Principal Component Analysis (PCA)

Master MAS, Université de Bordeaux

18 septembre 2019

Introduction

The aim is to explore numerical data.

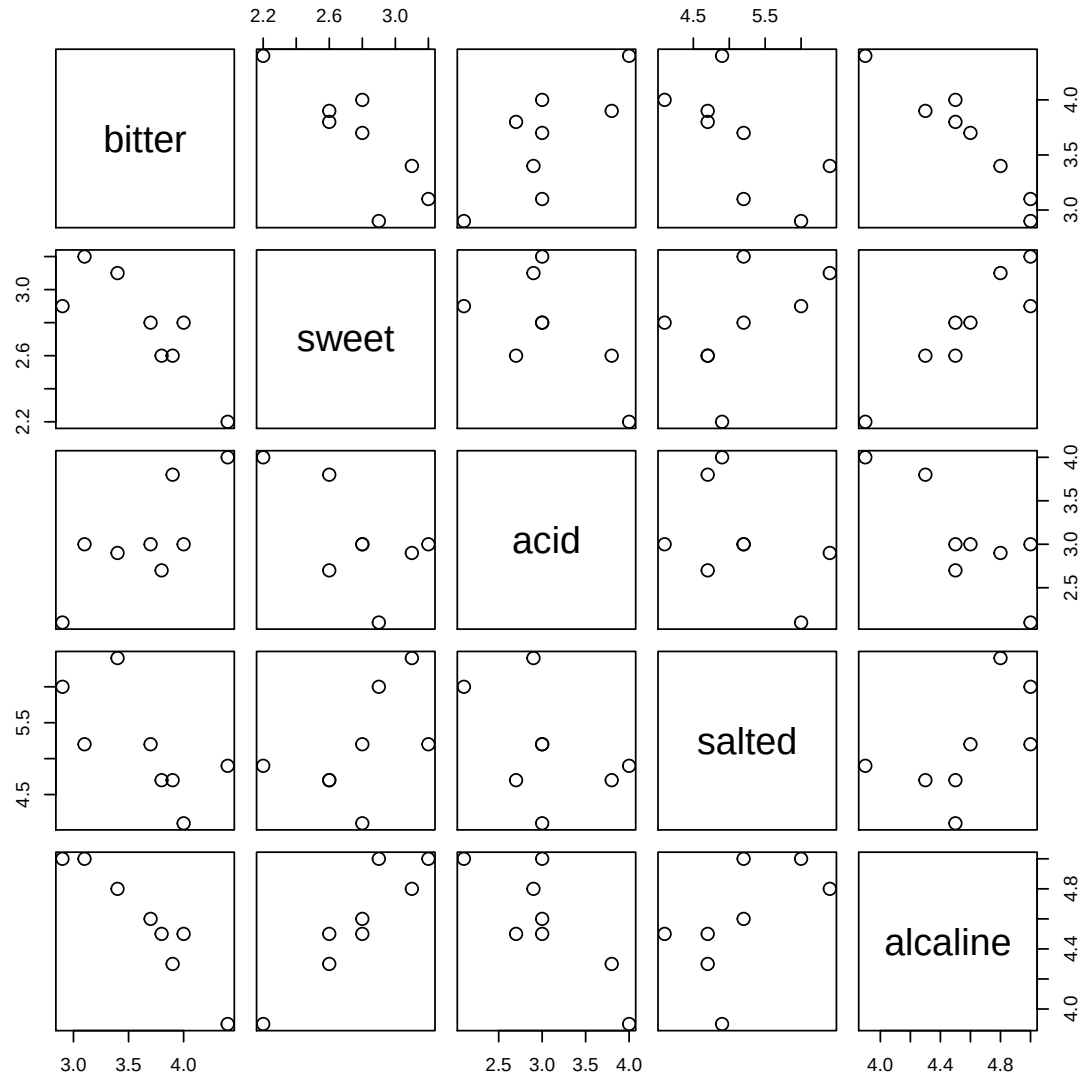
Example : 8 mineral waters described on 13 sensory descriptors.

```
##             bitter sweet acid salted alcaline
## St Yorre       3.4   3.1  2.9    6.4      4.8
## Badoit         3.8   2.6  2.7    4.7      4.5
## Vichy          2.9   2.9  2.1    6.0      5.0
## Quézac         3.9   2.6  3.8    4.7      4.3
## Arvie          3.1   3.2  3.0    5.2      5.0
## Chateauneuf    3.7   2.8  3.0    5.2      4.6
## Salvetat       4.0   2.8  3.0    4.1      4.5
## Perrier        4.4   2.2  4.0    4.9      3.9
```

The rows describe observations or individuals (the 8 mineral waters) and columns describe variables (the sensory descriptors).

The aim is to know :

- which observations are similar,
- quelles variables are linked.


bitter
sweet
acid
salted
alcaline
2.2
2.6
3.0
4.5
5.5
3.0
3.5
4.0
2.2
2.6
3.0
2.5
3.0
3.5
4.0
4.5
5.5
4.0
4.4
4.8
3.0
3.5
4.0
2.5
3.0
3.5
4.0
4.0
4.4
4.8

One can look at :

- the distance matrix between observations :

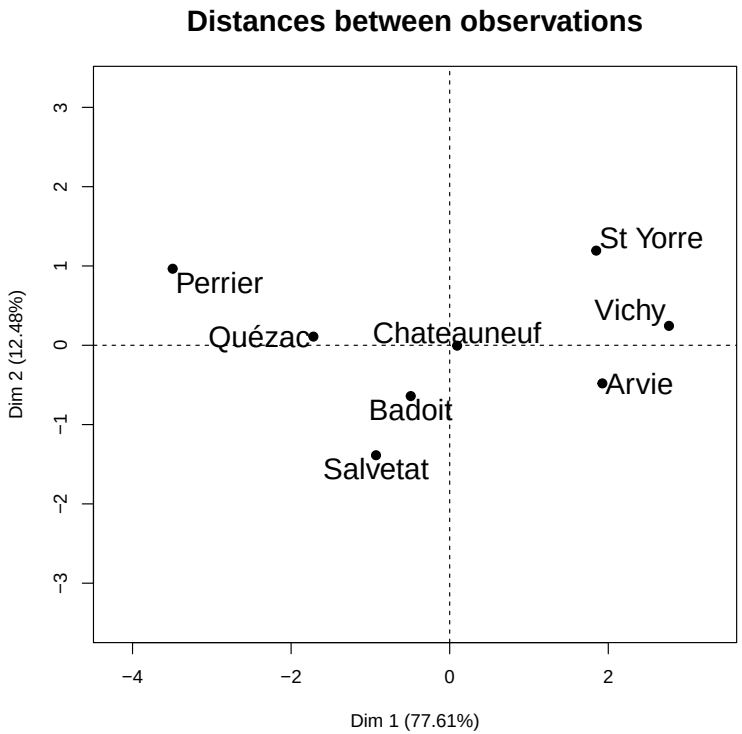
```
##             St Yorre Badoit Vichy Quézac Arvie Chateauneuf Salvetat
## Badoit           4.1
## Vichy            7.9    4.8
## Quézac           2.9    5.3   9.7
## Arvie            3.0    1.8   5.5    4.7
## Chateauneuf      2.9    1.8   5.7    4.3   1.3
## Salvetat         4.0    1.2   5.4    4.9   1.8         1.6
## Perrier          8.2   10.6  14.7    6.2  10.1         9.9     10.3
```

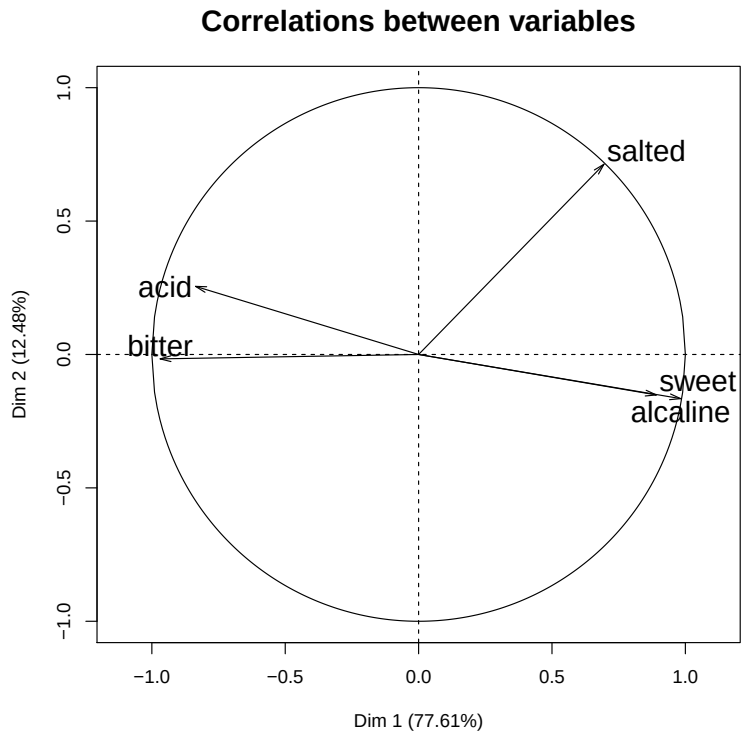
- the correlation matrix between variables :

```
##          bitter sweet  acid salted alcaline
## bitter     1.00 -0.83  0.78  -0.67    -0.96
## sweet     -0.83  1.00 -0.61   0.49     0.93
## acid       0.78 -0.61  1.00  -0.44    -0.82
## salted    -0.67  0.49 -0.44   1.00     0.56
## alcaline  -0.96  0.93 -0.82   0.56     1.00
```

It is also possible to use multivariate descriptive statistics like PCA in order to :

- visualize on graphics distances between observations or correlations between variables.





- Built new numerical variables "summarizing" as well as possible the original variables in order to reduce dimension.

TABLE: Original data

	bitter	sweet	acid	salted	alcaline
St Yorre	3.4	3.1	2.9	6.4	4.8
Badoit	3.8	2.6	2.7	4.7	4.5
Vichy	2.9	2.9	2.1	6.0	5.0
Quézac	3.9	2.6	3.8	4.7	4.3
Arvie	3.1	3.2	3.0	5.2	5.0
Chateauneuf	3.7	2.8	3.0	5.2	4.6
Salvetat	4.0	2.8	3.0	4.1	4.5
Perrier	4.4	2.2	4.0	4.9	3.9

TABLE: Two new synthetic variables

	PC1	PC2
St Yorre	1.85	1.19
Badoit	-0.49	-0.64
Vichy	2.77	0.24
Quézac	-1.72	0.11
Arvie	1.93	-0.48
Chateauneuf	0.09	0.00
Salvetat	-0.93	-1.39
Perrier	-3.49	0.97

Outline

Basic concepts

We consider a numerical datatable where n observations are described on p variables.

	$1 \ldots$	j	$\ldots p$
1			
$\vdots$		$\vdots$	
i	$\ldots$	x_{ij}	$\ldots$
$\vdots$		$\vdots$	
n			

Some notations :

$\mathbf{X} = (x_{ij})_{n \times p}$ is the numerical data matrix where $x_{ij} \in \mathbb{R}$ is the value of the i^{th} observation on the j^{th} variable.

$$\mathbf{x}_i = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix} \in \mathbb{R}^p$$

the description of the i^{th} observation (row of $\mathbf{X}$)

$$\mathbf{x}^j = \begin{pmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{pmatrix} \in \mathbb{R}^n$$

the description of the j^{th} variable (columun of $\mathbf{X}$).

Example : 6 patients described on 3 variables (diastolic pressure, systolic pressure and cholesterol).

