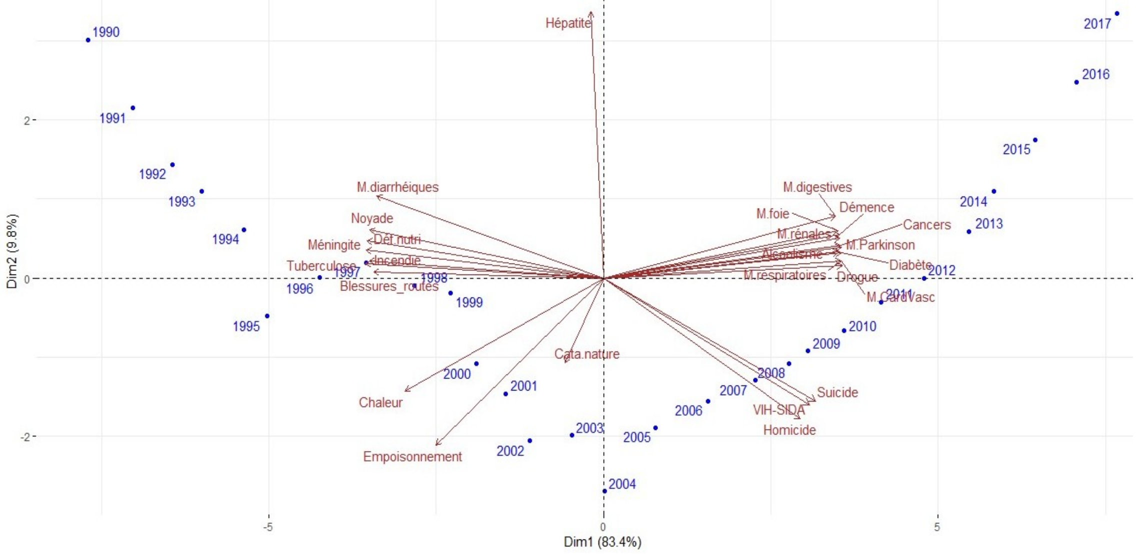

Figure 3.9: Biplot of individuals and variables in the first main plane

This double representation makes it easier for us to interpret the results obtained.

Indeed, we notice a separation of variables and individuals in the 4 sides of the main plan. This gives us information on the causes of death in the four periods: 1990-1999, 2000-2004, 2005-2011 and 2012-2017.

In the first period (1990-1999), causes of death took the form of diarrhea ("M.diarréhiques"), drowning ("Noyade"), nutritional deficiency, meningitis ("Méningite"), tuberculosis ("Tuberculose"), fires ("Incendies"), road accidents and hepatitis ("hépatite").

The second period (2000-2004), was not excluded from the diseases of the first period, but Morocco has known new causes with a significant number of deaths, we quote: heat, poisoning ("Empoisonnement") and natural disasters.

Thus, according to FIGURE 3.7, the "Cata.nature" variable is not well represented in the first main plane, but it is very well represented according to the third main axis.

In addition, FIGURE 3.6 shows that of the 28 years, the only years with a significant contribution, according to the third main axis, are 1995 and 2004. We would be interested to know the cause of this difference.

To do this, we use the double representation of individuals and variables according to the second main plane, generated by the first and third main axis, given by FIGURE 3.10.

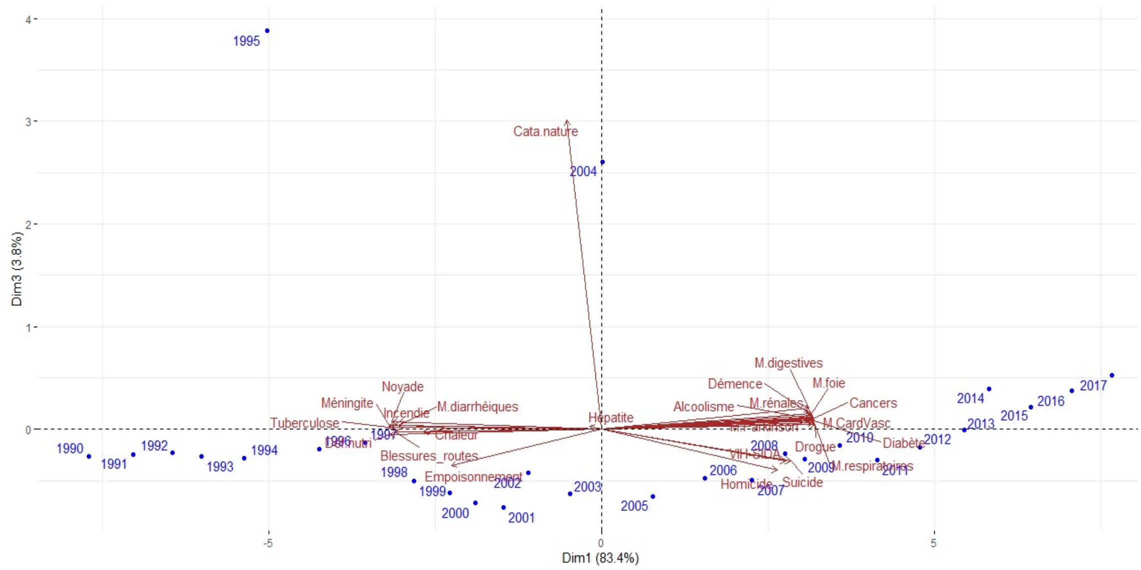

Figure 3.10: Biplot of individuals and variables in the second main plane

In the second main plan, we can clearly distinguish the two years 1995 and 2004 from the other years, as well as the variable "Cata.nature" from the other variables.

