

# L'Analyse des Correspondances Multiples (ACM)

Marie Chavent

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

2014-2015

Il s'agit d'une méthode de description statistique multidimensionnelle d'un tableau de **données qualitatives**. Comme l'ACP pour les données quantitatives, elle permet pour des données qualitatives :

- des représentations graphiques du contenu du tableau de données : représentation des “similitudes” entre les individus et entre les modalités des variables qualitatives.
- un recodage en données numériques, sur lesquelles on peut ensuite appliquer d'autres méthodes (par exemple des algorithmes de classification).

En pratique, l'ACM est une AFC appliquée au tableau disjonctif complet c'est à dire à la matrice des indicatrices des modalités des variables qualitatives.

## 1 Tableau de données, tableau disjonctif complet, tableau de Burt

Matrice des données qualitatives  $\mathbf{H}$  de dimension  $n \times p$  :

$$\mathbf{H} = \begin{array}{c|cccc} & 1 & \dots & j & \dots & p \\ \hline 1 & & & & & \\ \vdots & & & \vdots & & \\ i & & \dots & h_{ij} & \dots & \\ \vdots & & & \vdots & & \\ n & & & & & \\ \hline & & & & & \end{array}$$

où  $n$  est le nombre d'individus,  $p$  le nombre de variables qualitatives, et  $h_{ij} \in \mathcal{M}_j$  avec  $\mathcal{M}_j$  qui est l'ensemble des modalités de la  $j$ ème variable. On note  $m_j = \text{card}(\mathcal{M}_j)$  le nombre de modalités de  $j$  et  $m = m_1 + \dots + m_p$  le nombre total de modalités.

Tableau disjonctif complet  $\mathbf{K}$  de dimension  $n \times m$  :

Chaque colonne  $s$  est l'indicatrice de la modalité  $s$  avec :

	1 ... s ... m
1	
⋮	⋮
$\mathbf{K} =$ $i$	... $k_{is}$ ...
⋮	⋮
$n$	

$$\begin{cases} k_{is} = 1 \text{ si l'individu } i \text{ possède la modalité } s \\ k_{is} = 0 \text{ sinon} \end{cases}$$

Tableau de Burt  $\mathbf{B}$  de dimension  $m \times m$  :

Il s'agit d'un tableau symétrique qui rassemble les croisements deux à deux de toutes les variables, c'est à dire tous les tableaux de contingence des variables deux à deux.

	1 ... $s'$ ... m
1	
⋮	⋮
$\mathbf{B} = \mathbf{K}^t \mathbf{K} =$ $s$	... $b_{ss'}$ ...
⋮	⋮
$m$	

Avec :

- $b_{ss'} = \sum_{i=1}^n k_{is} k_{is'}$  qui est le nombre d'individus qui possèdent la modalité  $s$  et la modalité  $s'$
- et sur la diagonale  $b_{ss} = n_s$  le nombre d'individus qui possèdent la modalité  $s$ .

**Exemple.** On considère un tableau de données fictives où 27 races de chiens sont décrites avec 7 variables qualitatives (Exemple du livre de Gilbert Saporta).

```
load("chiens.rda")
chiens
```

```
##      taille poids velocite intellig affect agress fonction
## beauceron  T++  P+    V++    I+   Af+  Ag+  Utilite
## basset     T-   P-    V-    I-   Af-  Ag+  Chasse
## ber_alle  T++  P+    V++    I++  Af+  Ag+  Utilite
## boxer      T+   P+    V+    I+   Af+  Ag+  Compagnie
## bull-dog   T-   P-    V-    I+   Af+  Ag-  Compagnie
## bull-mass  T++  P++   V-    I++  Af-  Ag+  Utilite
## caniche    T-   P-    V+    I++  Af+  Ag-  Compagnie
## chihuahua  T-   P-    V-    I-   Af+  Ag-  Compagnie
## cocker     T+   P-    V-    I+   Af+  Ag+  Compagnie
## colley     T++  P+    V++    I+   Af+  Ag-  Compagnie
```

```
## dalmatien      T+   P+   V+   I+   Af+   Ag- Compagnie
## dobermann     T++  P+   V++  I++  Af-   Ag+  Utilite
## dogue_all     T++  P++  V++  I-   Af-   Ag+  Utilite
## epagn_bre     T+   P+   V+   I++  Af+   Ag-  Chasse
## epagn_fra     T++  P+   V+   I+   Af-   Ag-  Chasse
## fox_hound     T++  P+   V++  I-   Af-   Ag+  Chasse
## fox_terri     T-   P-   V+   I+   Af+   Ag+ Compagnie
## grand_ble     T++  P+   V+   I-   Af-   Ag+  Chasse
## labrador      T+   P+   V+   I+   Af+   Ag-  Chasse
## levrier       T++  P+   V++  I-   Af-   Ag-  Chasse
## mastiff       T++  P++  V-   I-   Af-   Ag+  Utilite
## pekinois      T-   P-   V-   I-   Af+   Ag- Compagnie
## pointer       T++  P+   V++  I++  Af-   Ag-  Chasse
## saint_ber     T++  P++  V-   I+   Af-   Ag+  Utilite
## setter        T++  P+   V++  I+   Af-   Ag-  Chasse
## teckel        T-   P-   V-   I+   Af+   Ag- Compagnie
## terre_neu     T++  P++  V-   I+   Af-   Ag-  Utilite
```

*#on supprime la variable Utility qui sera illustrative*

```
H <- subset(chiens,select=-fonction)
```

