

## Chapitre 6

# Analyse des Correspondances Multiples

### 6.1 Introduction

L'analyse factorielle des correspondances multiples (ACM ou AFCM) est la généralisation de l'analyse des correspondances multiples au cas de plusieurs variables. Elle consiste donc à représenter les modalités de variables qualitatives dans un espace euclidien dans lequel les distances du  $\chi^2$  entre deux modalités d'une même variable sont préservées au mieux. On considère donc dans cette section  $p$  variables qualitatives observées simultanément sur  $n$  individus de poids identiques  $1/n$ .

Exemple - Considérons le jeu de données de la Table 6.1 dans lequel on caractérise différentes races de chien en fonction de 7 variables portant sur des caractéristiques de physique, sur des points de caractère et une variable d'utilité.

La plupart des tableaux et figures liées à cet exemple sont empruntées à M. Tenenhaus.

### 6.2 Définitions et notations

#### 6.2.1 Tableau disjonctif complet

Il est difficile de travailler directement avec un tableau de données comme celui de l'utilité des races de chien. En effet, on ne peut pas considérer ces données comme des données quantitatives. Par exemple, ça n'a pas de sens de considérer qu'il y a une distance équivalente entre les classes - et + de la variable Poids et de la variable Intelligence. En conséquence, il est d'usage de recoder les données et de construire le *tableau disjonctif complet*.

Le *tableau disjonctif complet* est tel que chaque ligne correspond à un individu et chaque colonne à une modalité. On note  $K$  le nombre total de modalités. Et les observations  $x_{ij}$  sont codées 1 si l'individu  $i$  a la modalité  $j$  et 0 sinon. Notons  $X$  le tableau disjonctif complet.

Dans l'exemple, on obtient alors le tableau de la Table 6.2.

	<b>Race</b>	<b>Taille</b>	<b>Poids</b>	<b>Vitesse</b>	<b>Intell.</b>	<b>Affect.</b>	<b>Agress.</b>	<b>Fonction</b>
1	Beauceron	TA++	PO+	V++	INT+	AF+	AG+	Utilité
2	Basset	TA-	PO-	V-	INT-	AF-	AG+	Chasse
3	Berger-Allemand	TA++	PO+	V++	INT++	AF+	AG+	Utilité
4	Boxer	TA+	PO+	V+	INT+	AF+	AG+	Compagnie
5	Bull-Dog	TA-	PO-	V-	INT+	AF+	AG-	Compagnie
6	Bull-Mastiff	TA++	PO++	V-	INT++	AF-	AG+	Utilité
7	Caniche	TA-	PO-	V+	INT++	AF+	AG-	Compagnie
8	Chihuahua	TA-	PO-	V-	INT-	AF+	AG-	Compagnie
9	Cocker	TA+	PO-	V-	INT+	AF+	AG+	Compagnie
10	Colley	TA++	PO+	V++	INT+	AF+	AG-	Compagnie
11	Dalmatien	TA+	PO+	V+	INT+	AF+	AG-	Compagnie
12	Doberman	TA++	PO+	V++	INT++	AF-	AG+	Utilité
13	Dogue Allemand	TA++	PO++	V++	INT-	AF-	AG+	Utilité
14	Epagneul Breton	TA+	PO+	V+	INT++	AF+	AG-	Chasse
15	Epagneul Français	TA++	PO+	V+	INT+	AF-	AG-	Chasse
16	Fox-Hound	TA++	PO+	V++	INT-	AF-	AG+	Chasse
17	Fox-Terrier	TA-	PO-	V+	INT+	AF+	AG+	Compagnie
18	Grd Bleu de Gascogne	TA++	PO+	V+	INT-	AF-	AG+	Chasse
19	Labrador	TA+	PO+	V+	INT+	AF+	AG-	Chasse
20	Lévrier	TA++	PO+	V++	INT-	AF-	AG-	Chasse
21	Mastiff	TA++	PO++	V-	INT-	AF-	AG+	Utilité
22	Pékinois	TA-	PO-	V-	INT-	AF+	AG-	Compagnie
23	Pointer	TA++	PO+	V++	INT++	AF-	AG-	Chasse
24	Saint-Bernard	TA++	PO++	V-	INT+	AF-	AG+	Utilité
25	Setter	TA++	PO+	V++	INT+	AF-	AG-	Chasse
26	Teckel	TA-	PO-	V-	INT+	AF+	AG-	Compagnie
27	Terre-Neuve	TA++	PO++	V-	INT+	AF-	AG-	Utilité

TABLE 6.1 – Caractéristiques (physique, caractère, utilité) de différentes races de chien

Race	T-	T+	T++	P-	P+	P++	V-	V+	V++	I-	I+	I++	Af-	Af+	Ag-
Beauceron	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0
Basset	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0
Berger all	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0
Bower	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0
Bull-dog	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1
Bull Mastiff	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0
Caniche	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1
Chihuahua	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1
Cocker	0	1	0	1	0	0	1	0	0	0	1	0	0	1	0
Colley	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1
Dalmatien	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1
Doberman	0	0	1	0	1	0	0	0	1	0	0	1	1	0	0
Dogue all	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0
Epagneul br	0	1	0	0	1	0	0	1	0	0	0	1	0	1	1
Epagneul fr	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1
Fox-Hound	0	0	1	0	1	0	0	0	1	1	0	0	1	0	0
Fox-Terrier	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0
Grd Bl de G	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0
Labrador	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1
Lévrier	0	0	1	0	1	0	0	0	1	1	0	0	1	0	1
Mastiff	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0
Pékinois	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1
Pointer	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1
St-Bernard	0	0	1	0	0	1	1	0	0	0	1	0	1	0	0
Setter	0	0	1	0	1	0	0	0	1	0	1	0	1	0	1
Teckel	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1
Terre neuve	0	0	1	0	0	1	1	0	0	0	1	0	1	0	1

TABLE 6.2 – tableau disjonctif complet des caractéristiques (physique, caractère, utilité) de différentes races de chien.

