Multiple Correspondance Analysis (MCA)

Marie Chavent

Master MAS, Université de Bordeaux

11 octobre 2019

Introduction

The aim is to expore categorical data

Example : 27 dogs described on 6 categorical variables.

```
##                Size Weight Velocity Intelligence Affectivity Aggressivness
## Beauceron       S++     W+      V++           I+         Af+           Ag+
## BassetHound      S-     W-       V-           I-         Af-           Ag+
## GermanShepherd  S++     W+      V++          I++         Af+           Ag+
## Boxer            S+     W+       V+           I+         Af+           Ag+
## Bulldog          S-     W-       V-           I+         Af+           Ag-
## BullMastiff     S++    W++       V-          I++         Af-           Ag+
## Poodle           S-     W-       V+          I++         Af+           Ag-
## Chihuahua        S-     W-       V-           I-         Af+           Ag-
```

The rows describe observations or individuals (the 27 dogs) and columns describe variables (the descriptors).

The aim is to know :

- which observations are similar ?
- which variables are linked ?

One can look at :

- the distance matrix between observations :

```
##               Beauceron BassetHound GermanShepherd Boxer Bulldog
## Beauceron             0          NA             NA    NA      NA
## BassetHound          NA           0             NA    NA      NA
## GermanShepherd       NA          NA              0    NA      NA
## Boxer                NA          NA             NA     0      NA
## Bulldog              NA          NA             NA    NA       0
```

But how to measure the distance between two observations described by categorical variables ?

- the χ^2 of independance between pairs of variables.

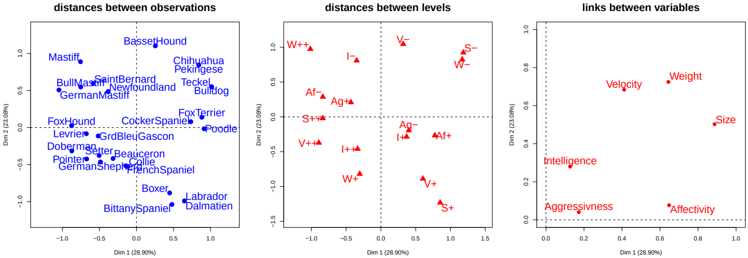
```
##              Size Weight Velocity Intelligence Affectivity Aggressivness
## Size         54.0   25.3    15.89          3.6       13.95          2.05
## Weight       25.3   54.0    18.47          1.4        9.48          2.55
## Velocity     15.9   18.5    54.00          3.2        2.97          0.57
## Intelligence  3.6    1.4     3.16         54.0        3.89          1.16
## Affectivity  14.0    9.5     2.97          3.9       23.14          0.91
## Aggressivness 2.1    2.6     0.57          1.2        0.91         23.14
```

The pvalues of the independance tests.

```
##               Size Weight Velocity Intelligence Affectivity Aggressivness
## Size         0.000  0.000    0.003         0.46       0.001          0.36
## Weight       0.000  0.000    0.001         0.85       0.009          0.28
## Velocity     0.003  0.001    0.000         0.53       0.227          0.75
## Intelligence 0.462  0.852    0.532         0.00       0.143          0.56
## Affectivity  0.001  0.009    0.227         0.14       0.000          0.34
## Aggressivness 0.359 0.279    0.750         0.56       0.339          0.00
```

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

- visualize on graphics distances between observations, distances between levels andlinks between categorical variables.



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

Categorical data

```
##                 Size Weight Velocity Intelligence
## Beauceron        S++     W+      V++           I+
## BassetHound       S-     W-       V-           I-
## GermanShepherd   S++     W+      V++          I++
## Boxer             S+     W+       V+           I+
## Bulldog           S-     W-       V-           I+
## BullMastiff      S++    W++       V-          I++
## Poodle            S-     W-       V+          I++
## Chihuahua         S-     W-       V-           I-
```

Numerical data

```
##                  PC1    PC2   PC3
## Beauceron      -0.32 -0.418 -0.10
## BassetHound     0.25  1.101 -0.19
## GermanShepherd -0.49 -0.464 -0.50
## Boxer           0.45 -0.882  0.69
## Bulldog         1.01  0.550 -0.16
## BullMastiff    -0.75  0.547  0.50
## Poodle          0.91 -0.016 -0.58
## Chihuahua       0.84  0.844 -0.47
```

- transforms categorical data into numerical data.

Outline

Basic concepts

We consider a categorical data table where n observations are described on p variables.

$$
\begin{array}{c|ccc|}
 & 1 \ldots & j & \ldots p \\
\hline
1 & & & \\
\vdots & & \vdots & \\
i & \ldots & x_{ij} & \ldots \\
\vdots & & \vdots & \\
n & & & \\
\hline
\end{array}
$$

Some notations :

- $\mathbf{X} = (x_{ij})_{n \times p}$ is the categorical data matrix whith $x_{ij} \in \mathcal{L}_j$ and $\mathcal{L}_j$ is the set of levels of the jth variable.
- $\ell_j = \text{card}(\mathcal{L}_j)$ is the number of levels of the jth variable.
- $\ell = \ell_1 + \ldots + \ell_p$ is the total number of levels.