This result comes from the fact that Morocco experienced natural disasters during 1995 and 2004.

At first, on the night of 17 August 1995, the flood of the wadi Ourika had washed away dozens of cars killing more than 200 people.

Second, February 24, 2004, was the day when the town of Al Hoceima was hit by an intense earthquake that cost the lives of more than 600 people.

In the third period (2005-2011), Morocco experienced the spread of three causes of death: homicide, suicide and AIDS ("SIDA").

Finally, the fourth period (2012-2017) recognized a transformation of causes of death.

Indeed, the impact of the causes of death in the 90s has decreased, with the emergence of new causes and diseases such as digestive diseases, cardiovascular diseases, drugs and cancers. Thus, the causes of death cited in the third period contribute to the increase in the number of deaths in the fourth period.

The causes of death from 1990 to 2017 have changed significantly. We cite, in the table below, the change of some causes.

Death causes	In 1990	In 2017	Change (in %)
Dementia	3 409,85	9 342,94	+ 174%
Parkinson's disease	477,11	1 260,77	+ 164,25%
Diabetes	3 216,64	8 062,25	+ 150,6%
Drugs ("Drogue")	68 113,48	11 5124,03	+ 324,25%
Cancer	11 665,68	24 504,9	+ 110%
HIV-AIDS	102,62	602,47	+ 487%
Meningitis ("Méningite")	2 613,17	715,7	- 72,61%
Nutritional deficiency	284,49	109,58	- 61,48%
Diarrhoeal disease	10 568,07	1 022,39	- 90,3%
Hepatitis ("Hépatite")	419,30	416,11	- 0,76%

The data in the above table explains a huge change in the number of deaths in Morocco over the 28 years. This makes it possible to say that Morocco has been able to control certain causes (meningitis, nutritional deficiency, diarrhoeal diseases), while others have increased the number of deaths by a significant percentage such as dementia, Parkinson's disease, diabetes, drugs and HIV-AIDS. This means that Morocco must be wary of the spread of certain causes through the development of the health system and the development of scientific research.

Conclusion

In this paper, we presented some general information on the Principal Component Analysis (PCA) method.

This method makes it possible to study a multivariate data set of any size and to give a graphical representation of it.

It offers, in a few mathematical operations, the existing relationships between the variables of study.

This flexibility of use translates into the diversity of applications of Principal Component Analysis, which affects all sectors (economics, biology, medicine, etc.).

As a method of data analysis, Principal Component Analysis applies to specific cases. The variables to be studied must be quantitative, which limits the use of this process.

On the other hand, Principal Component Analysis is performed on correlated variables, which is not always available in practice.

After all, Principal Component Analysis remains an **important** method in processing a quantitative data set, **practical** in many application areas and **simple** when handling such data.

Bibliography

[1] Baccini, A., 2010. Statistique Descriptive Multidimensionnelle (pour les nuls). *Institut de Mathématiques de Toulouse-UMR CNRS*, 5219.

[2] Escofier B., 2008. Analyses Factorielles Simples et Multiples: Objectifs, Méthodes Et Interprétation. $4^{ème}$ édition, Dunod, Paris, France, 328p.

[3] Jauregui J., 2012. Principal component analysis with linear algebra. *Philadelphia: Penn Arts & Sciences.*

[4] Ritchie H et Roser M : Causes of Death. [CSV] (Décembre 2019), disponible sur: https://ourworldindata.org/causes-of-death, page consultée le 28/05/2020.

Appendix

Proof
Let $A \in M_p(\mathbb{R})$. The application f_A is defined by:

$$\forall x \in \mathbb{R}^p, \quad f_A(x) =< x, Ax >$$

Let $x, h \in \mathbb{R}^p$.

$$\begin{aligned} f_A(x+h) - f_A(x) &=< x+h, A(x+h) > - < x, Ax > \\ &=< x, Ah > + < h, Ax > + < h, Ah > \\ &=< A^T x, h > + < Ax, h > + < h, Ah > \\ &=< (A^T + A)x, h > + < h, Ah > \\ &= L(h) + \theta(h) \end{aligned}$$

With $L(h) =< (A^T + A)x, h >$ is a continuous linear application and $\theta(h) =< h, Ah >$.

Indeed, for all $h \in \mathbb{R}^p$:

$$\begin{aligned} | < (A^T + A)x, h > | &\leq ||(A^T + A)x|| \, ||h|| \quad (\textit{Cauchy-Schwartz}) \text{ inequality} \\ &= K||h|| \end{aligned}$$

where $K = ||(A^T + A)x||$.

On the other hand, *Cauchy-Schwartz* inequality gives:

$$|\theta(h)| = | < h, Ah > | \leq ||A||||h||^2$$

Then: $\lim\limits_{||h|| \longrightarrow 0} \dfrac{|\theta(h)|}{||h||} = 0$

So, the application f_A is differentiable and its diferential is defined as:

$$\forall x \in \mathbb{R}^p, \forall h \in \mathbb{R}^p, \ df_A(x)(h) = L(h) =< (A^T + A)x, h >$$

As a result :

$$\forall x \in \mathbb{R}^p, \ df_A(x) = (A^T + A)x$$