## 6.2.2 Tableau de Burt

On appelle tableau de Burt le tableau  $\mathcal{B} = X^T X$ . On peut écrire  $\mathcal{B} = (B_{k,k'})_{k,k'=1,\dots,p}$  où

- $p$  est nombre total de variables
- si  $k \neq k'$ ,  $B_{k,k'}$  est la table de contingence des variables  $X_k$  et  $X_{k'}$ ,
- si  $k = k'$ ,  $B_{kk}$  est une matrice diagonale contenant les effectifs marginaux de  $X_k$  dans la diagonale, notés  $n_{c_1}^k, \dots, n_{c_k}^k$ .

Propriétés :

- $\mathcal{B}$  est symétrique.
- La somme des lignes (resp. des colonnes) de  $\mathcal{B}$  est  $pn_l^k, l = c_1, \dots, c_k$ .
- La somme des éléments de  $\mathcal{B}$  est  $p^2 n$ .

Remarque : si on considère les données du tableau disjonctif  $X$  comme des observations de variables qualitatives, alors le tableau de Burt représente la variance de  $X$  à un facteur multiplicatif près.

Dans l'exemple des chiens, le tableau de Burt prend la forme suivante. On observe que la diagonale représente les profils (ou distribution en effectif) des différentes variables tandis que les termes extra diagonaux donnent les effectifs croisés entre deux modalités.

TABLEAU DE BURTT

	TA-	TA+	TA0++	PO-	PO+	PO0++	VE-	VE+	VE0++	INT-	INT+	INT0++	AF-	AF+	AG-	AG+	Comp	Chas	Util
TA-	7	0	0																
TA+	0	5	0																
TA0++	0	0	15																
PO-	7	1	0	8	0	0													
PO+	0	4	10	0	14	0													
PO0++	0	0	5	0	0	5													
VE-	5	1	4	6	0	4	10	0	0										
VE+	2	4	2	2	6	0	0	8	0										
VE0++	0	0	9	0	8	1	0	0	9										
INT-	3	0	5	3	3	2	4	1	3	8	0	0							
INT+	3	4	6	4	7	2	5	5	3	0	13	0							
INT0++	1	1	4	1	4	1	1	2	3	0	0	6							
AF-	1	0	12	1	7	5	5	2	6	6	4	3	13	0					
AF+	6	5	3	7	7	0	5	6	3	2	9	3	0	14					
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			
Comp	6	3	1	7	3	0	5	4	1	2	7	1	0	10	7	3	10	0	0
Chas	1	2	6	1	8	0	1	4	4	4	3	2	7	2	6	3	0	9	0
Util	0	0	8	0	3	5	4	0	4	2	3	3	6	2	1	7	0	0	8

## 6.2.3 Tableau des $\chi^2$

Avant d'aller plus loin et pour aider à l'interprétation des résultats qu'on obtiendra par la suite, il est utile de générer aussi le tableau des statistiques du  $\chi^2$  entre les différentes variables. Ce

tableau garde un sens, en effet la distance entre deux modalités  $j$  et  $j'$

$$d^2(j, j') = \sum_{i=1}^n n \left( \frac{x_{ij}}{x_{\bullet j}} - \frac{x_{ij'}}{x_{\bullet j'}} \right)^2$$

Ainsi deux modalités choisies par les même individus coïncide. Par ailleurs, les modalités de faible effectif sont éloignées des autres.

La distance entre deux individus  $i$  et  $i'$  s'exprime

$$d^2(i, i') = \frac{1}{p} \sum_{j=1}^K \frac{n}{x_{\bullet j}} (x_{ij} - x_{i'j})^2$$

Deux individus sont proches s'ils ont répondu de la même manière.

Dans le tableau, les chiffres entre parenthèses représentent les degrés de significativité (p-value) du test du  $\chi^2$ . On remarque par exemple que la taille et le poids sont liés à la vélocité tandis que seule l'agressivité est lié à la fonction.

	<b>Poids</b>	<b>Vélocité</b>	<b>Intelligence</b>	<b>Affection</b>	<b>Agressivité</b>	<b>Fonction</b>
<b>Taille</b>	25.3 (.000)	15.9 (.000)	3.6 (.46)	14.0 (.001)	2.1 (.36)	16.35 (.003)
<b>Poids</b>		18.4 (.001)	1.35 (.85)	9.5 (.008)	2.6 (.28)	24.41 (.000)
<b>Vélocité</b>			3.16 (.53)	3.0 (.23)	.57 (.75)	8.49 (.08)
<b>Intelligence</b>				3.9 (.14)	1.15 (.56)	4.14 (.39)
<b>Affection</b>					1.8 (.18)	14.76 (.000)
<b>Agressivité</b>						7.07 (.03)

## 6.3 Analyse Factorielle des Correspondances Multiples

L'Analyse Factorielle des Correspondances Multiples des variables  $x_1, \dots, x_p$  est l'analyse factorielle des correspondances du tableau disjonctif complet ou du tableau de Burt.

On rappelle les notations définies plus haut.

- $n$  est le nombre d'individus.
- On a  $p$  variables qualitatives.
- La variable  $X_j$  admet  $n_{c_j}^j$  modalités.
- $K = n_{c_1}^1 + \dots + n_{c_j}^j$  est le nombre total de modalités.
- La modalité  $jl$  a une fréquence absolue  $n_{jl} = n_l^j$  et une fréquence relative  $\frac{n_{jl}}{np}$  dans le tableau de Burt.