Example : 27 dogs described on 6 categorical variables with a total of 16 levels.

```
##                Size Weight Velocity Intelligence Affectivity Aggressivness
## Beauceron       S++     W+      V++           I+         Af+           Ag+
## BassetHound      S-     W-       V-           I-         Af-           Ag+
## GermanShepherd  S++     W+      V++          I++         Af+           Ag+
## Boxer            S+     W+       V+           I+         Af+           Ag+
## Bulldog          S-     W-       V-           I+         Af+           Ag-
## BullMastiff     S++    W++       V-          I++         Af-           Ag+
## Poodle           S-     W-       V+          I++         Af+           Ag-
## Chihuahua        S-     W-       V-           I-         Af+           Ag-
```

Levels : S-,S+,S++ (size), W-,W+,W++ (weight), ...

$n =$ $p =$ $\mathbf{X} =$ $\ell_2 =$ $\ell =$

Two approaches for recoding categorical data into numerical data :

- the disjonctive table where each levels is coded as a binary variable,
- the Burt table (anglo-saxon approach) which gathers the contingency tables of all the pairs of variables.

The disjonctive table $\mathbf{K}$ describes the n observations on the ℓ levels :

$$\mathbf{K} = \begin{array}{c|ccc|}
 & 1 \ldots & s & \ldots \ell \\
\hline
1 & & & \\
\vdots & & \vdots & \\
i & \ldots & k_{is} & \ldots \\
\vdots & & \vdots & \\
\text{n} & & & \\
\hline\hline
\text{total} & & n_s &
\end{array}$$

Each column s is the indicator vector of the level s with :

$$\begin{cases} k_{is} = 1 \text{ if observation } i \text{ has level } s \\ k_{is} = 0 \text{ otherwise} \end{cases}$$

Let n_s denote the number of observations having level s.

Example of the dogs dataset :

Disjonctive table $\mathbf{K}$ of the $\ell = 16$ levels.

```
##                S- S+ S++ W- W+ W++ V- V+ V++ I- I+ I++ Af- Af+ Ag- Ag+
## Beauceron       0  0   1  0  1   0  0  0   1  0  1   0   0   1   0   1
## BassetHound     1  0   0  1  0   0  1  0   0  1  0   0   1   0   0   1
## GermanShepherd  0  0   1  0  1   0  0  0   1  0  0   1   0   1   0   1
## Boxer           0  1   0  0  1   0  0  1   0  0  1   0   0   1   0   1
```

Frequencies n_s of the $\ell = 16$ levels.

```
##  S-  S+ S++  W-  W+ W++  V-  V+ V++  I-  I+ I++ Af- Af+ Ag- Ag+
##   7   5  15   8  14   5  10   8   9   8  13   6  13  14  14  13
```

Relative frequencies $\frac{n_s}{n}$ of the $\ell = 16$ levels.

```
##   S-   S+  S++   W-   W+  W++   V-   V+  V++   I-   I+  I++  Af-  Af+  Ag-
## 0.26 0.19 0.56 0.30 0.52 0.19 0.37 0.30 0.33 0.30 0.48 0.22 0.48 0.52 0.52
##  Ag+
## 0.48
```

The centered disjonctive table $\mathbf{Z}$ describes the same n observations on the ℓ levels.

Matrix $\mathbf{K}$ of binary data

	$1\ldots$	s	$\ldots\ell$
1			
$\vdots$		$\vdots$	
i	$\ldots$	k_{is}	$\ldots$
$\vdots$		$\vdots$	
n			
mean	$\ldots$	$\frac{n_s}{n}$	$\ldots$

Matrix $\mathbf{Z}$ of centered binary data.

	$1\ldots$	s	$\ldots\ell$
1			
$\vdots$		$\vdots$	
i	$\ldots$	$z_{is} = k_{is} - \frac{n_s}{n}$	$\ldots$
$\vdots$		$\vdots$	
n			
mean	$\ldots$	0	$\ldots$
var	$\ldots$	$\frac{n_s}{n}(1 - \frac{n_s}{n})$	$\ldots$

One can check that $var(\mathbf{z}^s) = \frac{n_s}{n}(1 - \frac{n_s}{n})$ where $\mathbf{z}^s \in \mathbb{R}^n$ denotes s-th column of $\mathbf{Z}$.

Example of the dogs dataset :

Disjonctive table **K** of the $\ell = 16$ levels.

```
##                S- S+ S++ W- W+ W++ V- V+ V++ I- I+ I++ Af- Af+ Ag- Ag+
## Beauceron       0  0   1  0  1   0  0  0   1  0  1   0   0   1   0   1
## BassetHound     1  0   0  1  0   0  1  0   0  1  0   0   1   0   0   1
## GermanShepherd  0  0   1  0  1   0  0  0   1  0  0   1   0   1   0   1
## Boxer           0  1   0  0  1   0  0  1   0  0  1   0   0   1   0   1
```

Relative frequencies (means) $\frac{n_s}{n}$ of the $\ell = 16$ levels.

```
##   S-   S+  S++   W-   W+  W++   V-   V+  V++   I-   I+  I++  Af-  Af+  Ag-
## 0.26 0.19 0.56 0.30 0.52 0.19 0.37 0.30 0.33 0.30 0.48 0.22 0.48 0.52 0.52
##  Ag+
## 0.48
```