## Le tableau disjonctif complet :

```
library(FactoMineR)
K <- tab.disjonctif(H)
K

##          T-  T+  T++  P-  P+  P++  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
## basset     1  0  0  1  0  0  1  0  0  1  0  0  1  0  0  1
## ber_allem  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
## bull-dog   1  0  0  1  0  0  1  0  0  0  1  0  0  1  1  0
## bull-mass  0  0  1  0  0  1  1  0  0  0  0  1  1  0  0  1
## caniche    1  0  0  1  0  0  0  1  0  0  0  1  0  1  1  0
## chihuahua  1  0  0  1  0  0  1  0  0  1  0  0  0  1  1  0
## cocker     0  1  0  1  0  0  1  0  0  0  1  0  0  1  0  1
## colley     0  0  1  0  1  0  0  0  1  0  1  0  0  1  1  0
## dalmatien  0  1  0  0  1  0  0  1  0  0  1  0  0  1  1  0
## dobermann  0  0  1  0  1  0  0  0  1  0  0  1  1  0  0  1
## dogue_all  0  0  1  0  0  1  0  0  1  1  0  0  1  0  0  1
## epagn_bre  0  1  0  0  1  0  0  1  0  0  0  1  0  1  1  0
## epagn_fra  0  0  1  0  1  0  0  1  0  0  1  0  1  0  1  0
## fox_hound  0  0  1  0  1  0  0  0  1  1  0  0  1  0  0  1
## fox_terri  1  0  0  1  0  0  0  1  0  0  1  0  0  1  0  1
## grand_ble  0  0  1  0  1  0  0  1  0  1  0  0  1  0  0  1
## labrador   0  1  0  0  1  0  0  1  0  0  1  0  0  1  1  0
## levrier    0  0  1  0  1  0  0  0  1  1  0  0  1  0  1  0
## mastiff    0  0  1  0  0  1  1  0  0  1  0  0  1  0  0  1
## pekinois   1  0  0  1  0  0  1  0  0  1  0  0  0  1  1  0
## pointer    0  0  1  0  1  0  0  0  1  0  0  1  1  0  1  0
## saint_ber  0  0  1  0  0  1  1  0  0  0  1  0  1  0  0  1
## setter     0  0  1  0  1  0  0  0  1  0  1  0  1  0  1  0
## teckel     1  0  0  1  0  0  1  0  0  0  1  0  0  1  1  0
## terre_neu  0  0  1  0  0  1  1  0  0  0  1  0  1  0  1  0

apply(K,2,sum)

##  T-  T+  T++  P-  P+  P++  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
```

## Le tableau de Burt :

```

B <- t(K)%*%K
B
##      T-  T+  T++ P-  P+  P++ V-  V+  V++ I-  I+  I++ Af-  Af+  Ag-  Ag+
## T-    7  0   0  7  0   0  5  2   0  3  3   1   1   6   5   2
## T+    0  5   0  1  4   0  1  4   0  0  4   1   0   5   3   2
## T++   0  0  15  0 10   5  4  2   9  5  6   4  12   3   6   9
## P-    7  1   0  8  0   0  6  2   0  3  4   1   1   7   5   3
## P+    0  4  10  0 14   0  0  6   8  3  7   4   7   7   8   6
## P++   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

```

## 2 Le TDC : un tableau de contingence particulier

En ACM, on traite le TDC (tableau disjonctif complet)  $\mathbf{K}$  comme un tableau de contingence :

$$\mathbf{K} = \begin{array}{c|ccc|c} & 1 \dots & s & \dots m & \\ \hline 1 & & & & \\ \vdots & & \vdots & & \\ i & \dots & k_{is} & \dots & k_{i.} = p \\ \vdots & & \vdots & & \\ n & & & & \\ \hline & & k_{.s} = n_s & & k_{..} = np \end{array}$$

Dans ce cas particulier on a :

- les totaux en ligne sont constants et égaux au nombre  $p$  de variables.
- les totaux en colonne correspondent au nombre d'individus possédant la modalité  $s$ , noté  $n_s$ .
- le total général vaut ici  $np$ .

On déduit de ce tableau une matrice de fréquence  $\mathbf{F}$  particulière (associée à un TDC).

Matrice des fréquences  $\mathbf{F}$  :  $f_{is} = \frac{k_{is}}{np}$

On a donc :

- $f_{is} = \frac{1}{np}$  si l'individu  $i$  possède la modalité  $s$  et  $f_{is} = 0$  sinon.
- le poids des lignes (individus) est constant et vaut  $\frac{1}{n}$ .
- le poids des colonnes (modalités) vaut  $\frac{n_s}{np}$ . Il est d'autant plus grand que la modalité est fréquente.

$$\mathbf{F} = \begin{array}{c|cccc|c} & 1 & \dots & s & \dots & m & \\ \hline 1 & & & & & & \\ \vdots & & & & & & \\ i & \dots & & \frac{k_{is}}{np} & \dots & & \frac{1}{n} \\ \vdots & & & & & & \\ q & & & & & & \\ \hline & & & \frac{n_s}{np} & & & \\ \hline \end{array}$$

On notera donc :

- $\mathbf{r} = (\dots, \frac{1}{n}, \dots)^t \in \mathbb{R}^n$  le vecteur des poids des individus.
- $\mathbf{c} = (\dots, \frac{n_s}{np}, \dots)^t \in \mathbb{R}^m$  le vecteur des poids des modalités.
- $\mathbf{D}_r = \text{diag}(\mathbf{r})$
- $\mathbf{D}_c = \text{diag}(\mathbf{c})$

On déduit de ce tableau de fréquences une matrice de profil-lignes  $\mathbf{L}$  et une matrice des profil-colonnes  $\mathbf{C}$ .