### 6.3.1 AFC du tableau disjonctif complet relatif à 2 variables

On note toujours  $X_1$  et  $X_2$  les 2 variables qualitatives et on note  $r$  et  $c$  leur nombre respectif de modalités. Les matrices intervenant dans l'AFC usuelle sont reprises ici selon les mêmes notations

que dans le chapitre précédent mais surlignées. Ici,  $D_L$  (resp.  $D_C$ ) est la matrice diagonale qui contient les profils lignes (resp. colonnes) en fréquence.

$$\begin{aligned}\bar{N} &= X = [X_1|X_2] \\ \bar{D}_L &= \frac{1}{n}\mathbb{I}_n \\ \bar{D}_C &= \frac{1}{2} \begin{bmatrix} D_L & 0 \\ 0 & D_C \end{bmatrix} = \frac{1}{2}\Delta \\ \bar{A} &= \frac{1}{2n}\bar{N}^T\bar{D}_L^{-1} = \frac{1}{2}X^T, \text{ avec } N \text{ la table de contingence.} \\ \bar{B} &= \frac{1}{2n}\bar{N}\bar{D}_C^{-1} = \frac{1}{2}X\Delta^{-1}\end{aligned}$$

L'AFC est considérée comme une double ACP : celle des profils lignes de  $\bar{A}$  puis celle des profils colonne de  $\bar{B}$ .

**Proposition 6** - ACP des profils lignes

L'ACP des profils lignes issue de l'AFC réalisée sur le tableau disjonctif complet relatif à 2 variables qualitatives conduit à l'analyse spectrale de la matrice  $\bar{D}_C^{-1}$ -symétrique et positive :

$$\bar{A}\bar{B} = \frac{1}{2} \begin{bmatrix} \mathbb{I}_r & B \\ A & \mathbb{I}_c \end{bmatrix}.$$

Les  $r + c$  valeurs propres de  $\bar{A}\bar{B}$  s'écrivent

$$\mu_k = \frac{1 \pm \sqrt{\lambda_k}}{2}$$

où les  $|\lambda_k$  sont les valeurs propres de la matrice  $AB$  (celle de l'AFC classique de  $X$ ). Les vecteurs propres  $\bar{D}_C^{-1}$ -orthonormés associés peuvent se mettre sous la forme

$$\bar{V} = \frac{1}{2} \begin{bmatrix} U \\ V \end{bmatrix}$$

où  $U$  et  $V$  sont les matrices de vecteurs propres obtenues en faisant l'AFC de la table de contingence associée à  $X_1$  et  $X_2$ .

La matrice des composantes principales s'écrit

$$\bar{C}_L = \frac{1}{2}[X_1C_L + X_2C_C]\Lambda^{-1/2}$$

où  $C_L$  et  $C_C$  sont les matrices de l'AFC classique.

Dans la pratique on ne considère que les  $d = \inf(r - 1, c - 1)$  plus grandes valeurs propres différentes de 1.

$$M = \text{diag}(\mu_1, \dots, \mu_d) = \frac{1}{2}[\mathbb{I}_d + \Lambda^{1/2}]$$

Les autres valeurs propres non nulles sont des artéfacts liés à la construction de la matrice à diagonaliser. Elles n'ont donc pas de sens statistique.

**Proposition 7** - ACP des profils colonnes

L'ACP des profils colonnes issue de l'AFC réalisée sur le tableau disjonctif complet relatif à 2 variables qualitatives conduit à l'analyse spectrale de la matrice  $\bar{D}_L^{-1}$ -symétrique et positive :

$$\bar{B}\bar{A} = \frac{1}{2n} [X_1 D_L^{-1} X_1^T + X_2 D_C^{-1} X_2^T]$$

Les  $r + c$  valeurs propres non nulles de  $\bar{B}\bar{A}$  sont les  $\mu_k$ . Les vecteurs propres  $\bar{D}_L^{-1}$ -orthonormés associés peuvent se mettre sous la forme

$$\bar{U} = \frac{1}{n} \bar{C}_L M^{-1/2}.$$

La matrice des composantes principales s'écrit

$$\bar{C}_C = \begin{bmatrix} C_L \\ C_C \end{bmatrix} \Lambda^{-1/2} M^{1/2}.$$

L'AFC du tableau disjonctif complet permet, grâce aux coordonnées contenues dans  $\bar{C}_C$ , la représentation simultanée des modalités des deux variables. Cette représentation est très proche de celle de l'AFC classique. De plus cette approche permet une représentation des individus avec les coordonnées de la matrice  $\bar{C}_L$ . A un facteur près, l'individu apparait comme le barycentre des deux modalités qu'il a présentées.