Centered disjonctive table **Z** of the $\ell = 16$ levels.

```
##                   S-    S+   S++   W-    W+   W++    V-   V+   V++   I-
## Beauceron      -0.26 -0.19  0.44 -0.3  0.48 -0.19 -0.37 -0.3  0.67 -0.3
## BassetHound     0.74 -0.19 -0.56  0.7 -0.52 -0.19  0.63 -0.3 -0.33  0.7
## GermanShepherd -0.26 -0.19  0.44 -0.3  0.48 -0.19 -0.37 -0.3  0.67 -0.3
## Boxer          -0.26  0.81 -0.56 -0.3  0.48 -0.19 -0.37  0.7 -0.33 -0.3
##                   I+   I++   Af-   Af+   Ag-  Ag+
## Beauceron       0.52 -0.22 -0.48  0.48 -0.52 0.52
## BassetHound    -0.48 -0.22  0.52 -0.52 -0.52 0.52
## GermanShepherd -0.48  0.78 -0.48  0.48 -0.52 0.52
## Boxer           0.52 -0.22 -0.48  0.48 -0.52 0.52
```

Three sets are studied in MCA.

- The set of observations where each observation i is :
  - described by a vector $\mathbf{z}_i$ $\mathbb{R}^\ell$ (a row of $\mathbf{Z}$),
  - weighted by w_i with usually $w_i = \frac{1}{n}$.
- The set of levels where each level s is :
  - described by a vector $\mathbf{z}^s$ in $\mathbb{R}^{,}$ (a column of $\mathbf{Z}$),
  - weighted by m_s with $m_s = \frac{n}{n_s}$.
- The set of variables where each categorical variable j is described by a vector $\mathbf{x}^j$ in $\mathcal{L}^n$ (a column of $\mathbf{X}$).

Proximity between two observations is measured with the so called χ^2 distance.

- A weight m_s is associated with each level s in order to give more importance to rare levels :

$$m_s = \frac{n}{n_s}$$

- The χ^2 distance between two observations is the Euclidean distance with metric $\mathbf{M} = diag(\frac{n}{n_s}, s = 1 \ldots, \ell)$ on $\mathbb{R}^\ell$:

$$\begin{aligned} d^2_{\mathbf{M}}(\mathbf{z}_i, \mathbf{z}_{i'}) &= \sum_{s=1}^{\ell} \frac{n}{n_s}(z_{is} - z_{i's})^2 \\ &= \sum_{s=1}^{\ell} \frac{n}{n_s}(k_{is} - k_{i's})^2 \end{aligned}$$

Two observations are different if they have different levels, with more weight in the distance for rare levels (n_s small).

Example : distance between Beauceron and BassetHound

```
##             S- S+ S++ W- W+ W++ V- V+ V++ I- I+ I++ Af- Af+ Ag- Ag+
## Beauceron    0  0   1  0  1   0  0  0   1  0  1   0   0   1   0   1
## BassetHound  1  0   0  1  0   0  1  0   0  1  0   0   1   0   0   1
```

Relative frequencies of the levels :

```
##   S-   S+  S++   W-   W+  W++   V-   V+  V++   I-   I+  I++  Af-  Af+  Ag-
## 0.26 0.19 0.56 0.30 0.52 0.19 0.37 0.30 0.33 0.30 0.48 0.22 0.48 0.52 0.52
##  Ag+
## 0.48
```

$$d^2_{\mathbf{M}}(\mathbf{z}_1, \mathbf{z}_2) = \frac{1}{0.26}(0-1)^2 + \frac{1}{0.19}(0-0)^2 + \ldots + \frac{1}{0.48}(1-1)^2$$

The dispersion with metric $\mathbf{M}$ of the set of observations in $\mathbb{R}^{\ell}$ is measured by the inertia.

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

$$I(\mathbf{Z}) = \frac{1}{n} \sum_{i=1}^{n} d_{\mathbf{M}}{}^2(\mathbf{z}_i, \bar{\mathbf{z}}).$$

- Inertia is a generalization of the variance to the case of multivariate data (ℓ variables).

- One can show that :

$$I(\mathbf{Z}) = \sum_{s=1}^{\ell} m_s var(\mathbf{z}^s),$$

  where $m_s = \frac{n}{n_s}$ is the weight of the column (the level) s.

- The inertia of the set of observations is then the (weighted) sum of the variances of the columns.

- When the rows are weighted by $\frac{1}{n}$:

$$I(\mathbf{Z}) = \sum_{s=1}^{\ell}(1 - \frac{n_s}{n}).$$

In practice :

- The contribution of a level s to the inertia of $\mathbf{Z}$ is all the more important as the level is rare.
- Too rare levels are then avoided (by pre-processing for instance).

- This gives :

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

  In practice :

  - The contribution of a variable j to the inertia of $\mathbf{Z}$ is all the more imporant as its number of levels ℓ_j is high.
  - Variables with too different number of levels are then avoided (by pre-processing for instance).

- And also :

$$I(\mathbf{Z}) = \ell - p.$$

  Example of the dogs dataset :

  Number of variables $p = 6$, number of levels $\ell = 16$,

$$I(\mathbf{Z}) = 16 - 6 = 10.$$

The link between a numerical variable $\mathbf{y}$ and a categorical variable $\mathbf{x}^j$ is measured by the correlation ratio :

$$\eta^2(\mathbf{y}|\mathbf{x}^j) = \frac{var(\bar{\mathbf{y}}|\mathbf{x}^j)}{var(\mathbf{y})} = \frac{\sum_{s=1}^{\ell_j} \frac{n_s}{n}(\bar{\mathbf{y}}_s - \bar{\mathbf{y}})^2}{\sum_{i=1}^{n} \frac{1}{n}(y_i - \bar{\mathbf{y}})^2}$$

where ℓ_j is the number of levels of $\mathbf{x}^j$ and $\bar{\mathbf{y}}_s$ is the mean value of $\mathbf{y}$ performed with the observations having level s.