Matrice des profil-lignes  $\mathbf{L}$  :  $\frac{f_{is}}{f_{i.}} = \frac{k_{is}}{p}$

$$\mathbf{L} = \begin{array}{c|cccc|c} & 1 & \dots & s & \dots & m & \\ \hline 1 & & & & & & \\ \vdots & & & & & & \\ i & \dots & & \frac{k_{is}}{p} & \dots & & \\ \vdots & & & & & & \\ 1 & & & & & & \\ \hline \mathbf{c}^t & \dots & & \frac{n_s}{np} & \dots & & \\ \hline \end{array}$$

On a :

- $\mathbf{L} = \mathbf{D}_r^{-1} \mathbf{F}$
- Profil ligne moyen :  $\mathbf{c}$

Matrice des profil-colonnes  $\mathbf{C}$  :  $\frac{f_{is}}{f_{.s}} = \frac{k_{is}}{n_s}$

$$\mathbf{C} = \begin{array}{c|cccc||c} & 1 & \dots & s & \dots & m & \mathbf{r} \\ \hline 1 & & & & & & \\ \vdots & & & & & & \\ i & \dots & & \frac{k_{is}}{n_s} & \dots & & \frac{1}{n} \\ \vdots & & & & & & \\ q & & & & & & \\ \hline & & & & & & \\ \hline \end{array}$$

On a :

- $\mathbf{C} = \mathbf{F}\mathbf{D}_c^{-1}$
- Profil colonne moyen :  $\mathbf{r}$

Deux nuages de points pondérés :

On considère dans la suite les deux nuages **centrés** de point suivants :

- Le nuage des  $n$  point-individus centrés de  $\mathbb{R}^m$  c'est à dire les  $n$  lignes de la matrice des profil-lignes centrés  $\mathbf{L} = \mathbf{D}_r^{-1}(\mathbf{F} - \mathbf{r}\mathbf{c}^t)$ . Chaque point-individu est pondéré par  $\frac{1}{n}$
- Le nuage des  $m$  point-modalités centrés de  $\mathbb{R}^n$  c'est à dire les  $m$  lignes de la matrice des profil-colonnes centrés  $\mathbf{C} = (\mathbf{F} - \mathbf{r}\mathbf{c}^t)\mathbf{D}_c^{-1}$ . Chaque point-modalité est pondéré par  $\frac{n_s}{np}$ .

## 2.1 Distance du $\chi^2$

En ACM on utilise la distance du  $\chi^2$  pour comparer deux individus décrits par deux points de  $\mathbb{R}^m$  (deux profil-lignes) ou deux modalités décrites par deux points de  $\mathbb{R}^n$  ((deux profil-colonnes). En ACP, les individus et les variables étaient les lignes et les colonnes d'une même matrice (la matrice des données centrées-réduites). En ACM, les individus et les modalités sont les lignes et les colonnes de deux matrices différentes (resp. la matrice des profil-lignes et la matrice des profil-colonnes). Pour comparer deux individus ou deux modalités, on utilise en ACM la distance du  $\chi^2$ .

Distance du  $\chi^2$  entre deux individus : métrique  $\mathbf{D}_c^{-1}$

$$d^2(i, i') = \sum_{s=1}^m \frac{1}{f_{.s}} \left( \frac{k_{is} - k_{i's}}{p} \right)^2 = \frac{n}{p} \sum_{s=1}^m \frac{1}{n_s} (k_{is} - k_{i's})^2$$

Donc deux individus sont proches s'ils possèdent les mêmes modalités, sachant que l'on donne plus de "poids" dans cette distance au fait que ces deux individus ont en commun une modalité rare ( $n_s$  petit).

Distance du  $\chi^2$  entre deux modalités : métrique  $\mathbf{D}_r^{-1}$

$$d^2(s, s') = \sum_{i=1}^n \frac{1}{f_{.i}} \left( \frac{k_{is} - k_{is'}}{n_s} - \frac{k_{is'}}{n_{s'}} \right)^2 = n \sum_{i=1}^n \left( \frac{k_{is}}{n_s} - \frac{k_{is'}}{n_{s'}} \right)^2$$

Donc deux modalités sont proches si elles sont possédées par les mêmes individus.

## 2.2 Inertie totale

En AFC, on pouvait interpréter statistiquement l'inertie des nuages de points (profil-lignes et colonnes) en terme de  $\chi^2/n$  mesurant l'indépendance entre les deux variables qualitatives. En ACM ce n'est plus le cas puisque l'on a :

$$I(L) = I(C) = \frac{m}{p} - 1, \quad (1)$$

On a donc l'inertie qui dépend de  $\frac{m}{p}$ , le nombre moyen de catégories par variable.

**Exercice 2** : Démontrer (1).

## 2.3 Remarques pratiques

On a :

$$d^2(s, \mathbf{r}) = \frac{n}{n_s} - 1$$

$\Rightarrow$  la distance d'une modalité au centre du nuage est d'autant plus grande que la modalité est rare ( $n_s$  petit).

$$f_{.s}d^2(s, \mathbf{r}) = \frac{1}{p}\left(1 - \frac{n_s}{n}\right)$$

$\Rightarrow$  la part de l'inertie totale, due à une modalité  $s$  est d'autant plus grande que la modalité est rare.

$\Rightarrow$  On évite en pratique de conserver dans l'analyse des modalités trop rare (avec des pré-traitement).

$$\sum_{s \in \mathcal{M}_j} f_{.s}d^2(s, \mathbf{r}) = \frac{1}{p}(m_j - 1)$$

$\Rightarrow$  la part de l'inertie totale, due à une variable  $j$  est d'autant plus grande que le nombre de modalités  $m_j$  de cette variable est grand.

$\Rightarrow$  On évite en pratique de conserver dans l'analyse des variables ayant des nombre de modalités trop différents (avec des pré-traitement).

## 3 AFC du tableau disjonctif complet

Effectuer une ACM consiste à appliquer l'AFC au TDC c'est à dire à effectuer une ACP pondérée des nuages des point-individus et des point-modalités (centrés). On reprend donc les résultats du cours d'AFC.