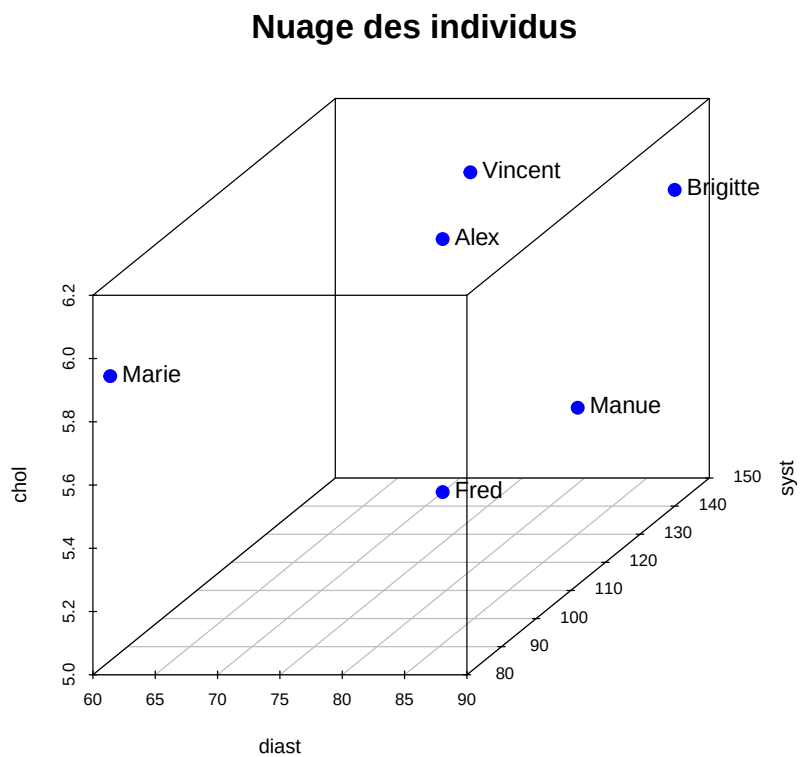
```
load("../data/chol.rda")
print(X)

##          diast syst chol
## Brigitte    90  140  6.0
## Marie       60   85  5.9
## Vincent     75  135  6.1
## Alex        70  145  5.8
## Manue       85  130  5.4
## Fred        70  145  5.0
```

$n =$ $p =$ $\mathbf{X} =$ $\mathbf{x}_3 =$ $\mathbf{x}^2 =$

$\Rightarrow$ Two sets of points.

The first set is the set of observations.

Example : the 6 patients define a set of $n = 6$ points in $\mathbb{R}^3$.



:

- Each observation i is a point $\mathbf{x}_i$ in $\mathbb{R}^p$ (a row of $\mathbf{X}$),
- A weight w_i is associated to each observation i. Usually :
  - $w_i = \frac{1}{n}$ for randomly drawn observations.
  - $w_i \neq \frac{1}{n}$ for ajusted samples, aggregated data...

A step of preproccessing is often applied to the data that might be :

- centered to have columns (variables) with mean zero,
- scaled to have columns (variables) of variance 1.

Original data matrix **X**

	1 ...	j	... p
1			
i	...	x_{ij}	...
n			
$\bar{\mathbf{x}}$	...	$\bar{x}^j$	...

Centered data matrix **Y**

	1 ...	j	... p
1			
i	...	y_{ij}	...
n			
$\bar{\mathbf{y}}$	...	0	...

Here :

- $\bar{x}^j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ est is the mean of the jth variable (column j of **X**),
- $y_{ij} = x_{ij} - \bar{x}^j$ is the general term of the centered data matrix **Y**.

The columns of the centered data matrix **Y** have zero mean :

$$\bar{y}^j = \frac{1}{n}\sum_{i=1}^{n} y_{ij} = 0.$$

Example : the set of 6 patients.

Original data matrix **X**

```
##          diast syst chol
## Brigitte    90  140  6.0
## Marie       60   85  5.9
## Vincent     75  135  6.1
## Alex        70  145  5.8
## Manue       85  130  5.4
## Fred        70  145  5.0
```

Means of the columns of **X**

```
## diast  syst  chol
##  75.0 130.0   5.7
```

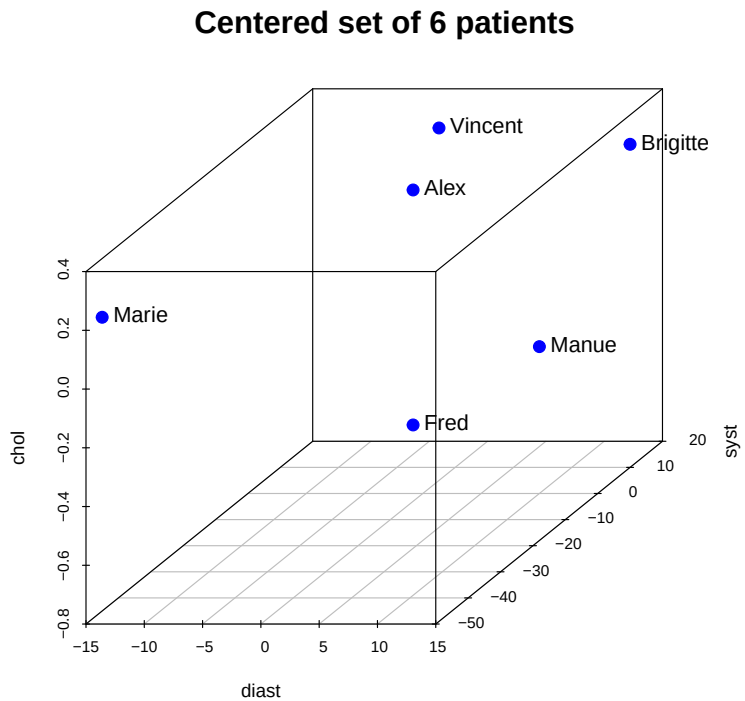
Centered data matrix **Y**

```
##          diast syst chol
## Brigitte    15   10  0.3
## Marie      -15  -45  0.2
## Vincent      0    5  0.4
## Alex        -5   15  0.1
## Manue       10    0 -0.3
## Fred        -5   15 -0.7
```

Means of the columns of **Y**

```
## diast  syst  chol
##     0     0     0
```

- Centering the data interprets as a translation of the set of observations in $\mathbb{R}^p$.



Original data matrix $\mathbf{X}$

	$1 \ldots$	j	$\ldots p$
1			
$\vdots$		$\vdots$	
i	$\ldots$	x_{ij}	$\ldots$
$\vdots$		$\vdots$	
n			
$\bar{\mathbf{x}}$	$\ldots$	$\bar{x}^j$	$\ldots$
s	$\ldots$	s_j	$\ldots$

Standardized data matrix $\mathbf{Z}$

	$1 \ldots$	j	$\ldots p$
1			
$\vdots$		$\vdots$	
i	$\ldots$	z_{ij}	$\ldots$
$\vdots$		$\vdots$	
n			
$\bar{\mathbf{z}}$	$\ldots$	0	$\ldots$
s	$\ldots$	1	$\ldots$

Here :

- $s_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}^j)^2$ is the variance of the jth variable (column j of $\mathbf{X}$),
- $z_{ij} = \frac{x_{ij} - \bar{x}^j}{s_j}$ is the general term of the standardized data matrix $\mathbf{Z}$.

The columns of the standardized data matrix $\mathbf{Z}$ have a mean equal to 0 and a variance equal to 1 :

$$\bar{z}^j = \frac{1}{n}\sum_{i=1}^{n} z_{ij} = 0, \ var(\mathbf{z}^j) = \frac{1}{n}\sum_{i=1}^{n}(z_{ij} - \bar{z}^j)^2 = 1.$$

Example : the set of 6 patients.

Original data matrix **X**

```
##          diast syst chol
## Brigitte    90  140  6.0
## Marie       60   85  5.9
## Vincent     75  135  6.1
## Alex        70  145  5.8
## Manue       85  130  5.4
## Fred        70  145  5.0
```

Means and sd of the columns of **X**

```
##      diast  syst  chol
## mean    75 130.0 5.700
## sd      10  20.8 0.383
```

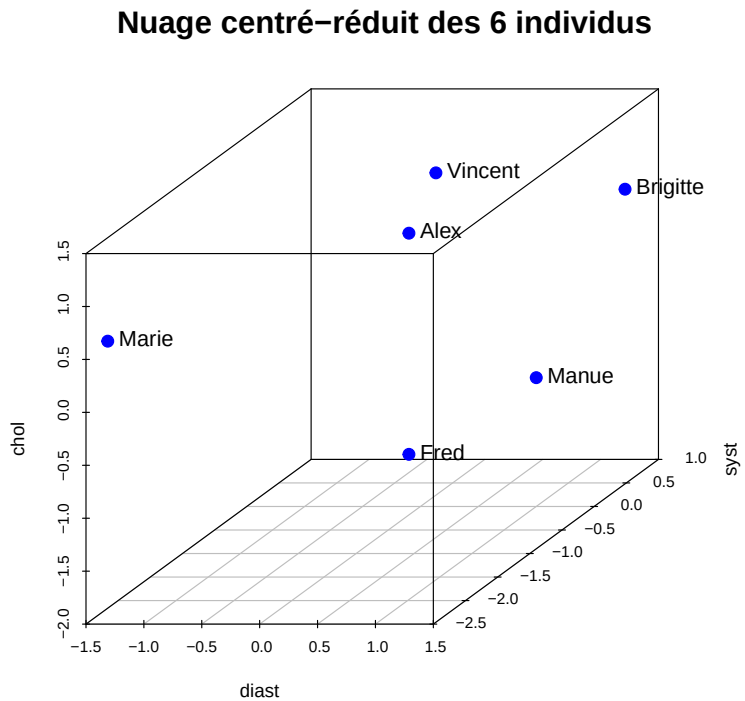
Standardized data matrix **Z**

```
##          diast  syst  chol
## Brigitte   1.5  0.48  0.78
## Marie     -1.5 -2.16  0.52
## Vincent    0.0  0.24  1.04
## Alex      -0.5  0.72  0.26
## Manue      1.0  0.00 -0.78
## Fred      -0.5  0.72 -1.83
```

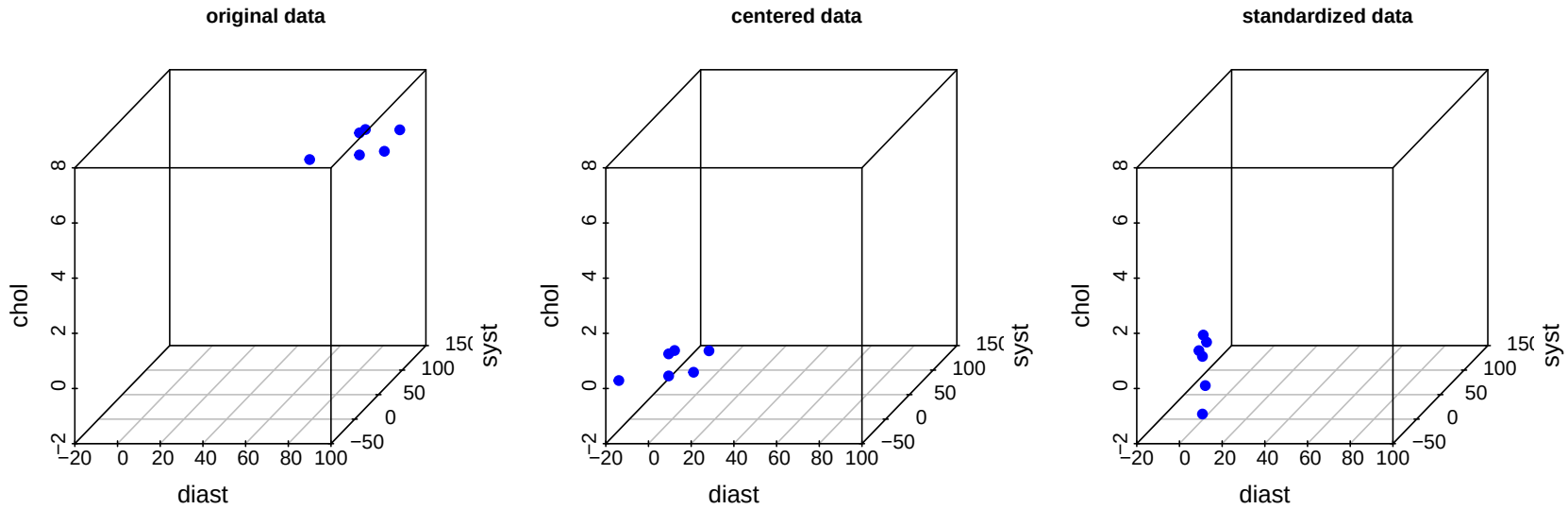
Means and sd of the columns of **Z**

```
## diast  syst  chol
##     0     0     0
## diast  syst  chol
##     1     1     1
```

- Standardization (centering and scaling) interprets as a translation and a normalisation of the set of observations in $\mathbb{R}^p$.