- This criterion takes its values in $[0, 1]$.
- It measures the proportion of the variance of the numerical variable y explained by the categorical variable $\mathbf{x}^j$.

In which situation is this criterion equal to 0, equal to 1 ?

Example : the Iris dataset.

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 1            5.1         3.5          1.4         0.2     setosa
## 2            4.9         3.0          1.4         0.2     setosa
## 50           5.0         3.3          1.4         0.2     setosa
## 51           7.0         3.2          4.7         1.4 versicolor
## 100          5.7         2.8          4.1         1.3 versicolor
## 101          6.3         3.3          6.0         2.5  virginica
```

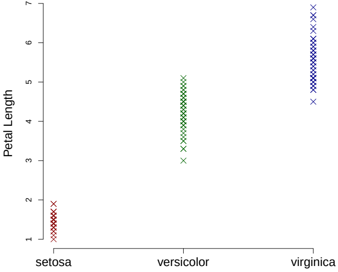
Correlation ratios between the categorical variable Species and the 4 numerical variables :

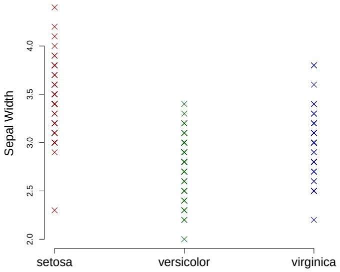
```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##         0.62         0.40         0.94         0.93
```

Species explains :

- 94 % of the variance of "Petal Length".
- 40 % of the variance of "Sepal Width".

Species is then more linked to Petal Length than to Sepal Width.
This can be visualized here :





MCA analysis :

- either the Burt table (anglo-saxon approach),
- or the centered disjonctive table $\mathbf{Z}$.

This leads to different methods of MCA :

- Correspondance Analysis (CA) of
  - either the Burt table,
  - or the disjonctive table,
- Principal Component Analysis with metrics of the centered disjonctive table $\mathbf{Z}$.

From now PCA with metrics of the centered disjonctive table $\mathbf{Z}$ is considered.

Outline

An MCA algorithm

MCA is presented here as a PCA with metrics (see lecture 1) of the centered disjonctive table.

Step 1 : the pre-processing step.

1. Build the centered disjonctive table $\mathbf{Z}$ of dimension $n \times \ell$.
2. Build the metrics $\mathbf{N}$ in $\mathbb{R}^n$ and $\mathbf{M}$ in $\mathbb{R}^\ell$:
   - $\mathbf{N}$ is the diagonal matrix of the weights of the observations i.e. when observations are weighted by $w_i = \frac{1}{n}$:
   $$\mathbf{N} = \frac{1}{n}\mathbb{I}_n.$$
   - $\mathbf{M}$ is the diagonal matrix of the weights of the levels i.e. when the levels are weighted by the inverse of the relative frequencies $m_s = \frac{n}{n_s}$:
   $$\mathbf{M} = \text{diag}(\frac{n}{n_1}, \ldots, \frac{n}{n_\ell})$$

Step 2 : the GSVD step.

The Generalized Singular Value Decomposition (GSVD) of $\mathbf{Z}$ with metrics $\mathbf{N}$ and $\mathbf{M}$ is :

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

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 $\ell \times r$. 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.

Step 3 : Analysis of the set of observations.

The $n \times r$ matrix $\mathbf{F}$ of the factor coordinates of the observations projected on the r axes $\Delta_1, \ldots, \Delta_r$ is given by :

$$\mathbf{F} = \mathbf{ZMV}, \tag{2}$$

and we deduce from (1) that :

$$\mathbf{F} = \mathbf{U}\Lambda. \tag{3}$$

$$
\mathbf{F} = \begin{array}{c|ccc|}
 & 1 \dots & \alpha & \dots r \\
\hline
1 & & & \\
\vdots & & \vdots & \\
i & \dots & f_{i\alpha} & \dots \\
\vdots & & \vdots & \\
n & & & \\
\hline\hline
\text{mean} & \dots & 0 & \dots \\
\text{var} & \dots & \lambda_\alpha & \dots
\end{array}
$$

- The columns $\mathbf{f}^\alpha$ of $\mathbf{F}$ are the principal components with :

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

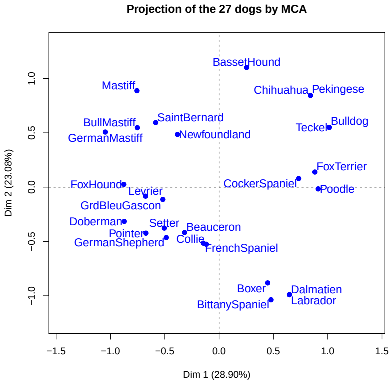
$$var(\mathbf{f}^\alpha) = \lambda_\alpha.$$

- The left singular vectors are the standardized principal components :

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

Example of the 27 dogs : $q = 2$ first principal components i.e. two first columns of $\mathbf{F}$.