### 3.1 Coordonnées factorielles des individus et des modalités

Rappels d'AFC :

On pose  $\mathbf{R} = \mathbf{D}_r^{-1}(\mathbf{F} - \mathbf{rc}^t)\mathbf{D}_c^{-1}$  et on effectue l'ACP du triplet  $(\mathbf{R}, \mathbf{D}_r, \mathbf{D}_c)$  :

1. On effectue la décomposition en valeurs singulières de  $\mathbf{R}$  avec les métriques  $\mathbf{D}_r$  et  $\mathbf{D}_c$  :

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^t$$

où

- $\Lambda = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_r})$  est la matrice des valeurs singulières de  $\mathbf{R}\mathbf{D}_r\mathbf{R}^t\mathbf{D}_c$  et  $\mathbf{R}^t\mathbf{D}_r\mathbf{R}\mathbf{D}_c$  et  $r = \text{rang}(\mathbf{R})$ .
- $\mathbf{U}$  est la matrice de dimension  $n \times r$  dont les colonnes sont les vecteurs propres de  $\mathbf{R}\mathbf{D}_c\mathbf{R}^t\mathbf{D}_r$  et  $\mathbf{U}^t\mathbf{D}_r\mathbf{U} = \mathbb{I}_r$  (les vecteurs propres sont  $\mathbf{D}_r$ -orthonormés).
- $\mathbf{V}$  est la matrice de dimension  $p \times r$  dont les colonnes sont les vecteurs propres de  $\mathbf{R}^t\mathbf{D}_r\mathbf{R}\mathbf{D}_c$  et  $\mathbf{V}^t\mathbf{D}_c\mathbf{V} = \mathbb{I}_r$  (les vecteurs propres sont  $\mathbf{D}_c$ -orthonormés).

2. On retient  $k \leq r$  dimensions et on déduit de cette DVS :

- La matrice  $\mathbf{X}_k$  de dimension  $n \times k$  des coordonnées factorielles des  $n$  individus sur les  $k$  premiers axes :

$$\begin{aligned}\mathbf{X} &= \mathbf{R}\mathbf{D}_c\mathbf{V}_k \\ &= \mathbf{U}_k\Lambda_k\end{aligned}\tag{2}$$

- La matrice  $\mathbf{Y}_k$  de dimension  $m \times k$  des coordonnées factorielles des  $m$  modalités sur les  $k$  premiers axes :

$$\begin{aligned}\mathbf{Y} &= \mathbf{R}^t\mathbf{D}_r\mathbf{U}_k \\ &= \mathbf{V}_k\Lambda_k\end{aligned}\tag{3}$$

Application en ACM :

On en déduit que l'ACM est l'analyse du triplet suivant :

$$\left(n\mathbf{K}\mathbf{D}^{-1} - \mathbf{1}_{n \times m}, \frac{1}{n}\mathbb{I}_n, \frac{1}{np}\mathbf{D}\right)\tag{4}$$

où  $\mathbf{D} = \text{diag}(\dots, n_s, \dots)$  est la matrice diagonale des effectifs des modalités et  $\mathbf{1}_{n \times m}$  est la matrice de de dimension  $n \times m$  de terme général 1.

Remarques :

- On définit parfois l'ACM comme l'analyse du triplet suivant :

$$\left(n\mathbf{K}\mathbf{D}^{-1}, \frac{1}{n}\mathbb{I}_n, \frac{1}{np}\mathbf{D}\right)$$

Cela correspond au cas où l'on considère les nuages de points **non centrés**. Dans ce cas, on observe en plus, un vecteur propre trivial  $\mathbf{1}_n$  associé à la valeur propre 1. Le fait de centrer les nuage de points permet donc de l'éliminer.

- En ACM il y a au plus  $r = \min(n - 1, m - p)$  valeurs propre non nulle dans le cas centré.

**Exercice 3 :** Démontrer (4).

## 3.2 Relations barycentriques

Rappels d'AFC :

On les relations quasi-barycentriques suivantes :

$$\begin{cases} x_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{s=1}^m \frac{f_{is}}{f_{i.}} y_{s\alpha} \\ y_{s\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \sum_{i=1}^n \frac{f_{is}}{f_{.s}} x_{i\alpha} \end{cases}$$

En ACM :

On en déduit que dans le cas particulier où  $\mathbf{K}$  est un TDC on a :

$$\begin{cases} x_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{p} \sum_{s=1}^m k_{is} y_{s\alpha} \\ y_{s\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{n_s} \sum_{i=1}^n k_{is} x_{i\alpha} \end{cases}$$

ce qui s'écrit encore :

$$\begin{cases} x_{i\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{p} \sum_{s \in S_i} y_{s\alpha} \\ y_{s\alpha} = \frac{1}{\sqrt{\lambda_\alpha}} \frac{1}{n_s} \sum_{i \in I_s} x_{i\alpha} \end{cases} \quad (5)$$

où  $S_i$  est l'ensemble des modalités prises par l'individu  $i$  et  $I_s$  est l'ensemble des individus qui possèdent la modalité  $s$ .

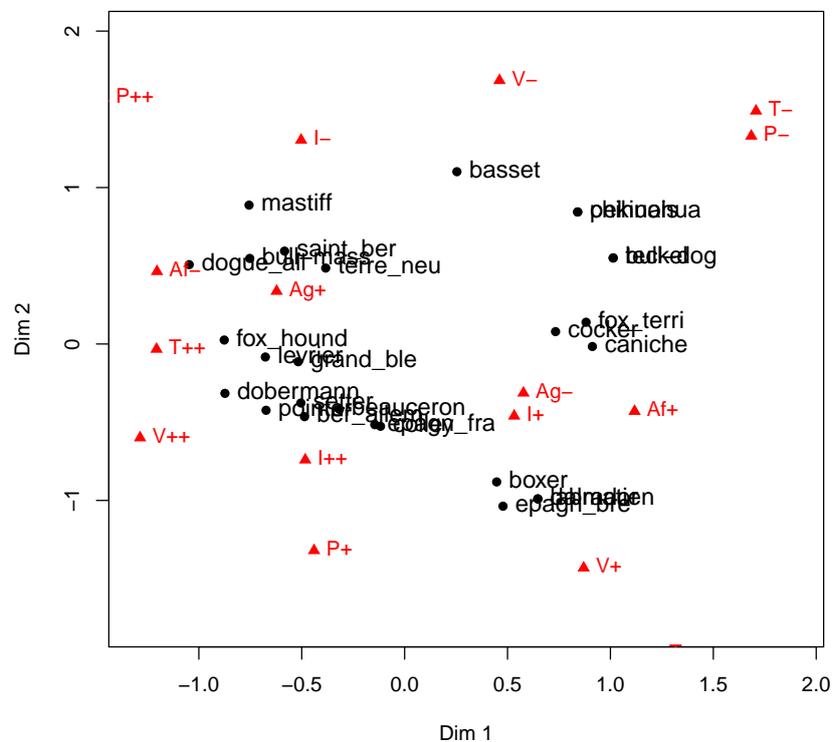
Ces relations quasi-barycentriques donnent trois modes de représentation simultanée des individus et des modalités :