**6.3.2 AFC du tableau disjonctif complet**

Comme dans le cas où  $p = 2$ , on reprend les notations de l'AFC classique en les surlignant

$$\begin{aligned} \bar{T} &= X = [X_1 | \dots | X_p] \\ \bar{D}_L &= \frac{1}{n} \mathbb{I}_n \\ \bar{D}_C &= \frac{1}{p} \Delta \\ \bar{A} &= \frac{1}{p} X^T \\ \bar{B} &= \frac{1}{n} X \Delta^{-1} \end{aligned}$$

**Proposition 8** - ACP des profils lignes

L'ACP des profils lignes issue de l'AFC réalisée sur le tableau disjonctif complet relatif à  $p$  variables qualitatives conduit à l'analyse spectrale de la matrice  $\bar{D}_C^{-1}$ -symétrique et positive :

$$\bar{A}\bar{B} = \frac{1}{np} \mathcal{B} \Delta^{-1}$$

Il y a  $m$  ( $m \leq c - p$ ) valeurs propres notées  $\mu_k$  comprises entre 0 et 1 rangées dans la matrice diagonale  $M$ . La matrice des vecteurs propres  $\bar{D}_C^{-1}$ -orthonormés associés se décompose par blocs de la façon suivante

$$\bar{V} = \begin{bmatrix} V_1 \\ \dots \\ V_p \end{bmatrix}$$

La matrice des composantes principales s'écrit

$$\bar{C}_L = \sum_{j=1}^p X_j D_j^{-1} V_j$$

Comme dans le cas où  $p = 2$ , chaque individu est positionné au barycentre des modalités qu'il a représentée. De plus, il faut noter que les modalités d'une même variable sont centrées : les facteurs opposent les modalités d'une même variable.

**Proposition 9** - ACP des profils colonnes

L'ACP des profils lignes issue de l'AFC réalisée sur le tableau disjonctif complet relatif à  $p$  variables qualitatives conduit à l'analyse spectrale de la matrice  $\bar{D}_L^{-1}$ -symétrique et positive :

$$\bar{B}\bar{A} = \frac{1}{np} \sum_{j=1}^p X_j D_j^{-1} X_j^T$$

La matrice des vecteurs propres  $\bar{D}_L^{-1}$ -orthonormés vérifie

$$\bar{U} = \bar{B}\bar{V}M^{-1/2}$$

La matrice des composantes principales s'écrit

$$\bar{C}_C = p\Delta^{-1}\bar{V}M^{1/2}$$

Chaque bloc  $C_j$  de  $\bar{C}_C$  fournit en lignes les coordonnées des modalités de la variable  $X_j$  et permet la représentation graphique simultanée.

### 6.3.3 AFC du tableau de Burt

Cas où  $p = 2$

Prenons le cas où  $p = 2$  et étudions ce que donne, dans ce cas, l'AFC du tableau de Burt. On se rappelle que l'AFC est une double ACP sur les profils-ligne d'une part et sur les profils-colonne d'autre part. Le tableau de Burt est symétrique, les profils ligne et colonne sont identiques : on s'intéresse donc à une seule des ACP.

On note

$$\tilde{T} = \mathcal{B} = \begin{bmatrix} nD_L & N \\ N^T & nD_C \end{bmatrix}$$

$$\tilde{D}_L = \tilde{D}_c = \frac{1}{2} \begin{bmatrix} D_L & 0 \\ 0 & D_C \end{bmatrix} = \frac{1}{2}\Delta = \bar{D}_c$$

$$\tilde{A} = \tilde{B} = \frac{1}{2} \begin{bmatrix} \mathbb{I}_L & B \\ A & \mathbb{I}_C \end{bmatrix} = \bar{A}\bar{B}$$

On fait l'AFC comme l'ACP des profils lignes de  $\tilde{A}$ .

**Proposition 10** *L'ACP des profils-lignes issue de l'AFC réalisée sur le tableau de Burt relatif à deux variables qualitatives conduit à l'analyse spectrale de la matrice  $\tilde{D}_C$ -symétrique et positive :*

$$\tilde{A}\tilde{B} = (\bar{A}\bar{B})^2.$$

*Elle admet pour matrice de vecteurs propres  $\tilde{D}_C^{-1}$ -orthonormés*

$$\tilde{U} = \tilde{V} = \bar{V}$$

*Les valeurs propres associées vérifient :  $\nu_k = \mu_k^2$ . La matrice des composantes principales s'écrit*

$$\tilde{C}_L = \tilde{C}_C = \begin{bmatrix} C_L \\ C_C \end{bmatrix} \Lambda^{1/2} M.$$

La matrice  $\tilde{C}_L$  permet de représenter simultanément les modalités des deux variables.

Remarques

- Les différentes AFC présentées ci-dessus conduisent à la même représentation simultanée des modalités des 2 variables.
- Dans l'AFC du tableau disjonctif complet comme dans celle du tableau de Burt, on obtient des valeurs propres non nulles qui n'ont pas de sens statistique. Ainsi les valeurs propres ne peuvent plus être interprétées comme une part d'inertie.
- L'AFC du tableau de Burt ne considère que des croisements de variables deux à deux, si on veut étudier des interactions d'ordre plus élevé, il faut recoder les variables.

### Cas où $p$ est quelconque

Le tableau de Burt est symétrique, on ne fera donc qu'une ACP. On note

$$\begin{aligned} \tilde{T} &= \mathcal{B} \\ \tilde{D}_L = \tilde{D}_C &= \frac{1}{p} \Delta = \bar{D}_C \\ \tilde{A} = \tilde{B} &= \frac{1}{np} \mathcal{B} \Delta^{-1} = \bar{A}\bar{B} \end{aligned}$$

**Proposition 11** *L'ACP des profils-lignes issue de l'AFC réalisée sur le tableau de Burt relatif à  $p$  variables qualitatives conduit à l'analyse spectrale de la matrice  $\tilde{D}_C$ -symétrique et positive :*

$$\tilde{A}\tilde{B} = (\bar{A}\bar{B})^2.$$

*Elle admet pour matrice de vecteurs propres  $\tilde{D}_C^{-1}$ -orthonormés*

$$\tilde{U} = \tilde{V} = \bar{V}$$

*Les valeurs propres associées vérifient :  $\nu_k = \mu_k^2$ . La matrice des composantes principales s'écrit*

$$\tilde{C}_L = \tilde{C}_C = \bar{C}_C M^{1/2}.$$

La matrice  $\tilde{C}_L$  permet de représenter simultanément les modalités de toutes les variables. En revanche on ne peut pas faire la représentation des individus quand on fait l'AFC du tableau de Burt.