In summary, three datasets of the same observations.



- Centering do not change the distances between the observations :

$$d^2(\mathbf{x}_i, \mathbf{x}_{i'}) = d^2(\mathbf{y}_i, \mathbf{y}_{i'}).$$

- Standardization changes the distances between the observations :

$$d^2(\mathbf{x}_i, \mathbf{x}_{i'}) \neq d^2(\mathbf{z}_i, \mathbf{z}_{i'}).$$

Proximity between two observations can be measured with the Euclidean distance.

- The Euclidean distance between two observations i and i' (two rows of $\mathbf{X}$) is :

$$d^2(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2.$$

- When data are standardized, the Euclidean distance between two observations i and i' (two rows of $\mathbf{Z}$) is :

$$d^2(\mathbf{z}_i, \mathbf{z}_{i'}) = \sum_{j=1}^{p} \frac{1}{s_j^2} (x_{ij} - x_{i'j})^2.$$

It means :

- If variables (columns of $\mathbf{X}$) are measured on different scales, variables with larger variance are more important than variables with smaller variance when performing the Euclidean distance.
- Standardizing the data gives the same importance to all the variables when performing the Euclidean distance.

Example : distance between Brigitte and Marie

Original data ($\mathbf{X}$) :

```
##          diast syst chol
## Brigitte    90  140  6.0
## Marie       60   85  5.9
## Vincent     75  135  6.1
## Alex        70  145  5.8
## Manue       85  130  5.4
## Fred        70  145  5.0
```

Standardized data ($\mathbf{Z}$)

```
##          diast  syst  chol
## Brigitte   1.5  0.48  0.78
## Marie     -1.5 -2.16  0.52
## Vincent    0.0  0.24  1.04
## Alex      -0.5  0.72  0.26
## Manue      1.0  0.00 -0.78
## Fred      -0.5  0.72 -1.83
```

Mean and sd of the columns :

```
##      diast  syst  chol
## mean    75 130.0 5.700
## sd      10  20.8 0.383
```

Euclidean distance between the two first rows of $\mathbf{X}$:

$$\begin{aligned} d(\mathbf{x}_1, \mathbf{x}_2) &= \sqrt{(90-60)^2 + (140-85)^2 + (6-5.9)^2} \\ &= \sqrt{30^2 + 55^2 + 0.1^2} \end{aligned}$$

Euclidean distance between the two first rows of $\mathbf{Z}$:

$$\begin{aligned} d(\mathbf{z}_1, \mathbf{z}_2) &= \sqrt{\frac{1}{10^2}(90-60)^2 + \frac{1}{20.8^2}(140-85)^2 + \frac{1}{0.383^2}(6-5.9)^2} \\ &= \sqrt{(1.5+1.5)^2 + (0.48+2.16)^2 + (0.78-0.52)^2} \\ &= \sqrt{3^2 + 2.7^2 + 0.26^2} \end{aligned}$$

The dispersion of the set of observations in $\mathbb{R}^p$ is measured by the inertia.

- The inertia of the n observations (the n rows of $\mathbf{X}$) is defined by :

$$I(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^{n} d^2(\mathbf{x}_i, \bar{\mathbf{x}}).$$

- Inertia is a generalization of the variance to the case of multivariate data (p variables).
- One can show that :

$$I(\mathbf{X}) = \sum_{j=1}^{p} var(\mathbf{x}^j).$$

This means that :

- when the variables are centered, $I(\mathbf{Y}) = \sum_{i=1}^{p} s_j^2$,
- when the variabes are standardized, $I(\mathbf{Z}) = p$.

Example : Inertia of the set of 6 patients

Centered data ($\mathbf{Y}$) :

```
##          diast syst chol
## Brigitte    15   10  0.3
## Marie      -15  -45  0.2
## Vincent      0    5  0.4
## Alex        -5   15  0.1
## Manue       10    0 -0.3
## Fred        -5   15 -0.7
```

Variance of the columns :

```
##  diast   syst   chol
## 100.00 433.33   0.15
```

Standardized data ($\mathbf{Z}$)

```
##          diast  syst  chol
## Brigitte   1.5  0.48  0.78
## Marie     -1.5 -2.16  0.52
## Vincent    0.0  0.24  1.04
## Alex      -0.5  0.72  0.26
## Manue      1.0  0.00 -0.78
## Fred      -0.5  0.72 -1.83
```

Variance of the columns :

```
## diast  syst  chol
##     1     1     1
```

- Inertia of the centered dataset :

$$I(\mathbf{Y}) = 100 + 433.33 + 0.15$$

- Inertia if the standardized dataset :

$$I(\mathbf{Z}) = 1 + 1 + 1 = 3$$

The second set of points associated with a numerical data matrix is the set of variables.

Example : the variables diastolic pressure, systolic pressure and cholesterol define a set of $p = 3$ points in $\mathbb{R}^6$.

```
##       Brigitte Marie Vincent  Alex Manue Fred
## diast       90  60.0    75.0  70.0  85.0   70
## syst       140  85.0   135.0 145.0 130.0  145
## chol         6   5.9     6.1   5.8   5.4    5
```

Can't be visualized !

- Each variable j is a point $\mathbf{x}^j$ in $\mathbb{R}^n$ (a column $\mathbf{X}$),
- A weight m_j is associated with each variabe j. Usually :
  - $m_j = 1$ in PCA,
  - $m_j \neq 1$ in MCA (Multiple Correspondance Analysis).

When data are centered :

- each variable j is a point denoted $\mathbf{y}^j$ in $\mathbb{R}^n$ (a column $\mathbf{Y}$),
- we talk about the set of centered variables.

When data are standardized :

- each variable j is a point denoted $\mathbf{z}^j$ in $\mathbb{R}^n$ (column of $\mathbf{Z}$),
- we talk about the set of standardized variables.

The link between two variables is measured by the covariance or the correlation.

To define covariance and correlation, a metric is associated with $\mathbb{R}^n$:

$$\mathbf{N} = diag(\frac{1}{n}, \ldots, \frac{1}{n}).$$

- The scalar product between $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ is defined by :

$$< \mathbf{x}, \mathbf{y} >_{\mathbf{N}} = \mathbf{x}^T \mathbf{N} \mathbf{y} = \frac{1}{n} \mathbf{x}^T \mathbf{y} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i.$$

- The norm of $\mathbf{x}$ in $\mathbb{R}^n$ is then :

$$\|\mathbf{x}\|_{\mathbf{N}} = \sqrt{< \mathbf{x}, \mathbf{x} >_{\mathbf{N}}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2}.$$

With this metric, the variance writes as a squared norm :

- $var(\mathbf{x}^j) = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}^j)^2 = \|\mathbf{y}^j\|^2_{\mathbf{N}}$,
- $var(\mathbf{z}^j) = \frac{1}{n}\sum_{i=1}^{n}(z_{ij} - \bar{z}^j)^2 = \|\mathbf{z}^j\|^2_{\mathbf{N}}$.

The set of the p standardized variables is then on the unit ball of $\mathbb{R}^n$ with $\|\mathbf{z}^j\|_{\mathbf{N}} = 1$.

Moreover the covariance and the correlation write as scalar product :

- $c_{jj'} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \bar{x}^j)(x_{ij'} - \bar{x}^{j'}) =< \mathbf{y}^j, \mathbf{y}^{j'} >_{\mathbf{N}}$,
- $r_{jj'} = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_{ij} - \bar{x}^j}{s_j})(\frac{x_{ij'} - \bar{x}^{j'}}{s_{j'}}) =< \mathbf{z}^j, \mathbf{z}^{j'} >_{\mathbf{N}}$

This leads to a simple expression of the covariance matrix denoted **C** and of the correlation matrix denoted **R** :

- $\mathbf{C} = \mathbf{Y}^T\mathbf{N}\mathbf{Y}$,
- $\mathbf{R} = \mathbf{Z}^T\mathbf{N}\mathbf{Z}$.

Example :

Covariance matrix :

```
##        diast  syst  chol
## diast 100.00 112.5  0.25
## syst  112.50 433.3 -2.17
## chol    0.25  -2.2  0.15
```

Correlation matrix

```
##       diast  syst   chol
## diast 1.000  0.54  0.065
## syst  0.540  1.00 -0.272
## chol  0.065 -0.27  1.000
```

With this metric, the correlation writes as a cosine :

- $r_{jj'} = \frac{<\mathbf{y}^j, \mathbf{y}^{j'}>_\mathbf{N}}{\|\mathbf{y}^j\|_\mathbf{N} \|\mathbf{y}^{j'}\|_\mathbf{N}} = \cos\theta_\mathbf{N}(\mathbf{y}^j, \mathbf{y}^{j'})$,
- $r_{jj'} =< \mathbf{z}^j, \mathbf{z}^{j'} >_\mathbf{N} = \cos\theta_\mathbf{N}(\mathbf{z}^j, \mathbf{z}^{j'})$.

This lead to a geometrical interpretation of the correlation between variables :

- an angle of 90 degrees between two standardized variables corresponds to a null correlation (cosine equals to 0) and then to the absence of linear link,
- an angle of 0 degrees corresponds to a correlation of 1 (cosine equals to 1) and then to a positive linear link,
- an angle of 180 degrees corresponds to a correlation of -1 (cosinus equals to -1) and then to a negative linear link.

PCA analyses :

- either the centered data matrix $\mathbf{Y}$,
- or the standardized data matrix $\mathbf{Z}$.

This lead to two different methods of PCA :

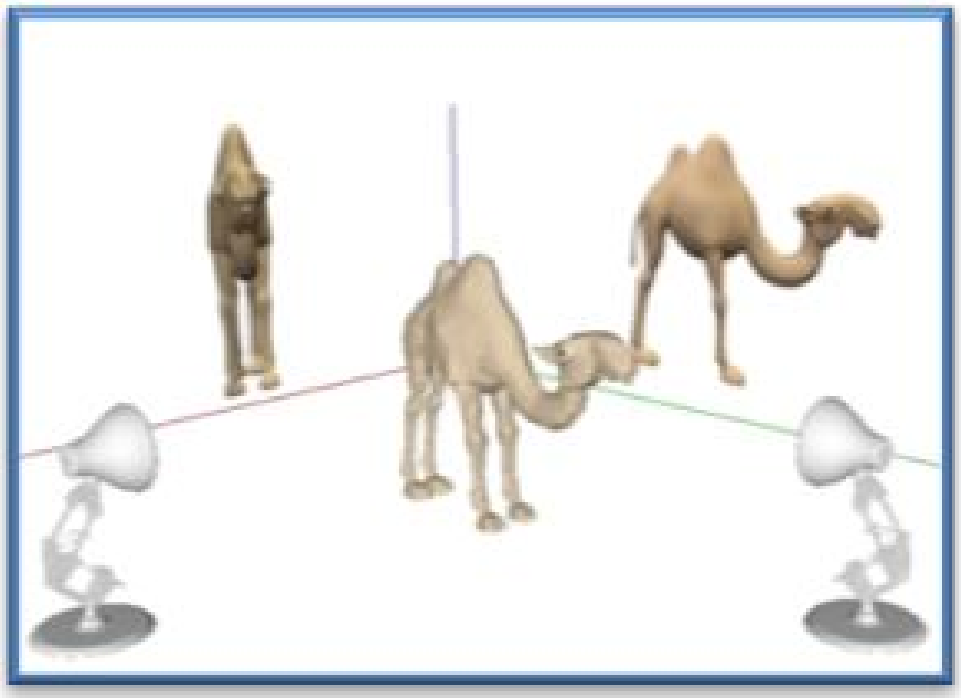
- non normalized PCA (or PCA on covariance matrix) which analyses $\mathbf{Y}$,
- normalized PCA (or PCA or correlation matrix) which analyses $\mathbf{Z}$.