```
##                PC1    PC2
## Beauceron     -0.32 -0.418
## BassetHound    0.25  1.101
## GermanShepherd -0.49 -0.464
## Boxer          0.45 -0.882
## Bulldog        1.01  0.550
## BullMastiff   -0.75  0.547
## Poodle         0.91 -0.016
## Chihuahua      0.84  0.844
## CockerSpaniel  0.73  0.079
## Collie        -0.12 -0.526
## Dalmatien      0.65 -0.990
## Doberman      -0.87 -0.315
## GermanMastiff -1.05  0.507
## BittanySpaniel 0.48 -1.037
## FrenchSpaniel -0.14 -0.516
## FoxHound      -0.88  0.025
## FoxTerrier     0.88  0.139
## GrdBleuGascon -0.52 -0.113
## Labrador       0.65 -0.990
## Levrier       -0.68 -0.083
## Mastiff       -0.76  0.888
## Pekingese      0.84  0.844
## Pointer       -0.67 -0.424
## SaintBernard  -0.58  0.594
## Setter        -0.50 -0.377
## Teckel         1.01  0.550
## Newfoundland  -0.38  0.485
```



Step 4 : Analysis of the set of levels.

The $\ell \times r$ matrix $\mathbf{A}$ of the factor coordinates of the levels projected on the r axes $G_1, \ldots, G_r$ is given by :

$$\mathbf{A} = \mathbf{M}\mathbf{Z}^T\mathbf{N}\mathbf{U}, \tag{4}$$

and we deduce from (1) that :

$$\mathbf{A} = \mathbf{M}\mathbf{V}\Lambda. \tag{5}$$

$$\mathbf{A}= \begin{array}{c|ccc|} & 1\ldots & \alpha & \ldots r \\ \hline 1 & & & \\ \vdots & & \vdots & \\ s & \ldots & a_{s\alpha} & \ldots \\ \vdots & & \vdots & \\ \ell & & & \\ \hline \end{array}$$

- One can show that :

$$a_{s\alpha} = \frac{1}{n_s} \sum_{i:k_{is}=1} \frac{f_{i\alpha}}{\sqrt{\lambda_\alpha}}$$

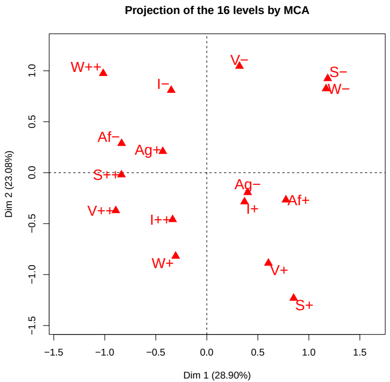
  i.e. $a_{s\alpha}$ is the mean value of the (standardized) factor coordinates of observations having level s ?

- This relation is called the (quasi) barycentric property.

This property is crucial for MCA results interpretation.

Example of the 16 levels of the dogs dataset : $q = 2$ first columns of $\mathbf{A}$

```
##      Dim 1  Dim 2
## S-    1.18  0.924
## S+    0.85 -1.232
## S++  -0.84 -0.021
## W-    1.17  0.824
## W+   -0.31 -0.819
## W++  -1.02  0.974
## V-    0.32  1.045
## V+    0.60 -0.888
## V++  -0.89 -0.372
## I-   -0.35  0.809
## I+    0.37 -0.286
## I++  -0.34 -0.459
## Af-  -0.84  0.287
## Af+   0.78 -0.267
## Ag-   0.40 -0.194
## Ag+  -0.43  0.209
```



Check the barycentric property for the level W++ knowing that the 5 dogs of having a weight W++ are : BullMastiff, GermanMastiff, Mastiff, SaintBernard, Newfoundland and that the variances of the two first PC are 0.48 and 0.38. .

Step 5 : Analysis of the set of variables.

The contribution $c_{j\alpha}$ of the variable $\mathbf{x}^j$ (jth column of $\mathbf{X}$) to the variance of the principal component $\mathbf{f}^\alpha$ is defined by :

$$c_{j\alpha} = \sum_{s \in \mathcal{L}_j} \frac{n_s}{n} a_{s\alpha}^{*2}. \tag{6}$$

The contributions matrix

$$\mathbf{C} = (c_{j\alpha})_{p \times r},$$

is also called the squared loadings matrix to draw an analogy with squared loadings in PCA.

$$
\mathbf{C}= \begin{array}{c|ccc|}
 & 1 \ldots & \alpha & \ldots r \\
\hline
1 & & & \\
\vdots & & \vdots & \\
j & \ldots & c_{j\alpha} & \ldots \\
\vdots & & \vdots & \\
p & & & \\
\hline
\end{array}
$$

The contribution $c_{j\alpha}$ of the categorical variable $\mathbf{x}^j$ to the variance of the principal component $\mathbf{f}^{\alpha}$ is equal to the correlation ratio between $\mathbf{x}^j$ and $\mathbf{f}^{\alpha}$:

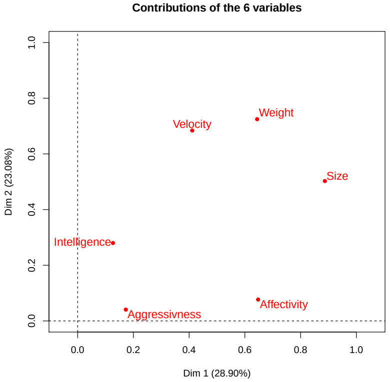
$$
\begin{aligned}
c_{j\alpha} &= \eta^2(\mathbf{f}^{\alpha}|\mathbf{x}^j) \\
&= \frac{var(\bar{\mathbf{f}}^{\alpha}|\mathbf{x}^j)}{var(\mathbf{f}^{\alpha})} = \frac{\sum_{s\in\mathcal{L}_j} \frac{n_s}{n}(\bar{\mathbf{f}}^{\alpha}_s - \bar{\mathbf{f}}^{\alpha})^2}{\sum_{i=1}^{n} \frac{1}{n}(f_{i\alpha} - \bar{\mathbf{f}}^{\alpha})^2}
\end{aligned}
$$

where $\bar{\mathbf{f}}^{\alpha}_s$ is the mean value of the principal component scores of observations having level s of the variable j.

This property is crucial for MCA results interpretation.

Example of the 6 categorical variables of the dogs dataset : $q = 2$ first columns of $\mathbf{C}$.