1. Comme  $\text{card}(S_i) = p$ , la première relation dans (5) se lit : “La coordonnées factorielle de l'individu  $i$  sur l'axe  $\alpha$  est égale, à  $\frac{1}{\sqrt{\lambda_\alpha}}$  près, à la moyenne arithmétique simple des coordonnées des modalités qu'il possède”.  
 $\Rightarrow$  Première possibilité de représentation simultanée : les individus au barycentre des modalités.
2. Comme  $\text{card}(I_s) = n_s$ , la seconde relation dans (5) se lit : “La coordonnées factorielle de la modalité  $s$  sur l'axe  $\alpha$  est égale, à  $\frac{1}{\sqrt{\lambda_\alpha}}$  près, à la moyenne arithmétique simple des coordonnées des individus qui la possèdent”.  
 $\Rightarrow$  Seconde possibilité de représentation simultanée : les modalités au barycentre des individus.
3. Une troisième représentation simultanée des individus et des modalités consiste à représenter sur un même graphique les moyennes arithmétiques dilatées par  $\frac{1}{\sqrt{\lambda_\alpha}}$ . Sur ces graphiques, à  $\frac{1}{\sqrt{\lambda_\alpha}}$  près :
  - les individus sont au centre des modalités qu'ils ont choisis,
  - les modalités sont au centre des individus qui les ont choisis.

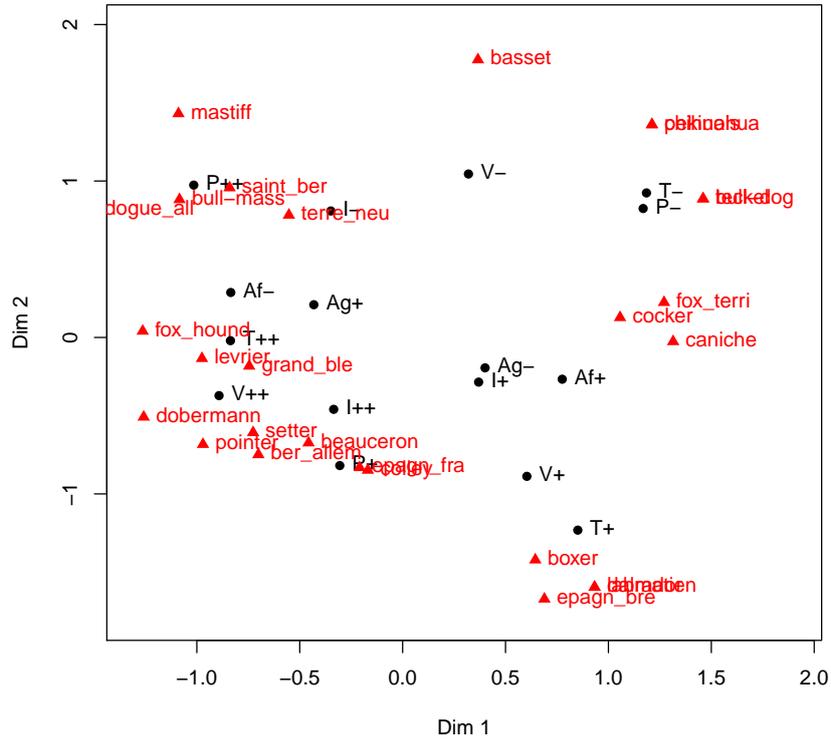
## Exemple.

```
res <- MCA(H,ncp=2,graph=FALSE)
X <- res$ind$coord[,1:2] #coord. des chiens sur le 1er plan fact.
Y <- res$var$coord[,1:2] #coord. des modalites sur le 1er plan fact.
d <- sqrt(res$eig[1:2,1]) #ecart-types des 2 premieres composantes
Xstand <- X %%% diag(1/d) #coordonnees factorielles standardisees
Ystand <- Y %%% diag(1/d)
xlim <- c(min(c(min(Xstand[,1]),min(Ystand[,1])),max(c(max(Xstand[,1]),max(Ystand[,1]))))+0.2
ylim <- c(min(c(min(Xstand[,2]),min(Ystand[,2])),max(c(max(Xstand[,2]),max(Ystand[,2]))))+0.2
plot(X,pch=16,xlim=xlim,ylim=ylim,main="individus au barycentre des modalites")
text(X,rownames(X),pos=4,cex=1.2)
points(Ystand,pch=17,col=2)
text(Ystand,rownames(Y),pos=4,col=2)
```