From now on, normalized PCA is considered.

Outline

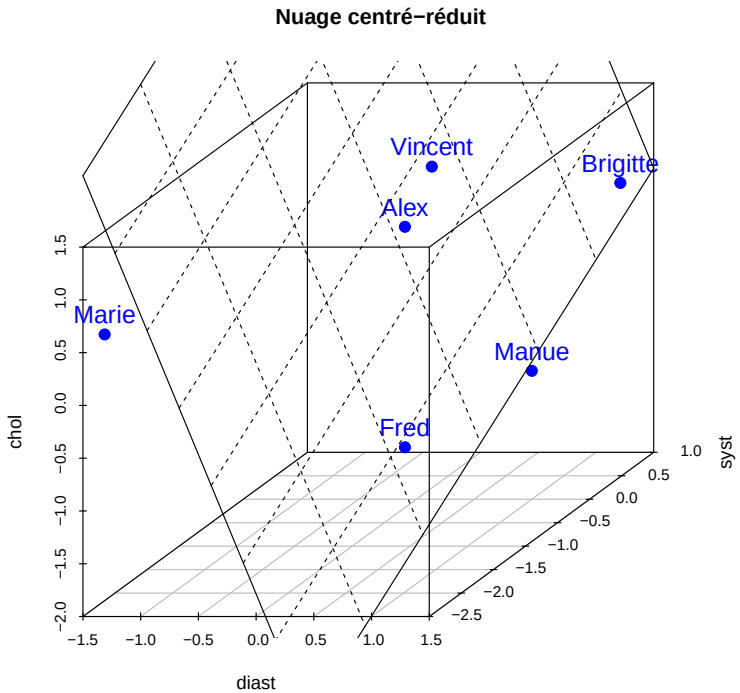
Analysis of the set of observations

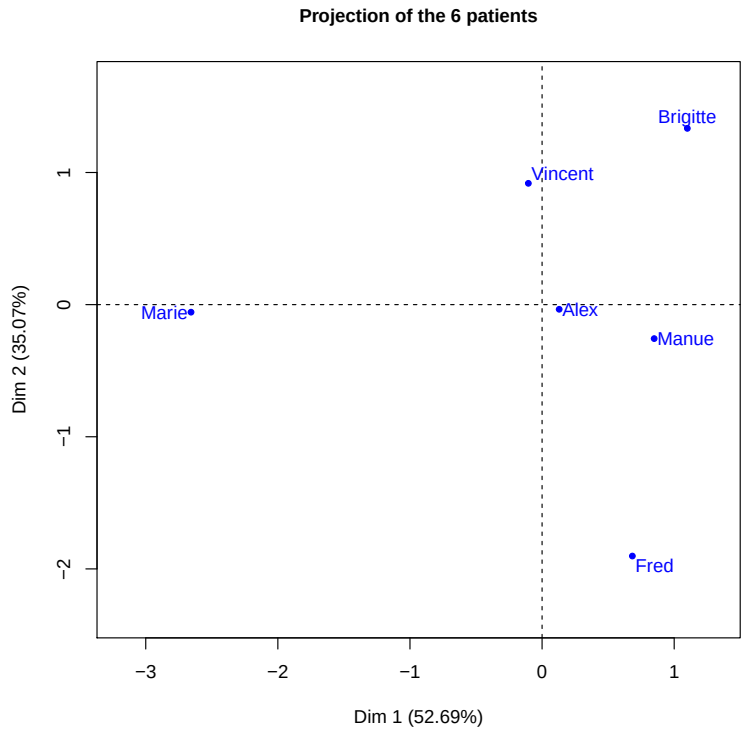
Find the subspace which gives the best representation of the observations.



- Best approximation of the data by projection.
- Best representation of the variability of the observations.

Example : the set of the 6 patients descibed on the 3 standardized variables.





The aim is to find the projection plane which keeps as good as possible the distances between the patients i.e. their variability and then their inertia.

Projection of an observation (a point in $\mathbb{R}^p$) on an axis.

The coordinate of the orhogonal projection of a point $\mathbf{z}_i \in \mathbb{R}^p$ on an axis Δ_α with orientation vector $\mathbf{v}_\alpha$ ($\mathbf{v}_\alpha^T \mathbf{v}_\alpha = 1$) is :

$$f_{i\alpha} =< \mathbf{z}_i, \mathbf{v}_\alpha >= \mathbf{z}_i^T \mathbf{v}_\alpha,$$

The vector of coordinates of the projections of the n observations is :

$$\mathbf{f}^\alpha = \begin{pmatrix} f_{1\alpha} \\ \vdots \\ f_{n\alpha} \end{pmatrix} = \mathbf{Z}\mathbf{v}_\alpha = \sum_{j=1}^{p} v_{j\alpha}\mathbf{z}^j.$$

- $\mathbf{f}^\alpha$ is a linear combination of the columns of $\mathbf{Z}$.
- $\mathbf{f}^\alpha$ is centered if the columns of $\mathbf{Z}$ are centered.

Example : the 6 patients are the rows of the following standardized data matrix

$$\mathbf{Z} = \begin{pmatrix} 1.50 & 0.48 & 0.78 \\ -1.50 & -2.16 & 0.52 \\ 0.00 & 0.24 & 1.04 \\ -0.50 & 0.72 & 0.26 \\ 1.00 & 0.00 & -0.78 \\ -0.50 & 0.72 & -1.83 \end{pmatrix}$$

Let us project the 6 "standardized" patients on two orthogonal axes Δ_1 and Δ_2 with orientation vectors :

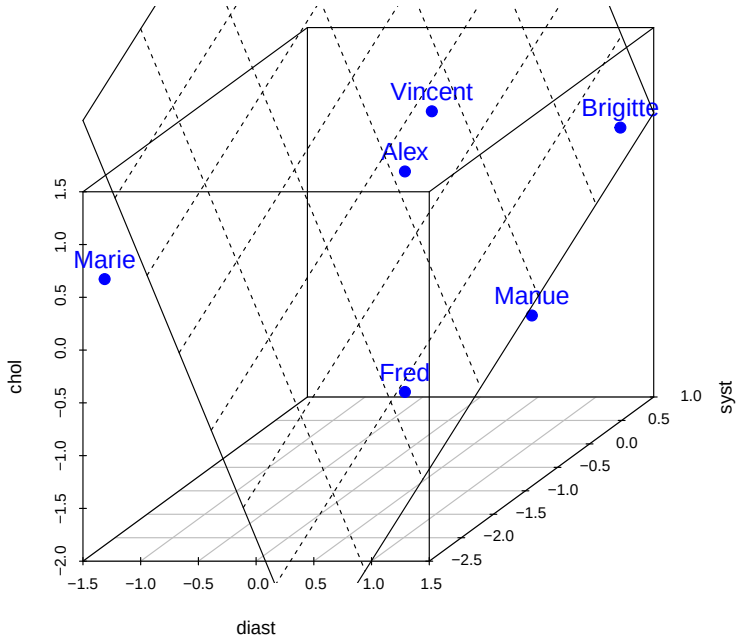
$$\mathbf{v}_1 = \begin{pmatrix} 0.641 \\ 0.72 \\ -0.265 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 0.4433 \\ -0.0652 \\ 0.894 \end{pmatrix}.$$

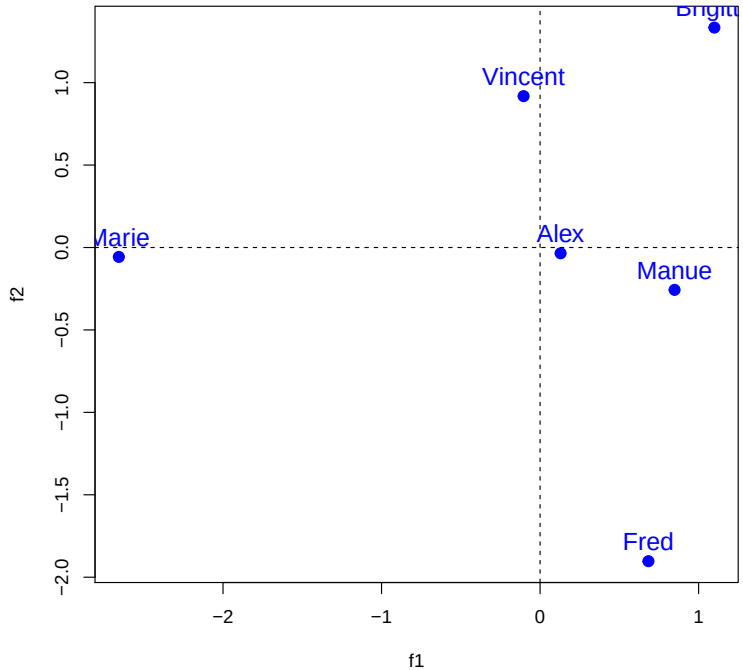
The vectors $\mathbf{f}^1$ and $\mathbf{f}^2$ of the coordinates of the projection of the 6 patients on Δ_1 and Δ_2 are :

$$\mathbf{f}^1 = \mathbf{Z}\mathbf{v}_1 = 0.641 \begin{pmatrix} 1.5 \\ \vdots \\ -0.5 \end{pmatrix} + 0.72 \begin{pmatrix} 0.48 \\ \vdots \\ 0.72 \end{pmatrix} - 0.265 \begin{pmatrix} 0.78 \\ \vdots \\ -1.82 \end{pmatrix} = \begin{pmatrix} 1.09 \\ \vdots \\ 0.683 \end{pmatrix}$$

$$\mathbf{f}^2 = \mathbf{Z}\mathbf{v}_2 = 0.4433 \begin{pmatrix} 1.5 \\ \vdots \\ -0.5 \end{pmatrix} - 0.0652 \begin{pmatrix} 0.48 \\ \vdots \\ 0.72 \end{pmatrix} - 0.894 \begin{pmatrix} 0.78 \\ \vdots \\ -1.82 \end{pmatrix} = \begin{pmatrix} 1.333 \\ \vdots \\ -1.9 \end{pmatrix}$$

$\mathbf{f}^1$ and $\mathbf{f}^2$ are two new synthetic and centered variables.





In PCA the orientation vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are defined to maximize the inertia of the set of projections of the observations and then keep as goog as possible the distances between the observations.

Axes of projection of the observations in PCA.

Δ_1 is the axis with orientation vector $\mathbf{v}_1 \in \mathbb{R}^p$ which maximises the variance of the n projected observations :

$$\begin{aligned}\mathbf{v}_1 &= \arg\max_{\|\mathbf{v}\|=1} var(\mathbf{Z}\mathbf{v}) \\ &= \arg\max_{\|\mathbf{v}\|=1} \mathbf{v}^T \mathbf{R}\mathbf{v}\end{aligned}$$

where

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

is the $p \times p$ correlation matrix.

One can show that :

- $\mathbf{v}_1$ is the eigenvector associated the largest eigenvalue λ_1 of $\mathbf{R}$,
- The first principal component (PC) $\mathbf{f}^1 = Z\mathbf{v}_1$ is centered :

$$\bar{\mathbf{f}^1} = 0,$$

- λ_1 is the variance of the first PC :

$$var(\mathbf{f}^1) = \lambda_1.$$

Δ_2 is the axis of orientation vector $\mathbf{v}_2 \perp \mathbf{v}_1$ which maximised the variance of the n projected observations :

$$\mathbf{v}_2 = \arg \max_{\|\mathbf{v}\|=1, \mathbf{v} \perp \mathbf{v}_1} var(\mathbf{Z}\mathbf{v}).$$

One can show that :

- $\mathbf{v}_2$ is the eigenvector associated with the second largest eigenvalue λ_2 of $\mathbf{R}$,
- The second principal component (PC) $\mathbf{f}^2 = Z\mathbf{v}_2$ is centered :

$$\bar{\mathbf{f}^2} = 0,$$

- λ_2 is the variance of the second PC :

$$var(\mathbf{f}^2) = \lambda_2,$$

- The principal components $\mathbf{f}^1$ and $\mathbf{f}^2$ are not correlated.