VALEURS PROPRES			
INERTIE TOTALE . . . .			1.6667
NUMERO	VALEUR PROPRE	POURCENT .	POURCENT . CUMULE
1	0.4816	28.90	28.90
2	0.3847	23.08	51.98
3	0.2110	12.66	64.64
4	0.1576	9.45	74.09
5	0.1501	9.01	83.10
6	0.1233	7.40	90.50
7	0.0815	4.89	95.38
8	0.0457	2.74	98.12
9	0.0235	1.41	99.54
10	0.0077	0.46	100.00

TABLE 6.3 – Valeurs propres associées à l’AFCM des caractéristiques de différentes races de chiens.

### 6.3.4 Interprétation

Comme en ACM, on définit le nuage de points associé aux profils-ligne. L’inertie totale est  $K/p-1$  et la dimension maximum du nuage de points  $K-p$ . La moyenne des valeurs propres sera égale à

$$\frac{K/p-1}{K-p} = 1/p$$

et on retient les axes associées à des valeurs propres supérieures à  $1/p$ ; on peut aussi utiliser la règle du coude. Attention, les valeurs propres ne peuvent pas être interprétées comme des parts d’inertie.

Dans l’exemple, l’inertie totale est égale à 1.67 et il y a 3 valeurs propres supérieures à  $1/p = 1/6$  (voir Table 6.3).

### 6.3.5 Représentation des individus

La variance de l’axe  $h$  est  $\lambda_h$  ce qui est classique en analyse factorielle et la contribution de l’individu  $i$  à l’axe  $h$  est donnée par

$$\frac{\frac{1}{n}c_{ih}}{\lambda_h}$$

où  $c_{ih}$  est la coordonnée de l’individu  $i$  sur l’axe  $h$  (voir 6.4).

### 6.3.6 Représentation des variables

L’inertie apportée par une modalité  $jl$  au nuage de points est

$$\frac{1}{p} \left( 1 - \frac{n_{jl}}{n} \right)$$

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRÉS DES INDIVIDUS  
 AXES 1 A 3

INDIVIDUS			COORDONNEES			CONTRIBUTIONS			COSINUS CARRÉS		
IDENTIFICATEUR	P.REL	DISTO	1	2	3	1	2	3	1	2	3
beauceron	3.70	1.14	-0.32	0.42	0.10	0.8	1.7	0.2	0.09	0.15	0.01
basset	3.70	1.91	0.25	-1.10	0.19	0.5	11.7	0.6	0.03	0.63	0.02
berger allemand	3.70	1.54	-0.49	0.46	0.50	1.8	2.1	4.4	0.15	0.14	0.16
boxer	3.70	1.80	0.45	0.88	-0.69	1.5	7.5	8.4	0.11	0.43	0.27
bull-dog	3.70	1.64	1.01	-0.55	0.16	7.9	2.9	0.5	0.62	0.18	0.02
bull-mastiff	3.70	2.09	-0.75	-0.55	-0.50	4.4	2.9	4.3	0.27	0.14	0.12
caniche	3.70	2.16	0.91	0.02	0.58	6.4	0.0	5.8	0.39	0.00	0.15
chihuahua	3.70	1.86	0.84	-0.84	0.47	5.4	6.9	3.9	0.38	0.38	0.12
cocker	3.70	1.93	0.73	-0.08	-0.66	4.1	0.1	7.7	0.28	0.00	0.23
colley	3.70	1.11	-0.12	0.53	0.33	0.1	2.7	2.0	0.01	0.25	0.10
dalmatien	3.70	1.77	0.65	0.99	-0.46	3.2	9.4	3.7	0.24	0.55	0.12
doberman	3.70	1.56	-0.87	0.32	0.45	5.9	1.0	3.6	0.49	0.06	0.13
dogue allemand	3.70	1.95	-1.05	-0.51	-0.17	8.4	2.5	0.5	0.56	0.13	0.01
epagneul breton	3.70	2.18	0.48	1.04	-0.06	1.8	10.4	0.1	0.10	0.49	0.00
epagneul français	3.70	1.20	-0.14	0.52	-0.12	0.2	2.6	0.2	0.02	0.22	0.01
fox-hound	3.70	1.38	-0.88	-0.03	0.36	5.9	0.0	2.3	0.56	0.00	0.10
fox-terrier	3.70	1.78	0.88	-0.14	-0.05	6.0	0.2	0.1	0.44	0.01	0.00
grand bleu de gascogne	3.70	1.44	-0.52	0.11	-0.04	2.1	0.1	0.0	0.19	0.01	0.00
labrador	3.70	1.77	0.65	0.99	-0.46	3.2	9.4	3.7	0.24	0.55	0.12
levrier	3.70	1.35	-0.68	0.08	0.60	3.5	0.1	6.2	0.34	0.01	0.26
mastiff	3.70	1.90	-0.76	-0.89	-0.59	4.4	7.6	6.1	0.30	0.41	0.18
pékinois	3.70	1.86	0.84	-0.84	0.47	5.4	6.9	3.9	0.38	0.38	0.12
pointer	3.70	1.54	-0.67	0.42	0.69	3.5	1.7	8.3	0.29	0.12	0.31
saint-bernard	3.70	1.69	-0.58	-0.59	-0.89	2.6	3.4	14.0	0.20	0.21	0.47
setter	3.70	1.14	-0.50	0.38	0.29	2.0	1.4	1.5	0.22	0.13	0.07
teckel	3.70	1.64	1.01	-0.55	0.16	7.9	2.9	0.5	0.62	0.18	0.02
terre-neuve	3.70	1.66	-0.38	-0.49	-0.66	1.1	2.3	7.7	0.09	0.14	0.26

TABLE 6.4 – Coordonnées et contributions des individus (modalités) associées à l'AFCM des caractéristiques de différentes races de chiens.