individus au barycentre des modalites

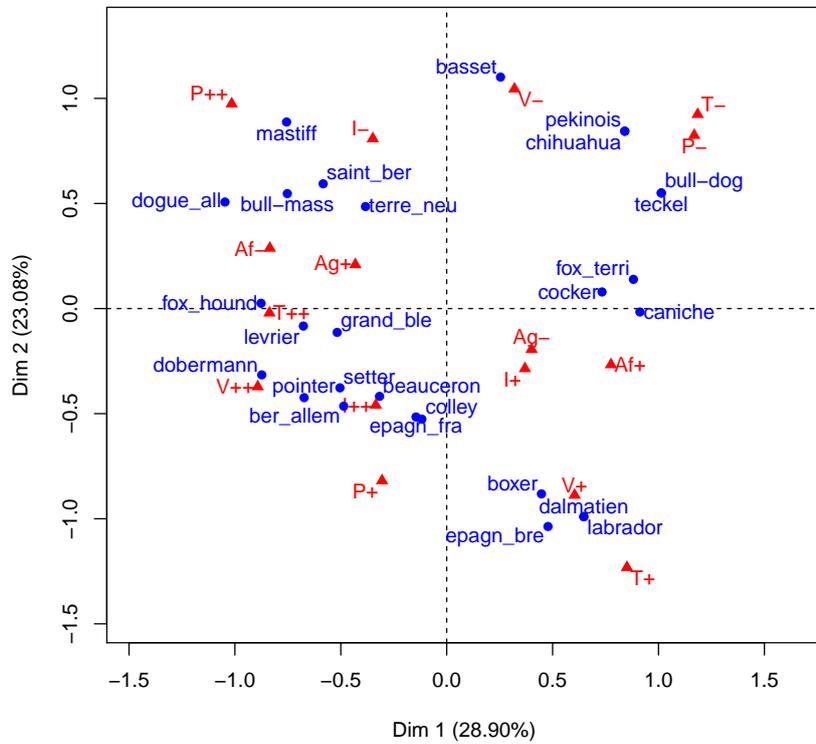


### modalites au barycentre des individus



```
plot(res,title="representation simultanee")
```

### representation simultanee



## 4 Règles d'interprétation

Les individus et les modalités sont représentés sur ses plans de projection dont la lecture nécessite des règles d'interprétation.

### 4.1 Inertie expliquée

On a vu qu'en ACM, l'inertie totale du nuage des individus et du nuage des modalités vaut  $\frac{m}{p} - 1$  et ne dépend donc que du nombre moyen de modalités par variable. De plus l'inertie totale est égale à  $\lambda_1 + \dots + \lambda_r$  où  $r = \min(n - 1, m - p)$  est le nombre de valeurs propre non nulles. Le pourcentage d'inertie expliquée par un axe  $\alpha$  est donc :

$$\frac{\lambda_\alpha}{\lambda_1 + \dots + \lambda_r} * 100$$

```
res$eig

##          eigenvalue percentage of variance cumulative percentage of variance
## dim 1    0.481606                28.89637                28.896
## dim 2    0.384737                23.08424                51.981
## dim 3    0.210954                12.65724                64.638
## dim 4    0.157554                 9.45324                74.091
## dim 5    0.150133                 9.00796                83.099
## dim 6    0.123295                 7.39772                90.497
## dim 7    0.081462                 4.88775                95.385
## dim 8    0.045670                 2.74019                98.125
## dim 9    0.023542                 1.41251                99.537
## dim 10   0.007713                 0.46278                100.000

#nombre de modalités
m <- sum(unlist(lapply(H,function(x){length(levels(x))})))
m

## [1] 16

p <- ncol(H)
p

## [1] 6

#nombre de chiens
n <- nrow(H)
n

## [1] 27

#nombre de valeurs propres non nulles
min(n-1,m-p)

## [1] 10

#inertie totale
sum(res$eig[,1])

## [1] 1.6667

#nombre moyen de modalités par variable
m/ncol(H)-1

## [1] 1.6667
```

Attention : en ACM, les pourcentages d’inertie expliquées par les axes sont **par construction “petits”** (cf. section 5) et ne peuvent donc pas être interprétés comme en AFC ou en ACP. Le nombre d’axes retenus pour l’interprétation ou le recodage ne peut pas être choisi à partir de ces pourcentages d’inertie expliquées.

## 4.2 Contributions

On reprend les formules de l’AFC et on trouve qu’en ACM la contribution d’un individu  $i$  et la contribution d’une modalité  $s$  à l’inertie de l’axe  $\alpha$  s’écrivent :

$$\begin{cases} \text{Ctr}_\alpha(i) = \frac{1}{n} \frac{x_{i\alpha}^2}{\lambda_\alpha} \\ \text{Ctr}_\alpha(s) = \frac{n_s}{np} \frac{y_{s\alpha}}{\lambda_\alpha} \end{cases}$$

On en déduit qu’en pratique :

- Les individus les plus excentrés sur les plans factoriels sont ceux qui contribuent le plus.
- En revanche, les modalités les plus excentrées ne sont pas nécessairement celles qui contribuent le plus. En effet, leur contribution dépend de leur fréquence.

En ACM, la **contribution (absolue) d’une variable  $j$**  à de l’axe  $\alpha$  est définie comme la somme des contributions de ses modalités :

$$\text{Ctr}_\alpha(j) = \sum_{s \in \mathcal{M}_j} \text{Ctr}_\alpha(s) = \frac{1}{p} \eta_{\mathbf{x}^\alpha | j}^2 \quad (6)$$

où  $\eta_{\mathbf{x}^\alpha | j}^2$  est le rapport de corrélation entre la composante  $\mathbf{x}^\alpha$  (quantitative) et la variable  $j$  (qualitative). Le rapport de corrélation mesure la part de la variance de  $\mathbf{x}^\alpha$  expliquée par la variable qualitative  $j$  :

$$\eta_{\mathbf{x}^\alpha | j}^2 = \frac{\sum_{s \in \mathcal{M}_j} n_s (\bar{\mathbf{x}}_s^\alpha - \bar{\mathbf{x}}^\alpha)^2}{\sum_{i=1}^n (x_{i\alpha} - \bar{\mathbf{x}}^\alpha)^2}$$

où  $\bar{\mathbf{x}}_s^\alpha$  est la moyenne de  $\mathbf{x}^\alpha$  calculé avec les individus qui possèdent la modalité  $s$ .