In the same way, we can get $q \leq r$ (r is the rank of $\mathbf{Z}$) orthogonal axes $\Delta_1, \ldots, \Delta_q$ on which observations are projected.

In summary :

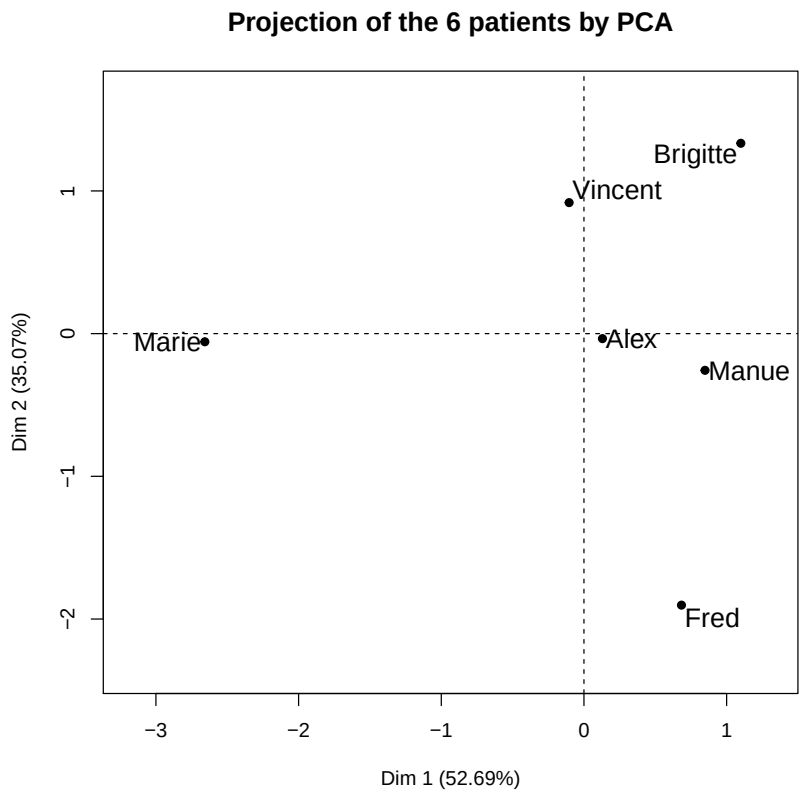
1. The eigen decomposition of the correlation matrix $\mathbf{R}$ is performed and $q \leq r$ is chosen.
2. The $n \times q$ matrix $\mathbf{F} = \mathbf{ZV}$ of the q principal components is obtained with the matrix $\mathbf{V}$ of the q first eigenvectors of $\mathbf{R}$.
   - The principal components $\mathbf{f}^\alpha = \mathbf{Z}\mathbf{v}_\alpha$ (column of $\mathbf{F}$) are centered and of variance λ_α.
   - The elements $f_{i\alpha}$ are called the factor coordinates or the observations or also the scores of the observations on the principal components.

$\mathbf{F}=$

	$1 \ldots$	α	$\ldots q$
1			
$\vdots$		$\vdots$	
i	$\ldots$	$f_{i\alpha}$	$\ldots$
$\vdots$		$\vdots$	
n			
mean	$\ldots$	0	$\ldots$
var	$\ldots$	λ_α	$\ldots$

Example of the 6 patients : matrix **F** of the $q = 2$ first PC

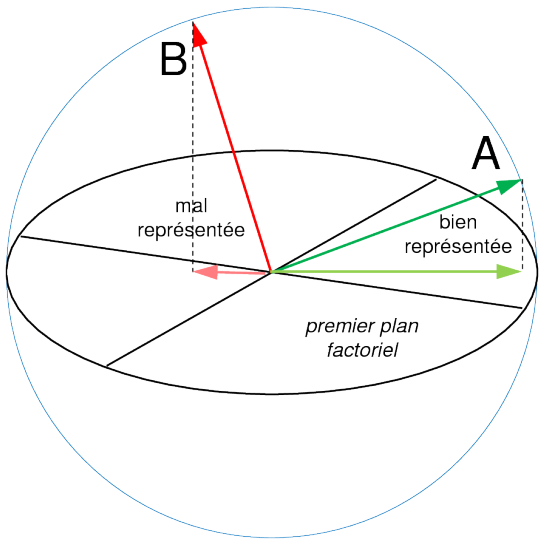
```
##             f1     f2
## Brigitte  1.10  1.334
## Marie    -2.66 -0.057
## Vincent  -0.10  0.918
## Alex      0.13 -0.035
## Manue     0.85 -0.257
## Fred      0.68 -1.903
```



Outline

Analysis of the set of variables

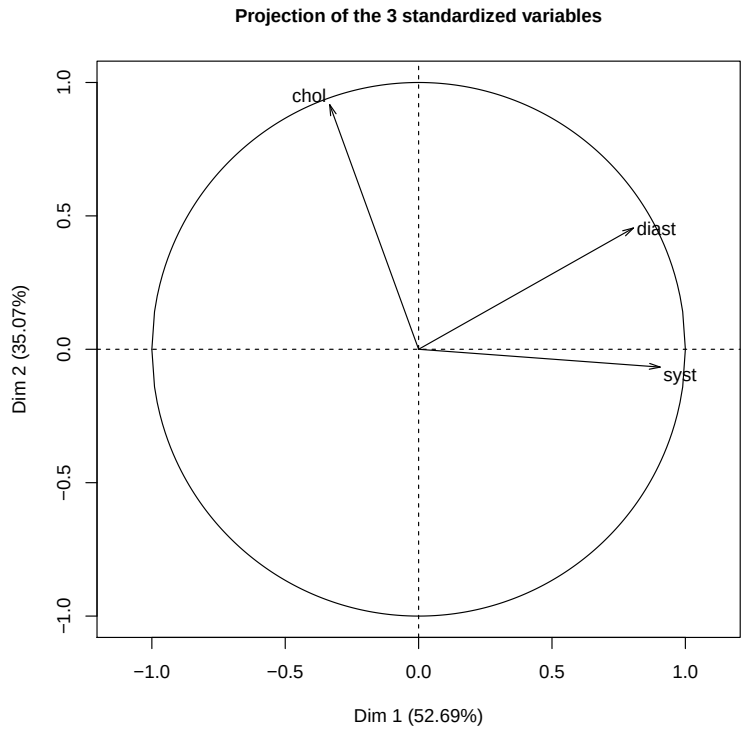
Find the subspace which gives the best representation of the variables.



Example : the set of 3 standardized variables.

3 variables on the unit ball of $\mathbb{R}^6$.

```
##       Brigitte Marie Vincent Alex Manue Fred
## diast      1.5  -1.5     0.0 -0.5   1.0 -0.5
## syst       0.5  -2.2     0.2  0.7   0.0  0.7
## chol       0.8   0.5     1.0  0.3  -0.8 -1.8
```



The aim is to find the projection plane which represents best the variables and then keeps as good as possible the angles between the variables i.e. their correlation.

Projection of a variable (a point in $\mathbb{R}^n$) on an axis.

The coordinate of the N-orthogonal projection of a point $\mathbf{z}^j \in \mathbb{R}^n$ on an axis G_α with orientation vector $\mathbf{u}_\alpha$ ($\mathbf{u}_\alpha^T \mathbf{N} \mathbf{u}_\alpha = 1$) is :

$$a_{j\alpha} = < \mathbf{z}^j, \mathbf{u}_\alpha >_{\mathbf{N}} = (\mathbf{z}^j)^T \mathbf{N} \mathbf{u}_\alpha,$$

and the vector of coordinates of the projections of the p variables is :

$$\mathbf{a}^\alpha = \begin{pmatrix} a_{1\alpha} \\ \vdots \\ a_{p\alpha} \end{pmatrix} = \mathbf{Z}^T \mathbf{N} \mathbf{u}_\alpha$$

Warning : a metric $\mathbf{N}$ in $\mathbb{R}^n$ is used.

- A metric in $\mathbb{R}^n$ is a $n \times n$ positive semidefinite matrix.
- Here in PCA, $\mathbf{N}$ is the diagonal matrix of the weight of the observations :

$$\mathbf{N} = diag(w_1, \ldots, w_n).$$

- When all observations are weighted by $\frac{1}{n}$ (usually by default) :

$$\mathbf{N} = \frac{1}{n} \mathbb{I}_n.$$

Example : the three variables (diast, syst, chol) are columns of the following standardized data matrix

$$\mathbf{Z} = \begin{pmatrix} 1.50 & 0.48 & 0.78 \\ -1.50 & -2.16 & 0.52 \\ 0.00 & 0.24 & 1.04 \\ -0.50 & 0.72 & 0.26 \\ 1.00 & 0.00 & -0.78 \\ -0.50 & 0.72 & -1.83 \end{pmatrix}$$

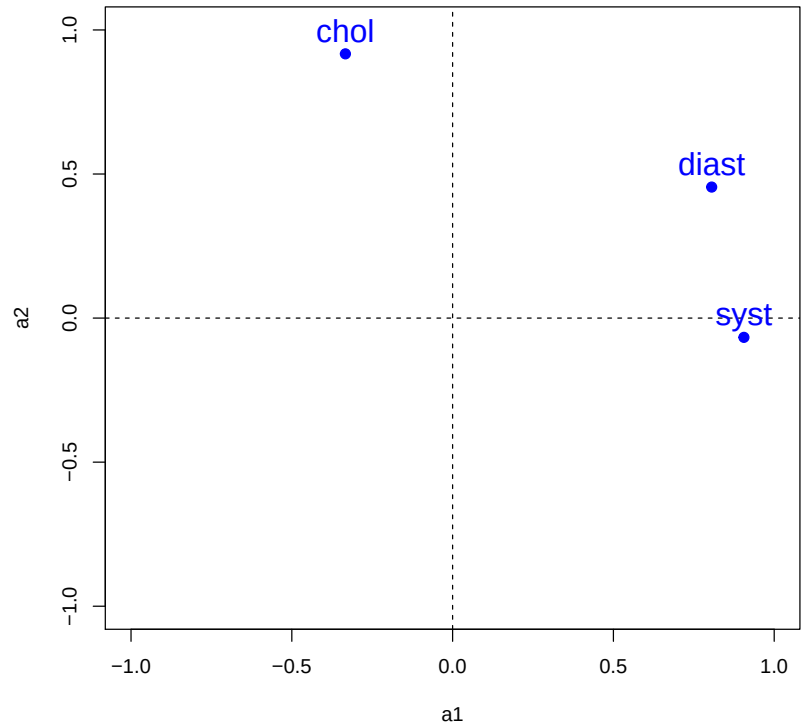
Let us project the 3 standardized variables on two N-orthogonal axes G_1 and G_2 with orientation vectors (here $\mathbf{N} = \frac{1}{6}\mathbb{I}_6$) :

$$\mathbf{u}_1 = \begin{pmatrix} 0.87 \\ -2.11 \\ -0.08 \\ 0.10 \\ 0.67 \\ 0.54 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} 1.30 \\ -0.06 \\ 0.90 \\ -0.03 \\ -0.25 \\ -1.8 \end{pmatrix}.$$

The vectors $\mathbf{a}^1$ and $\mathbf{a}^2$ of coordinates of the projection of the 3 variables on G_1 and G_2 are :

$$\mathbf{a}^1 = \mathbf{Z}^T\mathbf{N}\mathbf{u}_1 = \frac{0.87}{6}\begin{pmatrix}1.5\\0.48\\0.78\end{pmatrix}\frac{-2.11}{6}\begin{pmatrix}-1.5\\-2.16\\0.52\end{pmatrix} + \ldots \frac{+0.54}{6}\begin{pmatrix}-0.5\\0.72\\-1.83\end{pmatrix} = \begin{pmatrix}0.81\\0.91\\-0.33\end{pmatrix}$$

$$\mathbf{a}^2 = \mathbf{Z}^T\mathbf{N}\mathbf{u}_2 = \frac{1.30}{6}\begin{pmatrix}1.5\\0.48\\0.78\end{pmatrix}\frac{-0.06}{6}\begin{pmatrix}-1.5\\-2.16\\0.52\end{pmatrix} + \ldots \frac{-1.80}{6}\begin{pmatrix}-0.5\\0.72\\-1.83\end{pmatrix} = \begin{pmatrix}0.45\\-0.07\\0.92\end{pmatrix}$$



In PCA the orientation vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are defined to maximize the sum of the square cosine of angles between the variables and the projection axis.