```
##               Dim 1 Dim 2
## Size           0.89 0.502
## Weight         0.64 0.725
## Velocity       0.41 0.684
## Intelligence   0.13 0.280
## Affectivity    0.65 0.077
## Aggressivness  0.17 0.041
```



Outline

Interpretation of the MCA results

Variance of the principal components.

Principal components (columns of $\mathbf{F}$) are q new synthetic numerical 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 MCA (matrix $\mathbf{F}_q$ of the q first columns of $\mathbf{F}$) is :

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

Example of the set of 27 dogs : the $r = \min(n-1, \ell - p)$ non null eigenvalues are :

```
##  lambda1  lambda2  lambda3  lambda4  lambda5  lambda6  lambda7  lambda8
##     2.89     2.31     1.27     0.95     0.90     0.74     0.49     0.27
##  lambda9 lambda10
##     0.14     0.05
```

then

$$\begin{aligned} var(\mathbf{f}^1) &= 2.89 \\ var(\mathbf{f}^2) &= 2.31 \end{aligned}$$

and the inertia of the 27 dogs projected on the $q = 2$ first dimensions of MCA is :

$$\lambda_1 + \lambda_2 = 2.89 + 2.31 = 5.19.$$

Total inertia.

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

$$I(\mathbf{Z}) = \sum_{s=1}^{\ell} \frac{n}{n_s} var(\mathbf{z}^s) = \ell - p.$$

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

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

Example :

Inertia of the 27 dogs projected on the $q = 10$ (all) principal components :

$$I(\mathbf{F}) = \lambda_1 + \ldots + \lambda_{10} = 2.89 + \ldots + 0.05 = 10$$

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 :

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

Warning : In MCA, the percentages of inertia explained by the principal components are "small" by construction. Some authors have proposed corrections of the eigenvalues in MCA (Greenacre, 1993).

Example : the set of 27 dogs.

Original data ($p = 6$ et ℓ=16)

```
##                  Size Weight Velocity Intelligence
## Beauceron         S++     W+      V++           I+
## BassetHound        S-     W-       V-           I-
## GermanShepherd    S++     W+      V++          I++
## Boxer              S+     W+       V+           I+
## Bulldog            S-     W-       V-           I+
```

Reduction to 3 PCs

```
##                 Dim 1 Dim 2 Dim 3
## Beauceron       -0.32 -0.42 -0.10
## BassetHound      0.25  1.10 -0.19
## GermanShepherd -0.49 -0.46 -0.50
## Boxer            0.45 -0.88  0.69
## Bulldog          1.01  0.55 -0.16
```

What is the quality of this reduction ?