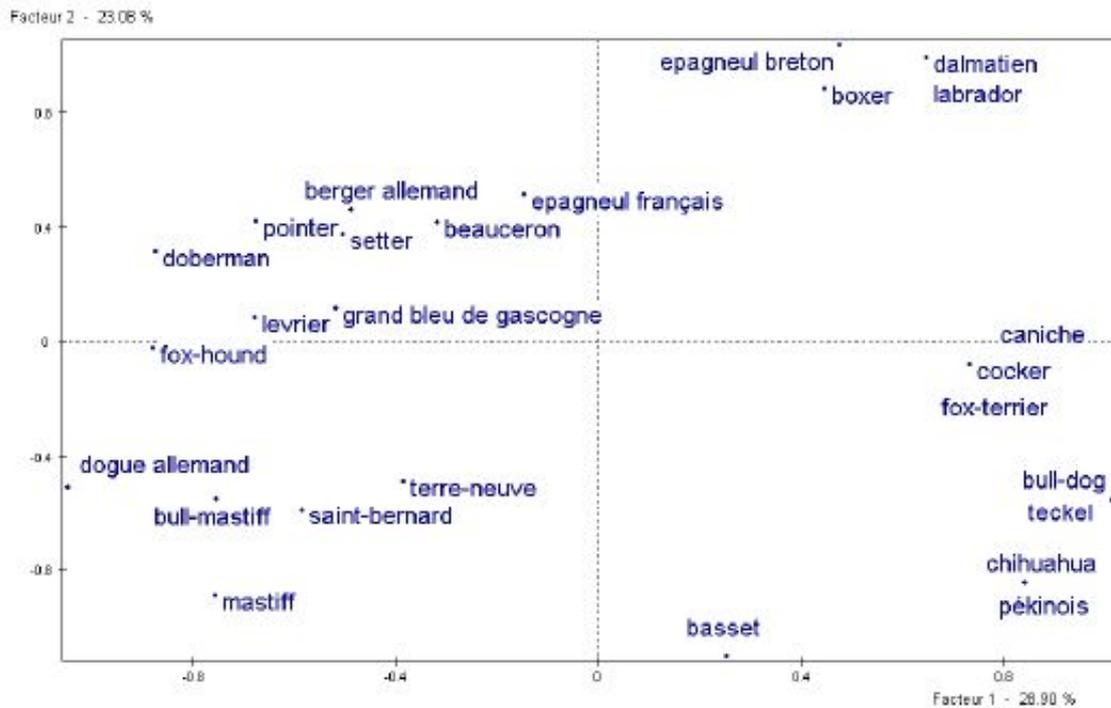


FIGURE 6.1 – Représentation des différentes races de chiens (individus).

Elle est donc d'autant plus forte que l'effectif de la modalité est faible. De nombreuses modalités à faible effectif peuvent donc déséquilibrer une AFCM. Et il est préférable de limiter le nombre de modalités à faible effectif, quitte à redéfinir les modalités.

Par ailleurs, l'inertie apportée par une variable  $j$  est

$$\frac{c_j - 1}{p}$$

Elle est donc d'autant plus importante que le nombre de modalités de la variable est important. Il est donc conseillé de travailler avec des variables ayant des modalités en nombre comparable. La contribution de la modalité  $k$  de la variable  $X_j$  à l'inertie de l'axe  $h$  est donnée par

$$\frac{\frac{n_{jk}}{pn} c_{jk}^2}{\lambda_h}$$

Voir Table 6.5.

En pratique,

- On interprète les proximités et les oppositions entre les modalités des différentes variables
- On privilégie les interprétations sur les modalités suffisamment éloignées du centre du graphique
- Les rapports de valeurs propres ne sont pas interprétables mais on peut regarder la décroissance des valeurs propres pour choisir la dimension.