Axes of projection of the variables in PCA.

G_1 is the axis with orientation vector $\mathbf{u}_1 \in \mathbb{R}^n$ which maximises the sum of the square cosine of angles with the variables.

$$\mathbf{u}_1 = \arg \max_{\|\mathbf{u}\|_\mathbf{N}=1} \sum_{j=1}^{p} \cos^2 \theta_\mathbf{N}(\mathbf{z}^j, \mathbf{u})$$

$$= \arg \max_{\|\mathbf{u}\|_\mathbf{N}=1} \|\mathbf{Z}^T \mathbf{N} \mathbf{u}\|^2$$

One can show that with $\mathbf{N} = \frac{1}{n}\mathbb{I}_n$:

- $\mathbf{u}_1$ is the eigenvector associated with the largest eigenvalue of the matrix

$$\frac{1}{n}\mathbf{Z}\mathbf{Z}^T,$$

- the largest eigenvalue of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$ is also the largest eigenvalue λ_1 of $\mathbf{R} = \frac{1}{n}\mathbf{Z}^T\mathbf{Z}$,
- λ_1 is the sum of the square cosines between the variables and $\mathbf{u}_1$:

$$\lambda_1 = \sum_{j=1}^{p} cos^2 \theta_\mathbf{N}(\mathbf{z}^j, \mathbf{u}_1)$$

G_2 is the axis with orientation $\mathbf{u}_2 \perp_{\mathbf{N}} \mathbf{u}_1$ which maximises the sum of the square cosine of angles with the variables :

$$\mathbf{u}_2 = \arg \max_{\|\mathbf{u}\|_{\mathbf{N}}=1, \mathbf{u}_2 \perp_{\mathbf{N}} \mathbf{u}_1} \sum_{j=1}^{p} \cos^2 \theta_{\mathbf{N}}(\mathbf{z}^j, \mathbf{u})$$

One can show that :

- $\mathbf{u}_2$ is the eigenvector associated with the second largest eigenvalue of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$.
- the second largest eigenvalue of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$ is also the second largest eigenvalue λ_2 of $\mathbf{R} = \frac{1}{n}\mathbf{Z}^T\mathbf{Z}$,
- λ_2 is the sum of the square cosines between the variables and $\mathbf{u}_2$:

$$\lambda_2 = \sum_{j=1}^{p} \cos^2 \theta_{\mathbf{N}}(\mathbf{z}^j, \mathbf{u}_2)$$

In the same way, we can get $q \leq r$ (r is the rank of $\mathbf{Z}$) orthogonal axes $G_1, \ldots, G_q$ on which variables are projected.

In summary :

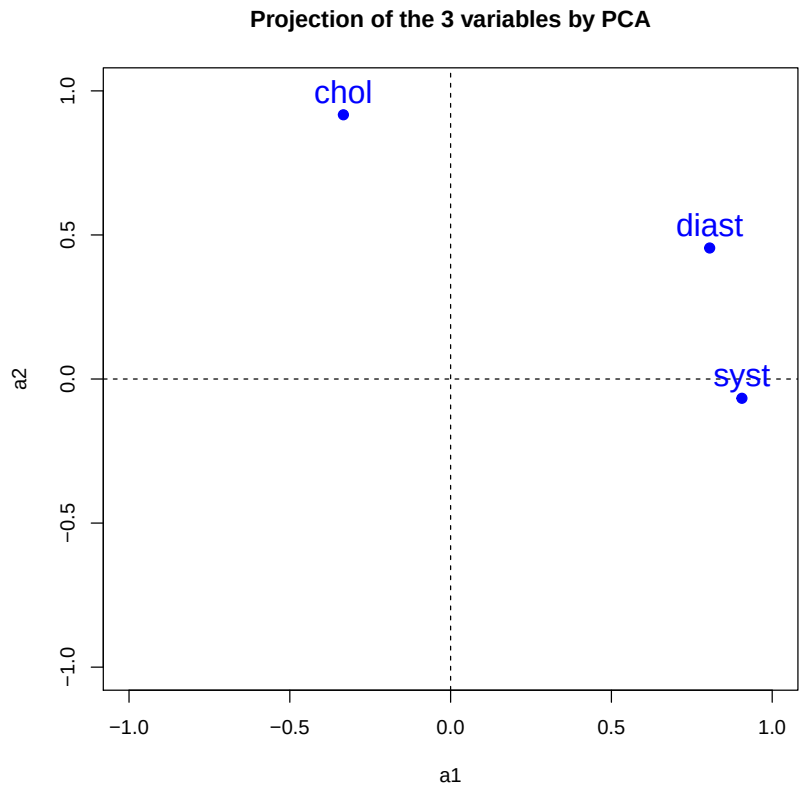
1. The eigen decomposition of the matrix $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$ is performed and $q \leq r$ is chosen.
2. The $p \times q$ matrix $\mathbf{A} = \mathbf{Z}^T\mathbf{N}\mathbf{U}$ of the q loading vectors is obtained with the matrix $\mathbf{U}$ of the q first eigenvectors of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$.
   - The loading vector $\mathbf{a}^\alpha = \mathbf{Z}^T\mathbf{N}\mathbf{u}_\alpha$ (column of $\mathbf{A}$) contains the coordinates of the projections of the p variables on the axis G_α.
   - The elements $a_{i\alpha}$ are called the factor coordinates of the variables or also the loadings of the variables.

$\mathbf{A}=$

	$1 \ldots$	α	$\ldots q$	
1				
$\vdots$		$\vdots$		
j	$\ldots$	$a_{j\alpha}$	$\ldots$	
$\vdots$		$\vdots$		
p				
norme	$\ldots$	$\sqrt{\lambda_\alpha}$	$\ldots$	

Example of the three variables : matrice **A** of the $q = 2$ first loading vectors.

```
##          a1     a2
## diast  0.81  0.455
## syst   0.91 -0.067
## chol  -0.33  0.917
```



Transition formulas.

One can show that

- principal components can be performed directly by eigen decomposition of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$:

$$\mathbf{f}^\alpha = \mathbf{Z}\mathbf{v}_\alpha = \sqrt{\lambda_\alpha}\,\mathbf{u}_\alpha,$$

- loadings can be performed directly by eigen decomposition $\frac{1}{n}\mathbf{Z}^T\mathbf{Z}$:

$$\mathbf{a}^\alpha = \mathbf{Z}^T\mathbf{N}\mathbf{u}_\alpha = \sqrt{\lambda_\alpha}\,\mathbf{v}_\alpha$$

Then :

$$\mathbf{F} = \mathbf{U}\Lambda$$
$$\mathbf{A} = \mathbf{V}\Lambda$$

where $\Lambda = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_q})$

This means that

- the eigenvectors $\mathbf{u}_\alpha$ of $\frac{1}{n}\mathbf{Z}\mathbf{Z}^T$ are the standardized principal components :

$$\mathbf{u}_\alpha = \frac{\mathbf{f}^\alpha}{\sqrt{\lambda_\alpha}},$$

- The loadings are correlations between the variables and the principal components :

$$a_{j\alpha} = cor(\mathbf{x}^j, \mathbf{f}^\alpha).$$

This last property is in crucial for PCA results interpretation.

Outline

Interpretation of PCA results

Variance of the principal components.

Principal components (columns of $\mathbf{F}$) are q new synthetic variables which are non correlated and of maximum variance with

$$var(\mathbf{f}^{\alpha}) = \lambda_{\alpha}$$

This means that the inertia of the set of observations projected on the q first dimensions of PCA is :

$$I(\mathbf{F}) = \lambda_1 + \ldots + \lambda_q.$$

Example : the set of 6 patients :

The $p = 3$ non null eigenvalues of the correlation matrix $\mathbf{R}$ are :

```
##         eigenvalue
## lambda1       1.58
## lambda2       1.05
## lambda3       0.37
```

then

$$\begin{aligned} var(\mathbf{f}^1) &= 1.58 \\ var(\mathbf{f}^2) &= 1.05 \end{aligned}$$

and the inertia of the 6 patients projected on the $q = 2$ first dimensions of PCA is :

$$\lambda_1 + \lambda_2 = 1.58 + 1.05 = 2.63.$$

Total inertia.

Total inertia in normalized PCA is the sum of the variance of the columns of $\mathbf{Z}$:

$$I(\mathbf{Z}) = \sum_{j=1}^{p} var(\mathbf{z}^j) = p.$$

When $q = r$ the total inertia is equal to the sum of the variance of all the principal components :

$$I(\mathbf{F}) = \lambda_1 + \ldots + \lambda_r = I(\mathbf{Z}) = p$$

Example :

Inertia of the 6 patients projected on the $q = 3$ (all) principal components :

$$I(\mathbf{F}) = \lambda_1 + \lambda_2 + \lambda_3 = 1.58 + 1.05 + 0.37 = 3$$

Quality of the dimension reduction.

- The proportion of the inertia of the data explained by the αth principal component is :

$$\frac{var(\mathbf{f}^\alpha)}{I(\mathbf{Z})} = \frac{\lambda_\alpha}{\lambda_1 + \ldots + \lambda_r}.$$

- The proportion of the inertia of the data explained by the q first principal components is :

$$\frac{I(\mathbf{F})}{I(\mathbf{Z})} = \frac{\lambda_1 + \ldots + \lambda_q}{\lambda_1 + \ldots + \lambda_r}.$$

Example : the set of 6 patients.

Original data ($p = 3$ and n=6)

```
##          diast syst chol
## Brigitte    90  140  6.0
## Marie       60   85  5.9
## Vincent     75  135  6.1
## Alex        70  145  5.8
## Manue       85  130  5.4
## Fred        70  145  5.0
```

Reduction to 2 PC

```
##             f1     f2
## Brigitte  1.10  1.334
## Marie    -2.66 -0.057
## Vincent  -0.10  0.918
## Alex      0.13 -0.035
## Manue     0.85 -0.257
## Fred      0.68 -1.903
```

What is the quality of this reduction ?

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1       1.58                     53                                 53
## comp 2       1.05                     35                                 88
## comp 3       0.37                     12                                100
```

- $r = 3$ non numm eigenvalues because $r = \min(n - 1, p) = 3$,
- The sum of the eigenvalues is $p = 3$ (the total inertia),
- 53 % of the inertia is explained by the first PC.
- 88 % of the inertia is by the two first PC.
- 100 % of the inertia is explained by all the PC.

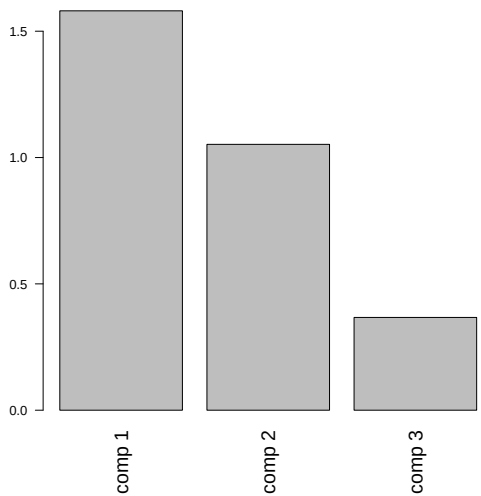
How many comonents to keep ?

- The number q of components can be chosen according to a fixed proportion of explained inertia.
- It can be chosen such that the components with variance λ_α greater than the mean variance (by variable) are kept. In normalized PCA, the mean variance is 1. Then q is chosen such that $\lambda_q > 1$ and $\lambda_{q+1} < 1$. This is the Kaiser rule.
- It can be chosen looking at the barplot of the eigenvalues and identifying a "break". This break can be also identified using the elbow rule :
  i. perform the first-differences : $\epsilon_1 = \lambda_1 - \lambda_2$, $\epsilon_2 = \lambda_2 - \lambda_2$, ...
  ii. perform the second-differences : $\delta_1 = \epsilon_1 - \epsilon_2$, $\delta_2 = \epsilon_2 - \epsilon_2$, ...
  iii. keep q such that $\delta_1, \ldots, \delta_{q-1}$ are positive and δ_q are negative.
- choose q according to a stability criterion estimated by bootstrap or cross-validation methods.

Example : the set of 6 patients.

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1       1.58                     53                                53
## comp 2       1.05                     35                                88
## comp 3       0.37                     12                               100
```

Ebouli des valeurs propres



- 88% of the inertia is explained by $q = 2$ componetss.
- Kaiser rule : two eigenvalues greater than 1.
- Elbow rule : "break" after 2 components.