```
##        Eigenvalue Proportion Cumulative
## dim 1       2.890      28.90         29
## dim 2       2.308      23.08         52
## dim 3       1.266      12.66         65
## dim 4       0.945       9.45         74
## dim 5       0.901       9.01         83
## dim 6       0.740       7.40         90
## dim 7       0.489       4.89         95
## dim 8       0.274       2.74         98
## dim 9       0.141       1.41        100
## dim 10      0.046       0.46        100
```

- $r = 10$ non nul eigenvalues because $r = \min(n-1, \ell - p) = 10$,
- The sum of the eigenvalues is $\ell - p = 10$ (total inertia),
- 64.6 % of the inertia is exaplined by the 3 first PCs.

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 $\mathbb{R}^\ell$ knowing that in MCA distances between observations are small when observations have same levels.

- The quality of the projection of an observation i on the projection 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 projection 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.

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

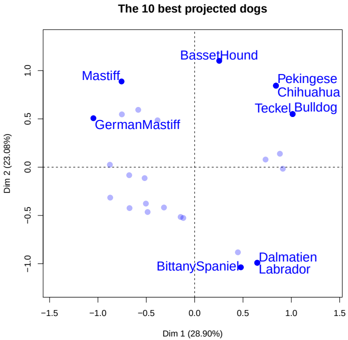
$$Ctr(i, \alpha) = \frac{w_i f_{i\alpha}^2}{\lambda_\alpha}.$$

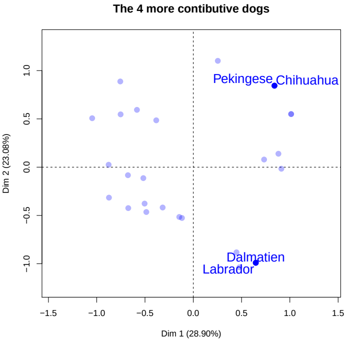
- The relative contribution of an observation i to the inertia on the plan $(\Delta_\alpha, \Delta'_\alpha)$ is

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

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 of the 27 dogs.





- Interpret the distances between Mastiff and German Mastiff, between Mastif and Labrador.
- Is any dog contributing excessively ?

Interpretation of the projection plans of the levels.

If two levels are well projected, then their distance on the projection plane can be interpreted using the barycentric property :

- two levels of different variables are close if they are owned by the same observations.
- two levels of the same variable are close if the two associated groups of observations are close.

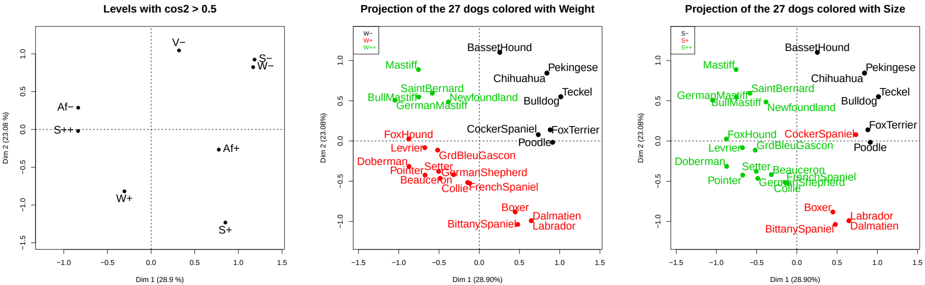
- The quality of the projection of a level s on the projection axis G_α is measured by the square cosine of the angle $\theta_{s\alpha}$ between the point $\mathbf{z}^s$ and the axis G_α :

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

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

$$\cos^2(\theta_{s(\alpha,\alpha')}) = \frac{a_{s\alpha}^2 + a_{s\alpha'}^2}{\|\mathbf{z}^s\|_{\mathbf{N}}^2}.$$

Example of the 27 dogs.



Interpret the distance between the levels W_ and S_ on this projection plan.

The levels 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 $\lambda_\alpha = \sum_{s=1}^{\ell} \frac{n_s}{n} a_{s\alpha}^2$.
- The relative contribution of a level s to the inertia on the axis Δ_α is :

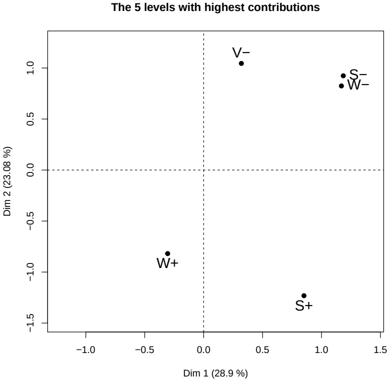
$$Ctr(s, \alpha) = \frac{n_s}{n} \frac{a_{s\alpha}^2}{\lambda_\alpha}.$$

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

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

Warning : the levels far from the center of projection plan are not necessary the one with highest contribution.

Example of the 27 dogs.



- Is here any level contributing excessively ?
- Why ?

Interpretation of the contribution map of the variables.

The abscissa and the ordinate are correlation ratios between the categorical variables and the principal components.

- The absolute contribution of a categorical variable j to the variance of the principal component $\mathbf{f}^\alpha$ is the sum of the contributions of its levels :

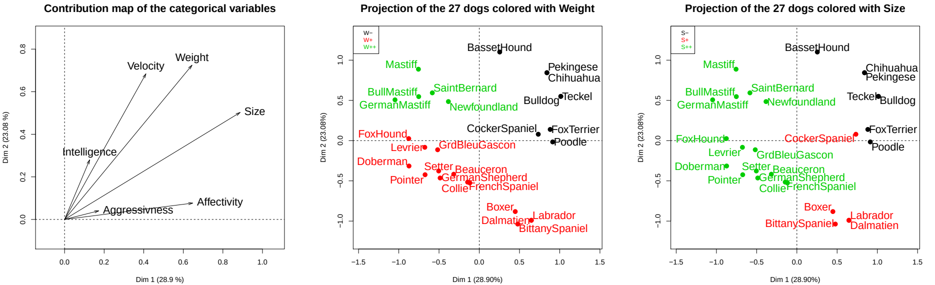
$$c_{j\alpha} = \sum_{s \in \mathcal{L}_j} \frac{n_s}{n} a_{s\alpha}^2$$

- Moreover, this absolute contribution is the correlation ratio between the categorical variable $\mathbf{x}^j$ and the principal component $\mathbf{f}^\alpha$:

$$c_{j\alpha} = \eta^2(\mathbf{f}^\alpha | \mathbf{x}^j)$$

The correlation ratio is a signless measure of links between categorical and numerical variables taking its values in $[0, 1]$.