En pratique :

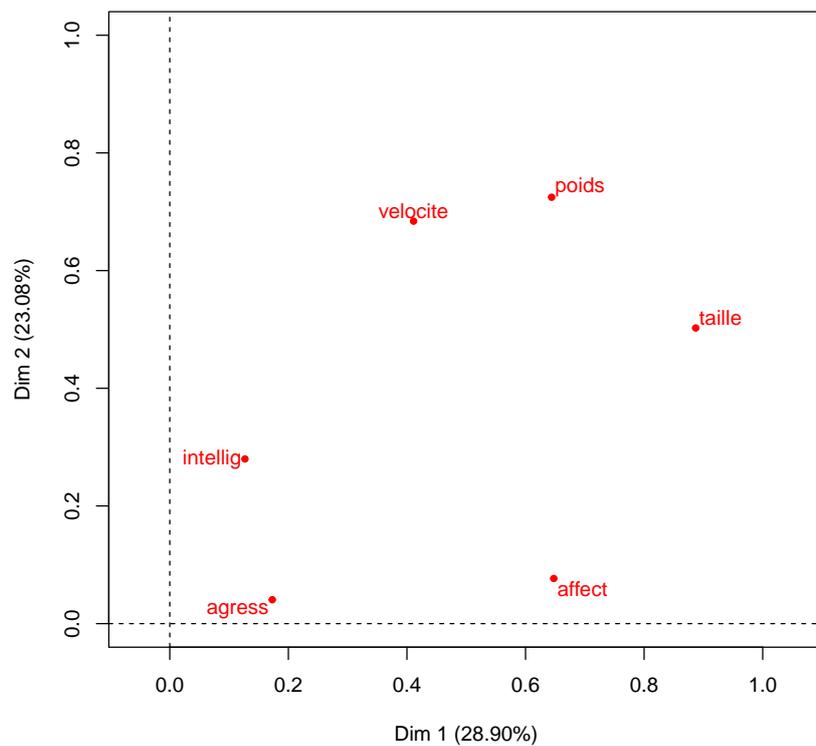
- La contribution d’une variable qualitative  $j$  à un axe  $\alpha$  donne une idée de la liaison entre cette variable et la composante principale  $\mathbf{x}^\alpha$ .
- On utilise ces contributions pour représenter graphiquement les variables qualitatives de l’analyse sur un plan factoriel  $(\alpha, \alpha')$ . On représente en abscisse les contributions des variables à l’axe  $\alpha$  et en ordonnée les contributions des variables à l’axe  $\alpha'$ . On dessine alors des “flèches” comme dans le cercle des corrélations en ACP.

### Exemple.

```
#contributions relatives
contrib <- res$var$contrib
contrib
```

```
##      Dim 1      Dim 2
## T-  12.59781  9.586617
## T+   4.64207 12.170670
## T++ 13.45855  0.010191
## P-  14.01042  8.722246
## P+   1.67369 15.062072
## P++  6.60404  7.608867
## V-   1.31199 17.517423
## V+   3.73685 10.117058
## V++  9.18042  1.996447
## I-   1.24924  8.391308
## I+   2.27421  1.700144
## I++  0.86338  2.032408
## Af- 11.62158  1.723646
## Af+ 10.79147  1.600529
## Ag-  2.88132  0.847587
## Ag+  3.10296  0.912786
```

```
plot(res, choix="var")
```



```
#contributions absolues
contrib_abs <- sweep(contrib,2,res$eig[1:2,1],"*")/100
#rapports de correlations
apply(contrib_abs[1:3,],2,sum)*p

##      Dim 1      Dim 2
## 0.88707  0.50249
```

### 4.3 Cosinus carrés

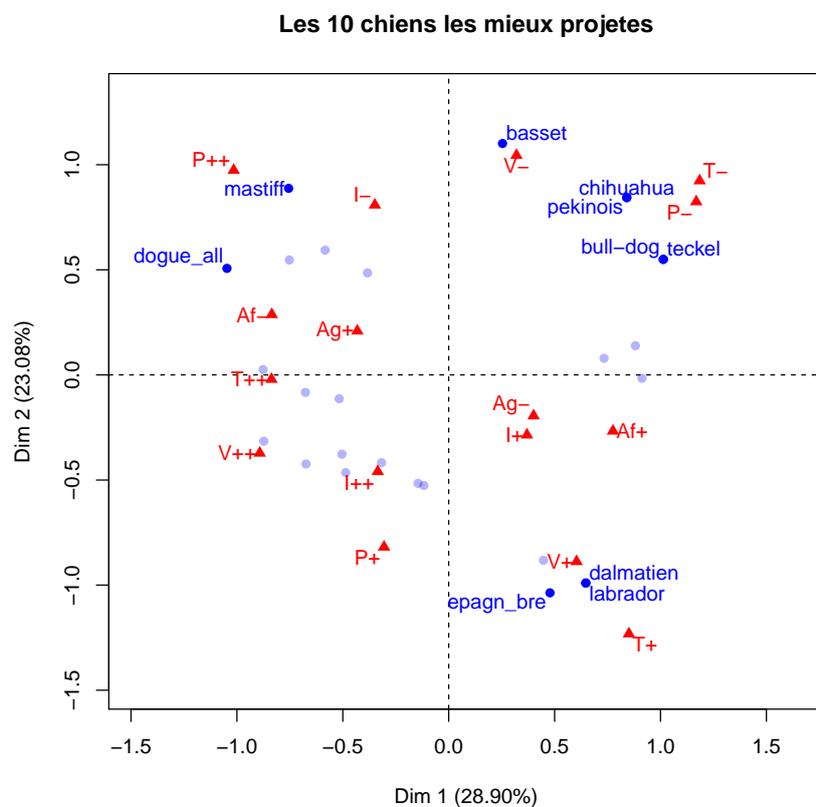
On utilise les mêmes formules de  $\cos_\alpha^2(i)$  et  $\cos_\alpha^2(s)$  pour mesurer la qualité de la projection des individus et des modalités sur les axes factoriels.

En pratique, si deux individus **sont bien projetés** alors s'ils sont proches en projections, ils sont effectivement proches dans leur espace d'origine et on peut alors interpréter leur proximité :

- La proximité entre deux individus s'interprète en terme de distance (du  $\chi^2$ ) : deux individus se ressemblent s'ils ont choisis les mêmes modalités. C'est cohérent avec la relation barycentrique qui dit que les individus sont au barycentre des modalités qu'ils possèdent.
- La proximité entre deux modalités de **deux variables différentes** s'interprète en terme de distance (du  $\chi^2$ ) : deux modalités se ressemblent si elles sont possédées par les mêmes individus. C'est cohérent avec la relation barycentrique qui dit que les modalités sont au barycentre des individus qui les possèdent.