We choose to keep $q = 2$ principal components to describe the 6 patients originally described by $p = 3$ variables.

Only 12% of the information (the inertia of the data) is lost.

Interpretation of the projection plans of the observations.

If two observations are well projected, then their distance on the projection plan is close to their distance in $\mathbb{R}^p$.

- The quality of the projection of an observation i on the axis Δ_α is measured by the square cosine of the angle $\theta_{i\alpha}$ between the point $\mathbf{z}_i$ and the axis Δ_α :

$$\cos^2(\theta_{i\alpha}) = \frac{f_{i\alpha}^2}{\|\mathbf{z}_i\|^2}$$

- The quality of the projection of an observation i on the plan $(\Delta_\alpha, \Delta_{\alpha'})$ is measured by the square cosine of the angle $\theta_{i(\alpha,\alpha')}$ between the point $\mathbf{z}_i$ and the plan $(\Delta_\alpha, \Delta_{\alpha'})$:

$$\cos^2(\theta_{i(\alpha,\alpha')}) = \frac{f_{i\alpha}^2 + f_{i\alpha'}^2}{\|\mathbf{z}_i\|^2}$$

The more $\cos^2$ is close to 1, the better the quality of the projection the observation i.

Example : the set of patients.

Factor coordinates of the patients

```
##          Dim.1  Dim.2
## Brigitte  1.10  1.334
## Marie    -2.66 -0.057
## Vincent  -0.10  0.918
## Alex      0.13 -0.035
## Manue     0.85 -0.257
## Fred      0.68 -1.903
```

cos^2 of the patients on the axes

```
##          Delta1  Delta2
## Brigitte 0.3907 0.57504
## Marie    0.9812 0.00045
## Vincent  0.0094 0.73386
## Alex     0.0200 0.00148
## Manue    0.4462 0.04094
## Fred     0.1137 0.88080
```

The quality of the projection of Brigitte on the first axis est 0.3907.

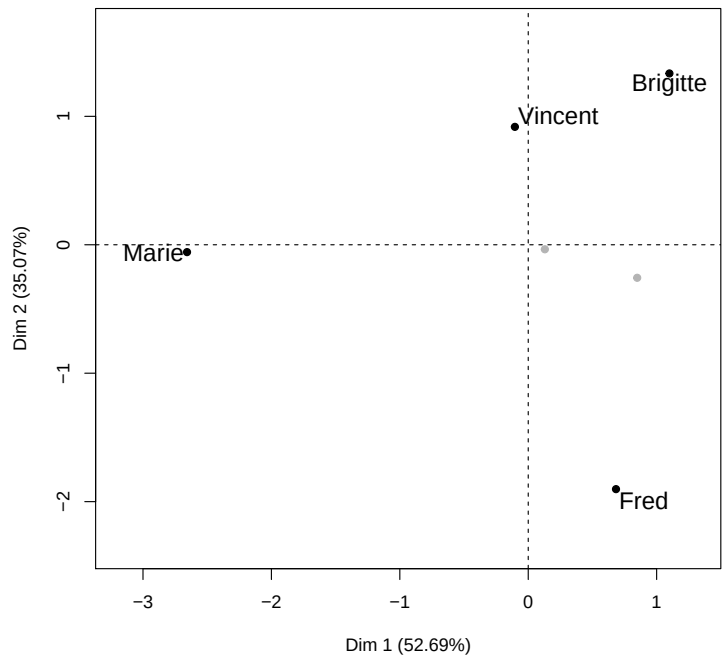
The *cos*2 of the patients on the axes can also calculated with the distances between the patients and the origin :

```
## Brigitte    Marie  Vincent     Alex    Manue     Fred
##     1.76     2.68     1.07     0.92     1.27     2.03
```

Brigitte is well projected on the first PCA plan as her cos^2 with that plan is $0.966 = 0.3907 + 0.57504$.

In the same way we get the cos^2 of the 6 patients on the first PCA plan :

```
##     Fred    Marie Brigitte  Vincent    Manue     Alex
##    0.994    0.982    0.966    0.743    0.487    0.022
```



- Are the observations globally well represented on this plan ?
- Interpret the distances between Vincent and Marie, between Vincent and Brigitte.

The observations having an important contribution to the inertia of the projected data is source of instabillity.

- The inertia (the variance) on the axis Δ_α is $\lambda_\alpha = \sum_{i=1}^{n} w_i f_{i\alpha}^2$ with usually $w_i = \frac{1}{n}$.
- The relative contribution of an observation i to the inertia on the axis Δ_α is

$$Ctr(i, \alpha) = \frac{w_i f_{i\alpha}^2}{\lambda_\alpha}.$$

- The relative contribution of an observation i to the inertia on the plan $(\Delta_\alpha, \Delta'_\alpha)$ is

$$Ctr(i, (\alpha, \alpha')) = \frac{w_i f_{i\alpha}^2 + w_i f_{i\alpha'}^2}{\lambda_\alpha + \lambda_{\alpha'}}.$$

When the weights w_i are all identical ($w_i = \frac{1}{n}$ for instance), the observations with a fringe location on the plan are those with the greater contribution.

Example : the set of 6 patients.

Factor coordinates of the patients

```
##             f1     f2
## Brigitte  1.10  1.334
## Marie    -2.66 -0.057
## Vincent  -0.10  0.918
## Alex      0.13 -0.035
## Manue     0.85 -0.257
## Fred      0.68 -1.903
```

Contributions of the patients

```
##          Delta1 Delta2
## Brigitte  12.75 28.186
## Marie     74.44  0.052
## Vincent    0.11 13.352
## Alex       0.18  0.020
## Manue      7.59  1.046
## Fred       4.93 57.345
```

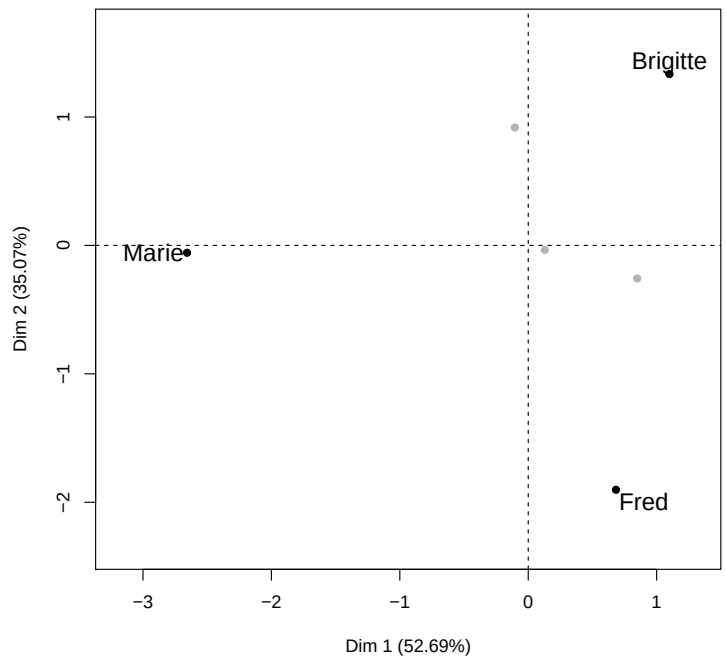
Two first eigenvalues

```
## lambda1 lambda2
##     1.6     1.1
```

- Perform the contribution of Brigitte to the inertia on the first PCA axis and then to the first plan.
- Check that for each axis, the sum of the relative contributions is equal to 1.

The contributions of the 6 patients to the inertia on the first PCA plan are :

```
##    Marie     Fred Brigitte  Vincent    Manue     Alex
##    44.71    25.87    18.92     5.40     4.98     0.11
```



The 3 patients with the highest contributions have fringe locations.

Correlation circle interpretation.

If two variables are well projected, then their angle in the projection plane is close to their angle in $\mathbb{R}^n$. The correlation between those two variables is then close to the cosine of their angle in the projection plane.

- The quality of the projection of a variable j on the axis G_α is measured by the square cosine of the angle $\theta_{j\alpha}$ between the point $\mathbf{z}^j$ and the axis G_α :

$$\cos^2(\theta_{j\alpha}) = \frac{a_{j\alpha}^2}{\|\mathbf{z}^j\|_{\mathbf{N}}^2} = a_{j\alpha}^2$$

- The quality of the projection of a variable j on the plane $(G_\alpha, G_{\alpha'})$ is measured by the square cosine of the angle $\theta_{j(\alpha,\alpha')}$ between the point $\mathbf{z}^j$ and the plan $(G_\alpha, G_{\alpha'})$:

$$\cos^2(\theta_{j(\alpha,\alpha')}) = a_{j\alpha}^2 + a_{j\alpha'}^2.$$

$\sqrt{\cos^2(\theta_{j(\alpha,\alpha')})}$ is then the "length of the arrow".

The closer to the unit circle, the better the quality of the projection of the variable.

Example : the set of 3 variables.

Factor coordinates of the variables

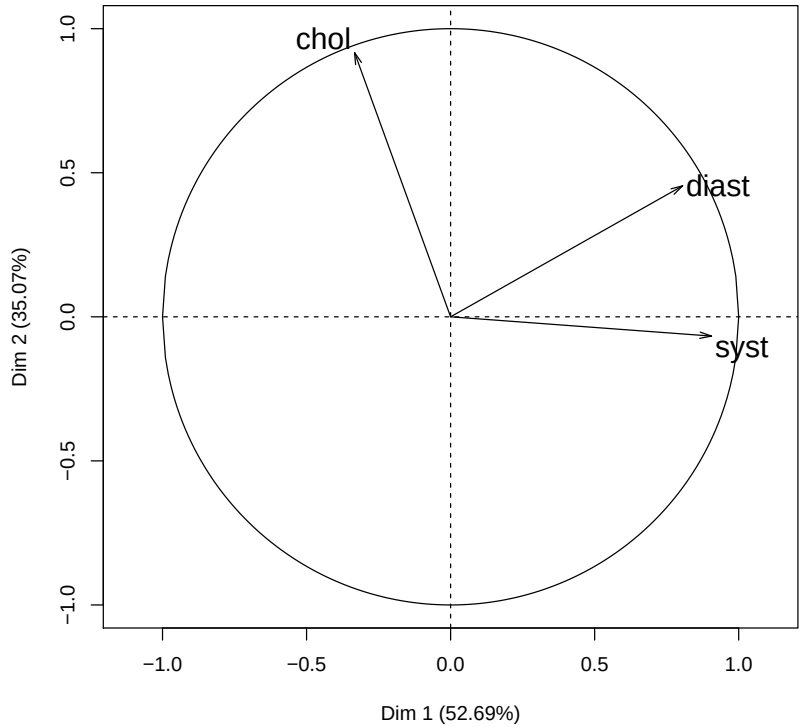
```
##          a1     a2
## diast  0.81  0.455
## syst   0.91 -0.067
## chol  -0.33  0.917
```

$\cos^2$ of the variables on the axes

```
##         G1     G2
## diast 0.65 0.2068
## syst  0.82 0.0045
## chol  0.11 0.8410
```

The $\cos^2$ of the variable diast on G_1 is 0.65 and the cos^2 of the variable diast on (G_1, G_2) is $0.65 + 0.2068 = 0.8568$.