COORDONNEES, CONTRIBUTIONS ET COSINUS CARRÉS DES MODALITES ACTIVES  
 AXES 1 A 3

MODALITES		COORDONNEES			CONTRIBUTIONS			COSINUS CARRÉS			
LIBELLE	P.REL	DISTO	1	2	3	1	2	3	1	2	3
1 . Taille											
ta-	4.32	2.86	1.18	-0.92	0.62	12.6	9.6	7.8	0.49	0.30	0.13
ta+	3.09	4.40	0.85	1.23	-1.02	4.6	12.2	15.1	0.16	0.34	0.23
ta++	9.26	0.80	-0.84	0.02	0.05	13.5	0.0	0.1	0.88	0.00	0.00
CONTRIBUTION CUMULEE =						30.7	21.8	23.0			
2 . Poids											
po-	4.94	2.38	1.17	-0.82	0.36	14.0	8.7	3.0	0.58	0.29	0.05
po+	8.64	0.93	-0.31	0.82	0.23	1.7	15.1	2.2	0.10	0.72	0.06
po++	3.09	4.40	-1.02	-0.97	-1.22	6.6	7.6	21.8	0.23	0.22	0.34
CONTRIBUTION CUMULEE =						22.3	31.4	27.0			
3 . Vitesse											
v-	6.17	1.70	0.32	-1.04	-0.40	1.3	17.5	4.7	0.06	0.64	0.09
v+	4.94	2.38	0.60	0.89	-0.36	3.7	10.1	3.0	0.15	0.33	0.050
v++	5.56	2.00	-0.89	0.37	0.76	9.2	2.0	15.3	0.40	0.07	0.29
CONTRIBUTION CUMULEE =						14.2	29.6	23.0			
4 . Intelligence											
int-	4.94	2.38	-0.35	-0.81	0.35	1.2	8.4	2.9	0.05	0.28	0.05
int+	8.02	1.08	0.37	0.29	-0.49	2.3	1.7	9.3	0.13	0.08	0.23
int++	3.70	3.50	-0.34	0.46	0.60	0.9	2.0	6.3	0.03	0.06	0.10
CONTRIBUTION CUMULEE =						4.4	12.1	18.5			
5 . Affection											
af-	8.02	1.08	-0.84	-0.29	-0.07	11.6	1.7	0.2	0.65	0.08	0.00
af+	8.64	0.93	0.78	0.27	0.06	10.8	1.6	0.2	0.65	0.08	0.00
CONTRIBUTION CUMULEE =						22.4	3.3	0.3			
6 . Aggressivité											
ag-	8.64	0.93	0.40	0.19	0.31	2.9	0.8	3.9	0.17	0.04	0.10
ag+	8.02	1.08	-0.43	-0.21	-0.33	3.1	0.9	4.2	0.17	0.04	0.10
CONTRIBUTION CUMULEE =						6.0	1.8	8.2			

TABLE 6.5 – Coordonnées et contributions des variables associées à l’AFCM des caractéristiques de différentes races de chiens.

COORDONNEES ET VALEURS-TEST DES MODALITES			
AXES 1 A 3			
MODALITES	VALEURS-TEST		
LIBELLE	1	2	3
1 . Taille			
ta-	3.6	-2.8	1.9
ta+	2.1	3.0	-2.5
ta++	-4.8	0.1	0.3
2 . Poids			
po-	3.9	-2.7	1.2
po+	-1.6	4.3	1.2
po++	-2.5	-2.4	-3.0
3 . Vélodité			
v-	1.3	-4.1	-1.6
v+	2.0	2.9	-1.2
v++	-3.2	1.3	2.8
4 . Intelligence			
int-	-1.2	-2.7	1.2
int+	1.8	1.4	-2.4
int++	-0.9	1.3	1.6
5 . Affection			
af-	-4.1	-1.4	-0.3
af+	4.1	1.4	0.3
6 . Aggressivité			
ag-	2.1	1.0	1.6
ag+	-2.1	-1.0	-1.6

TABLE 6.6 – Coordonnées et valeurs tests des associées à l’AFCM des caractéristiques de différentes races de chiens.

- Seules les contributions des modalités à l’inertie selon les axes sont interprétables

Une modalité  $jl$  a une position sur l’axe  $h$  qui est significativement différente du centre de gravité 0 si

$$\left| C_h(jl) \sqrt{\frac{n_{jl}(n-1)}{n-n_{jl}}} \right| > 2$$

où  $C_h(jl)$  est la coordonnée de la modalité  $l$  de la variable  $j$  sur l’axe  $h$ . Dans l’exemple des chiens on obtient le tableau ci-dessous. On observe par exemple que seule la modalité **int-** de l’intelligence est significative et seulement sur l’axe 2.

Le premier axe factoriel oppose les chiens de grande taille (à gauche) aux chiens de petite taille (à droite), il oppose aussi l’affection faible (à gauche) à l’affectivité (à droite). L’axe deux oppose les chiens lents et légers en bas aux chiens assez rapides en haut et gros.

### 6.3.7 Représentation simultanée

En AFCM, la représentation simultanée est assez naturelle puisqu’on cherche à représenter les individus au centre des modalités qu’il ont choisies et les modalités au centre des individus qui les ont choisies, aux facteurs de dilatation  $1/\sqrt{\lambda}$  près.