Example of the 27 dogs.



- Which variable is linked to the first PC ?
- Which variable is linked to the second PC ?

Simultaneous representation of the observations and the levels.

First possibility : plot the levels at the barycenter of the observations.

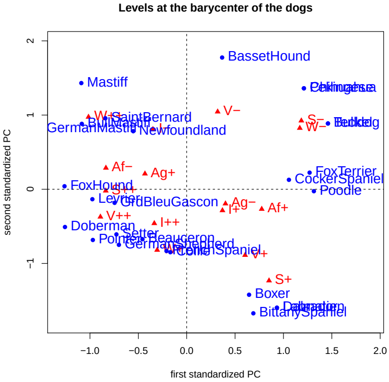
The barycentric property gives :

$$a_{s\alpha} = \frac{1}{n_s} \sum_{i:k_{is}=1} \frac{f_{i\alpha}}{\sqrt{\lambda_\alpha}},$$

Then

- observations are plotted with their standardized factor coordinates $\frac{f_{i\alpha}}{\sqrt{\lambda_\alpha}}$,
- levels are plotted with their factor coordinates $a_{s\alpha}$.

Example : levels at the barycenter of the dogs



For instance W++ is plotted at the barycenter of the dogs BullMastiff, GermanMastiff, Mastiff, SaintBernard, Newfoundland (of weight W++).

Second possibility : plot the levels at the quasi-barycenter of the observations.

The quasi-barycentric property is simply the barycentric property written as follows :

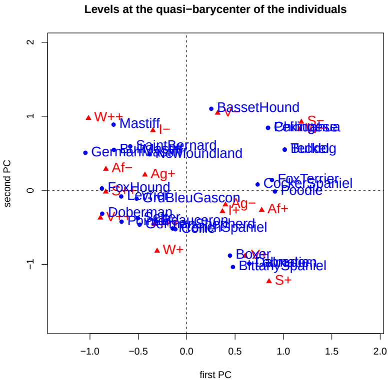
$$a_{s\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \left(\frac{1}{n_s} \sum_{i:k_{is}=1} f_{i\alpha} \right)$$

Then

- observations are plotted with their factor coordinates $f_{i\alpha}$,
- levels are plotted according to their factor coordinates $a_{s\alpha}$.

Levels are then at the barycenter of the observations with dilatation coefficient $\frac{1}{\sqrt{\lambda_\alpha}}$ in each dimension.

Example : levels at the quasi-barycenter of the dogs



For instance W++ is plotted at the quasi-barycenter of the dogs BullMastiff, GermanMastiff, Mastiff, SaintBernard, Newfoundland (of weight W++) i.e. the barycenter dilated by

- $\frac{1}{\sqrt{\lambda_1}} = \frac{1}{\sqrt{0.481}} = 1.44$ in the first dimension,
- $\frac{1}{\sqrt{\lambda_2}} = \frac{1}{\sqrt{0.384}} = 1.61$ in the second dimension.

Outline

MCA implementation

The procedure CORRESP of SAS.

- Implements correspondance analysis (CA) of the Burt table.
- Gives the factor coordinates for the levels but not for the observations by default.

The Burt table is a symmetric table of size $\ell \times \ell$ gathering contingency tables :

$$\mathbf{B} = \mathbf{K}^T\mathbf{K} = \begin{array}{c|ccc|c} & 1 \dots & s' & \dots \ell & \\ \hline 1 & & & & \\ \vdots & & \vdots & & \\ s & \dots & b_{ss'} & \dots & \\ \vdots & & \vdots & & \\ \ell & & & & \\ \hline & & & & \end{array}$$

where :

- $b_{ss'} = \sum_{i=1}^{n} k_{is} k_{is'}$ is the number of individual having both levels s and s'
- $b_{ss} = n_s$ is the number of individuals having s.

Example : Burt table of the $\ell = 16$ levels of the dogs dataset.

	S-	S+	S++	W-	W+	W++	V-	V+	V++	I-	I+	I++	Af-	Af+	Ag-	Ag+
S-	7	0	0	7	0	0	5	2	0	3	3	1	1	6	5	2
S+	0	5	0	1	4	0	1	4	0	0	4	1	0	5	3	2
S++	0	0	15	0	10	5	4	2	9	5	6	4	12	3	6	9
W-	7	1	0	8	0	0	6	2	0	3	4	1	1	7	5	3
W+	0	4	10	0	14	0	0	6	8	3	7	4	7	7	8	6
W++	0	0	5	0	0	5	4	0	1	2	2	1	5	0	1	4
V-	5	1	4	6	0	4	10	0	0	4	5	1	5	5	5	5
V+	2	4	2	2	6	0	0	8	0	1	5	2	2	6	5	3
V++	0	0	9	0	8	1	0	0	9	3	3	3	6	3	4	5
I-	3	0	5	3	3	2	4	1	3	8	0	0	6	2	3	5
I+	3	4	6	4	7	2	5	5	3	0	13	0	4	9	8	5
I++	1	1	4	1	4	1	1	2	3	0	0	6	3	3	3	3
Af-	1	0	12	1	7	5	5	2	6	6	4	3	13	0	5	8
Af+	6	5	3	7	7	0	5	6	3	2	9	3	0	14	9	5
Ag-	5	3	6	5	8	1	5	5	4	3	8	3	5	9	14	0
Ag+	2	2	9	3	6	4	5	3	5	5	5	3	8	5	0	13

The function MCA of the R package FactoMineR.

- Implement correspondance analysis (CA) of the of the disjonctive table.
- Implements then two PCA with metrics : one with the row profiles matrix, one with the column profiles matrix.
- Gives directly the factor coordinates of both levels and observations.

The function PCAmix of the R package PCAmixdata.

- Implement a single PCA with metrics of the of the centered disjonctive table.
- Gives almost similar results than the MCA function :
  - factor coordinates of the levels are identical.
  - factor coordinates of the observations are multiplied by $\sqrt{p}$.
  - total inertia is multiplied by p and is equal to $\ell - p$.