The variable diast is then well projected on this plan and the length of the arrow in the correlation circle will be $\sqrt{0.8568} = 0.92563$.



- Are the variables globally well projected on this plan ?
- Interpret the correlation circle.

The variables having an important contribution to the inertia of the projected data are used to interpret the axes.

- The inertia of the axis Δ_α is $\lambda_\alpha = \sum_{j=1}^{p} a_{j\alpha}^2 = \sum_{j=1}^{p} cos^2\theta_{j\alpha}$.
- The relative contribution of a variable j to the inertia on the axis Δ_α is

$$Ctr(j,\alpha) = \frac{a_{j\alpha}^2}{\lambda_\alpha}.$$

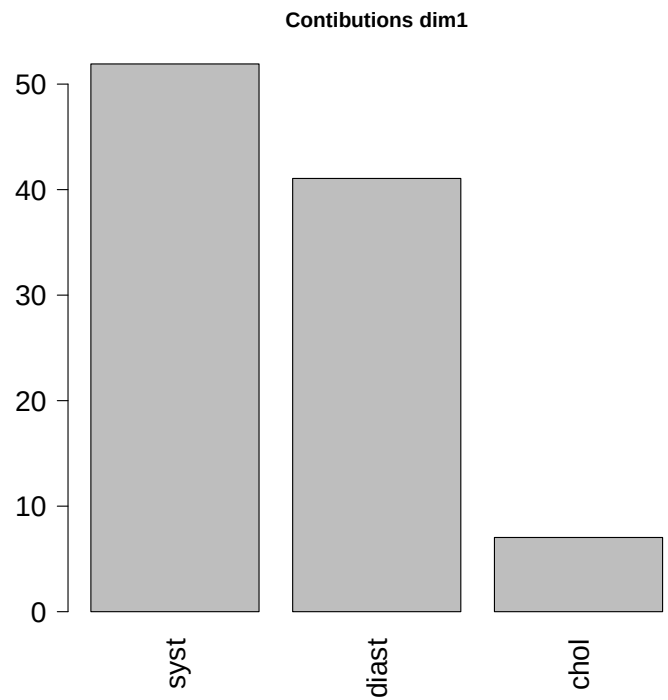
- The relative contribution of a variable j to the inertia on the plan $(\Delta_\alpha, \Delta'_\alpha)$ is

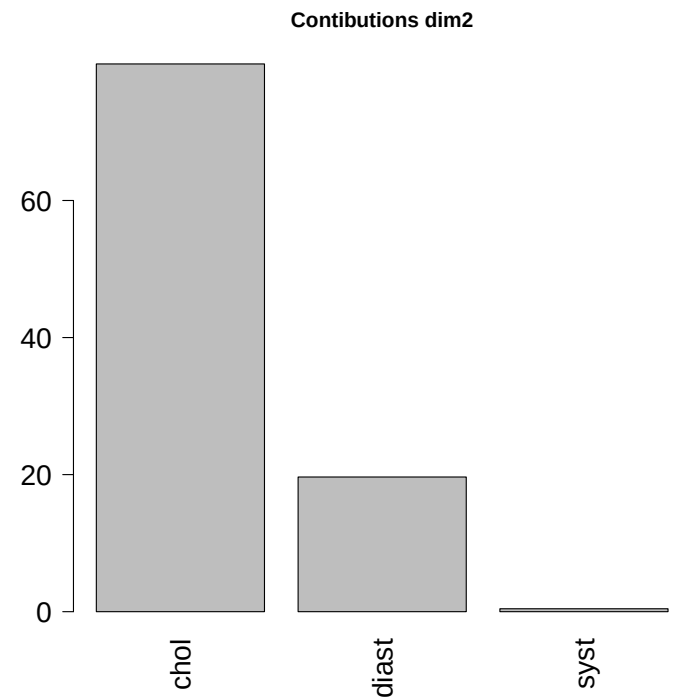
$$Ctr(j,(\alpha,\alpha')) = \frac{a_{j\alpha}^2 + a_{j\alpha'}^2}{\lambda_\alpha + \lambda'_\alpha}.$$

Example : the set of 3 variables.

Relative contributions (en percent) of the 3 variables :

```
##       G1    G2
## diast 41 19.65
## syst  52  0.42
## chol   7 79.92
```

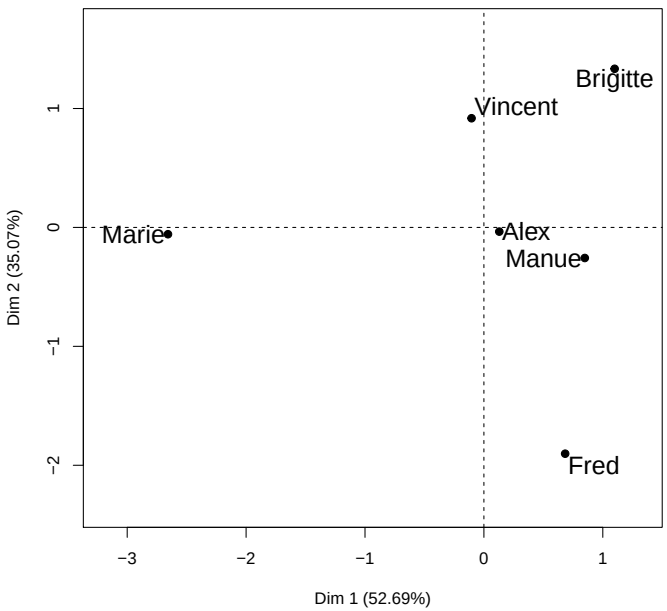


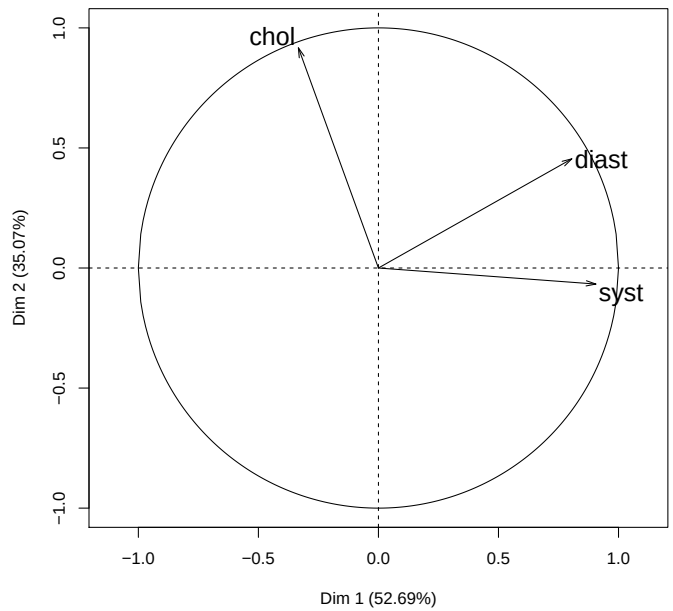


Interpretation of the projection plan of the observations using the correlation circle.

$$a_{j\alpha} = cor(\mathbf{x}^j, \mathbf{f}^\alpha)$$

```
##       Dim.1  Dim.2
## diast  0.81  0.455
## syst   0.91 -0.067
## chol  -0.33  0.917
```





Give an interpretation of the position (left, rigth, top, bottom) of the observations according to the position of the variables in the correlation circle.

Outline

PCA with metrics and GSVD

Let $\mathbf{Z}$ be a real data matrix of dimension $n \times p$. Let $\mathbf{N}$ (resp. $\mathbf{M}$) the diagonal matrix of the weights of the n rows (resp. the weights of the p columns).

The Generalized Singular Value Decomposition (GSVD) of $\mathbf{Z}$ with metrics $\mathbf{N}$ in $\mathbb{R}^n$ and $\mathbf{M}$ in $\mathbb{R}^p$ gives :

$$\mathbf{Z} = \mathbf{U}\Lambda\mathbf{V}^T \tag{1}$$

where

- $\Lambda = \text{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r})$ and $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_r}$ are the singular values defined as the square roots of the eigenvalues of $\mathbf{ZMZ}^T\mathbf{N}$ and $\mathbf{Z}^T\mathbf{NZM}$. Here r is the rank of $\mathbf{Z}$.
- $\mathbf{U}$ is the left singular vectors matrix of dimension $n \times r$. The left singular vectors are the r eigenvectors of $\mathbf{ZMZ}^T\mathbf{N}$ (with $\mathbf{U}^T\mathbf{NU} = \mathbb{I}_r$) ranked by decreasing order of the eigenvalues.
- $\mathbf{V}$ is the right singular vectors matrix of dimension $p \times r$. The right singular vectors are the r eigenvectors of $\mathbf{Z}^T\mathbf{NZM}$ (with $\mathbf{V}^T\mathbf{MV} = \mathbb{I}_r$) ranked by decreasing order of the eigenvalues.

Principal Components. The n rows of $\mathbf{Z}$ (observations) are $\mathbf{M}$-projected on the axes of orientation vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ in $\mathbb{R}^p$ obtained by solving the sequence (indexed by i) of the following optimization problems :

$$\begin{array}{lll} \max_{\mathbf{v}_i \in \mathbb{R}^p} & \|\mathbf{Z}\mathbf{M}\mathbf{v}_i\|^2_{\mathbf{N}} & \\ \text{subject to} & \mathbf{v}_i^T \mathbf{M} \mathbf{v}_j = 0 & \forall 1 \leq j < i, \\ & \mathbf{v}_i^T \mathbf{M} \mathbf{v}_i = 1. & \end{array} \tag{2}$$

The solutions are the eigenvectors of $\mathbf{Z}^T\mathbf{N}\mathbf{Z}\mathbf{M}$ i.e. the right singular vectors in (1).

The $n \times r$ matrix $\mathbf{F}$ of the coordinates of the $\mathbf{M}$-projections of the n observations on these r axes is by definition :

$$\mathbf{F} = \mathbf{Z}\mathbf{M}\mathbf{V}, \tag{3}$$

and it comes from (1) that :

$$\mathbf{F} = \mathbf{U}\Lambda. \tag{4}$$

The ith principal component (ith column of $\mathbf{F}$) is :

$$\mathbf{f}_i = \mathbf{Z}\mathbf{M}\mathbf{v}_i \in \mathbb{R}^n$$

and (2) gives :

$$\|\mathbf{f}_i\|^2_{\mathbf{N}} = \lambda_i.$$

Loadings. The p columns of $\mathbf{Z}$ (variables) are $\mathbf{N}$-projected on the axes of orientation vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r$ in $\mathbb{R}^n$ obtained by solving the sequence (indexed by i) of the following optimization problems :

$$\begin{array}{lll} \max_{\mathbf{u}_i \in \mathbb{R}^n} & \|\mathbf{Z}^T\mathbf{N}\mathbf{u}_i\|^2_{\mathbf{M}} & \\ \text{subject to} & \mathbf{u}_i^T\mathbf{N}\mathbf{u}_j = 0 & \forall 1 \leq j < i, \\ & \mathbf{u}_i^T\mathbf{N}\mathbf{u}_i = 1. & \end{array} \tag{5}$$

The solutions are the eigenvectors of $\mathbf{ZMZ}^T\mathbf{N}$ i.e. the left singular vectors in (1).

The $p \times r$ matrix $\mathbf{A}$ of the coordinates of the $\mathbf{N}$-projections of the p variables on these axes is by definition.

$$\mathbf{A} = \mathbf{Z}^T\mathbf{NU}, \tag{6}$$

It comes from (1) that :

$$\mathbf{A} = \mathbf{V}\Lambda. \tag{7}$$

The ith loadings vector (ith column of $\mathbf{A}$) is :

$$\mathbf{a}_i = \mathbf{Z}^T\mathbf{N}\mathbf{u}_i$$

and (5) gives :

$$\|\mathbf{a}_i\|^2_{\mathbf{M}} = \lambda_i.$$