**Exemple.**

```
#contributions relatives  
plot(res,select="cos2 10",title="Les 10 chiens les mieux projetes")
```



## 5 Cas particulier où $p = 2$ : lien entre l'AFC et l'ACM

On se place dans le cas particulier où  $p = 2$  c'est à dire le cas où l'on observe  $n$  individus sur deux variables qualitatives ayant respectivement  $m_1$  et  $m_2$  modalités.

On peut alors faire l'analyse factorielle d'un des deux tableaux suivant :

- le tableau disjonctif complet  $\mathbf{K}$  de dimension  $n \times (m_1 + m_2)$  comme en ACM
- le tableau de contingence  $\mathbf{N}$  de dimension  $m_1 \times m_2$  croisant les deux variables qualitatives comme en AFC

On a la relation suivante entre les premières valeurs propres des deux analyses :

$$\mu_\alpha = (2\lambda_\alpha - 1)^2$$

où  $\mu_\alpha$  est la  $\alpha$ -ème valeur propre de l'analyse de  $\mathbf{N}$  (l'AFC) et  $\lambda_\alpha$  est la  $\alpha$ -ème valeur propre de l'analyse de  $\mathbf{K}$  (l'ACM). On en déduit qu'à chaque valeur propre  $\mu_\alpha$  de l'AFC correspondent deux valeurs propres de l'ACM :

$$\begin{cases} \lambda_\alpha = \frac{1 + \sqrt{\mu_\alpha}}{2} \\ \lambda_* = \frac{1 - \sqrt{\mu_\alpha}}{2} \end{cases}$$

On a également la relation suivante entre les coordonnées factorielles des deux analyses :

$$\begin{aligned} \mathbf{y}^\alpha &= \begin{pmatrix} \mathbf{x}_N^\alpha \\ \mathbf{y}_N^\alpha \end{pmatrix} && \text{pour } \lambda_\alpha = \frac{1 + \sqrt{\mu_\alpha}}{2} \\ \mathbf{y}_* &= \begin{pmatrix} \mathbf{x}_N^\alpha \\ -\mathbf{y}_N^\alpha \end{pmatrix} && \text{pour } \lambda_* = \frac{1 - \sqrt{\mu_\alpha}}{2} \end{aligned}$$

et où  $\mathbf{x}_N^\alpha$  et  $\mathbf{y}_N$  sont les composantes principales des profil-lignes et colonnes de  $\mathbf{N}$ .

Les conséquences pratiques de ces résultats sont :

- Dans l'ACM de  $p = 2$  variables, on ne retiendra que les valeurs propres  $\lambda_\alpha > 1/2$  qui correspondent aux composantes  $\mathbf{y}^\alpha = \begin{pmatrix} \mathbf{x}_N^\alpha \\ \mathbf{y}_N^\alpha \end{pmatrix}$ . En effet, les composantes qui correspondent aux valeurs propres  $\lambda_*$  sont liées aux précédentes par la relation ci-dessus.
- Les pourcentages d'inertie expliqués par les axes en ACM sont souvent très faibles et ne peuvent donc pas être interprétés comme en AFC et en ACP.

Exemple du livre de G. Saporta : on croise deux variables avec  $m_1 = 9$  et  $m_2 = 8$

ACM	AFC
$\lambda_1=0.628$ (8 %)	$\mu_1 = 0.0657$ (61.24%)
$\lambda_2=0.580$ (7.73 %)	$\mu_2 = 0.0254$ (23.7%)
$\lambda_3=0.545$ (7.27 %)	$\mu_3 = 0.0081$ (7.55%)
$\sum_{i=1}^{15} \lambda_i = 7.5$	$\sum_{i=1}^7 \mu_i = 0.1073$

## 6 AFC du tableau de Burt

L'ACM peut être définie comme l'AFC :

- du tableau disjonctif complet  $\mathbf{K}$ . C'est la présentation "à la française".
- du tableau de Burt  $\mathbf{B} = \mathbf{K}^t\mathbf{K}$ . C'est la présentation "à l'anglo-saxonne".

Différence entre les deux analyses :

Le tableau de Burt est un assemblage particulier des tableaux de contingence qui sont les faces de l'hypercube de contingence. Il s'agit donc d'une matrice symétrique de dimension  $m \times m$ . Puisque cette matrice est symétrique, les profil-lignes et les profils-colonnes sont identiques et correspondent aux point-modalités que l'on veut analyser. On ne pourra donc pas effectuer d'analyse des points-individus ce qui est la principale différence avec l'analyse à partir du TDC.

Lien entre les deux analyses :

On peut montrer que :

$$\begin{cases} \mathbf{y}_{\mathbf{B}}^\alpha = \sqrt{\lambda_\alpha} \mathbf{y}^\alpha \\ \mu_\alpha = \lambda_\alpha^2 \end{cases}$$

où  $\mathbf{y}_{\mathbf{B}}^\alpha$  est le vecteur des coordonnées factorielles des points-modalité sur l'axe  $\alpha$  de l'analyse du tableau de Burt et  $\mu_\alpha = \text{Var}(\mathbf{y}_{\mathbf{B}}^\alpha)$  est la  $\alpha$ -ème valeur propre de cette analyse.

Cela veut dire que les coordonnées factorielles des points-modalités de l'analyse du tableau de Burt sont les mêmes que celles de l'analyse du TDC à  $\sqrt{\lambda_\alpha}$  près. Cela veut dire également qu'on peut obtenir les valeurs propres de l'analyse du tableau de Burt en élevant au carré les premières valeurs propres de l'analyse du TDC.

## 7 Références

- "Statistique exploratoire multidimensionnelle", Lebart & al., Dunod.
- "Probabilité, analyse des données, statistique", G. Saporta, Technip.