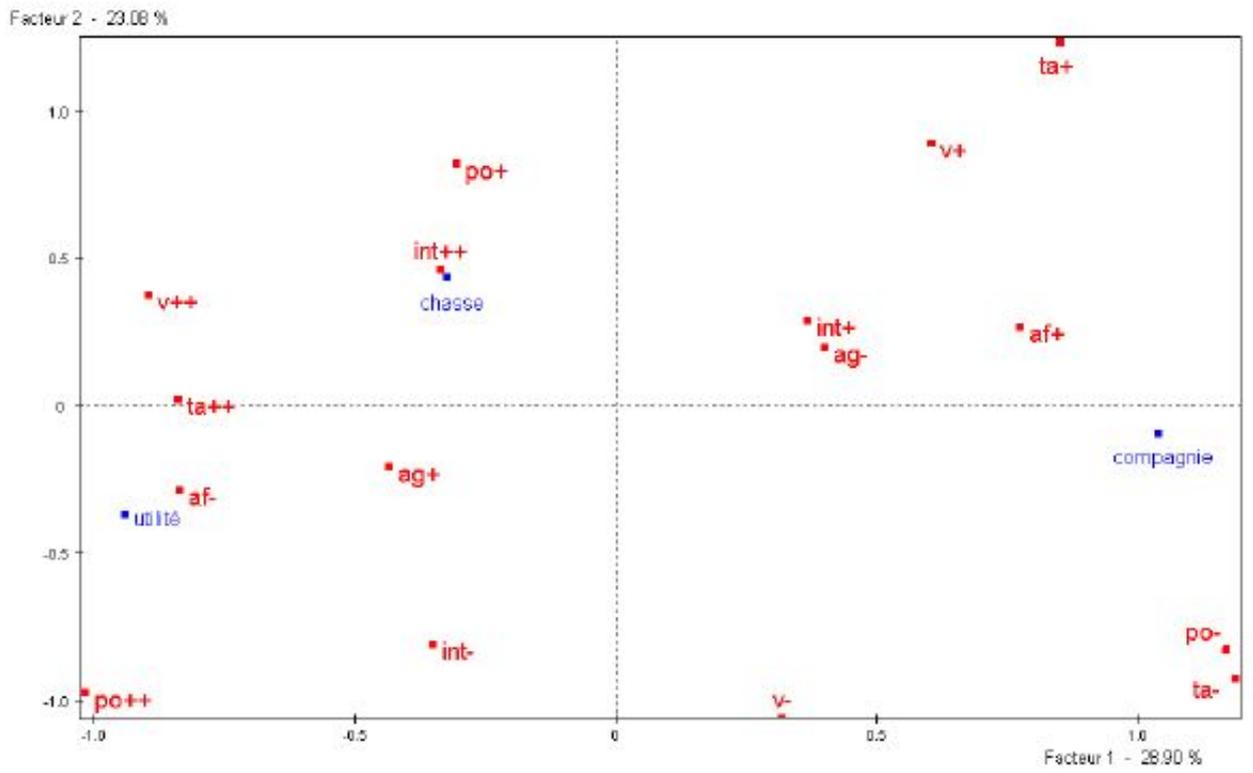


FIGURE 6.2 – Représentation des caractéristiques (variables).

Quand les individus sont nombreux et anonymes cette représentation n'est pas très utile.

Pour l'exemple des races de chiens, nous obtenons la figure 6.3.

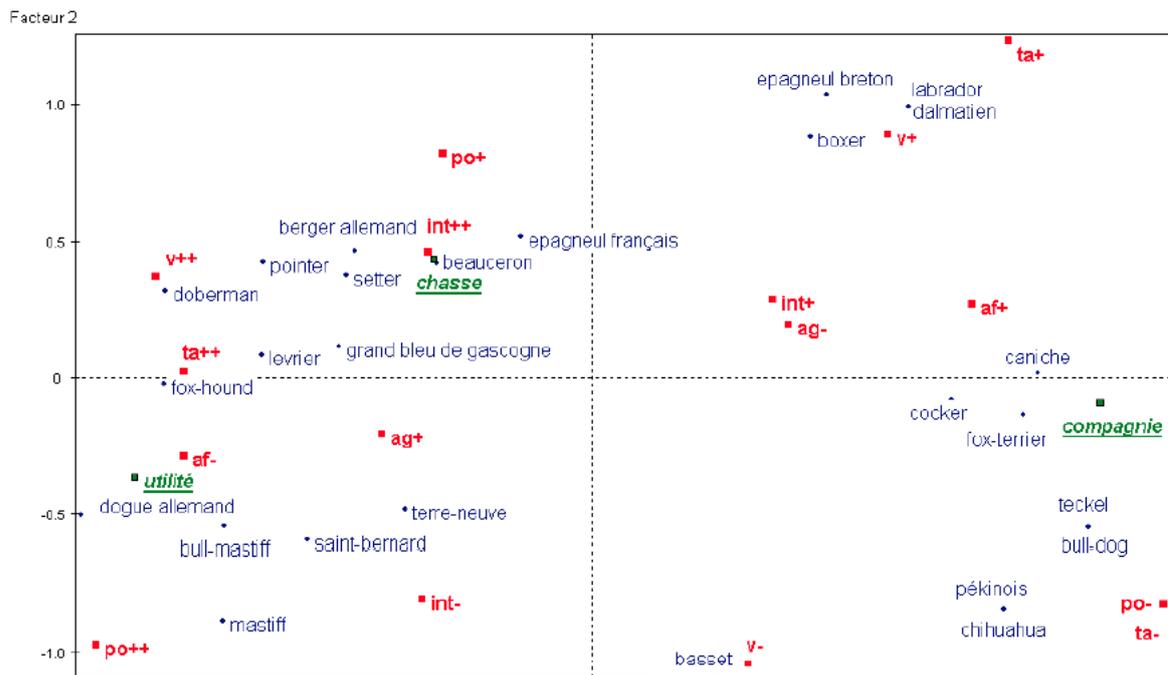


FIGURE 6.3 – AFCM des données sur les races de chiens, la variable d'utilité (en bleu) est considérée comme une variable illustrative (ou supplémentaire).

## 6.4 Individus et variables supplémentaires

La prise en compte d'éléments supplémentaires permet d'aider à la compréhension et à l'interprétation des résultats. Ceci est généralement intéressant quand les variables se décomposent en thèmes. En pratique, les modalités des variables supplémentaires seront prises en compte en projetant, dans les sous espaces factoriels, le centre des groupes d'individus ayant choisi ces modalités. On pourra aussi, dans une optique de prévision, projeter des individus ou des variables supplémentaires pour les situer par rapport aux individus et variables actives.

## 6.5 Les variables continues

Pour être actives dans une analyse factorielle des correspondances multiples, les variables continues doivent être rendues nominales. On devra s'attacher à retenir un nombre de modalités proche de celui des variables discrètes actives dans l'analyse et à avoir une répartition de l'effectif équilibré entre les différentes modalités (ie faire un découpage équiréparti). Cependant, on pourra éventuellement retenir une modalité à faible effectif si celle ci est importante pour l'étude, ou respecter des seuils naturels s'ils existent.

Quand une variable continue est introduite comme variable supplémentaire, on peut avoir intérêt à faire un découpage fin.

On doit être extrêmement vigilant quand on discrétise une variable continue car ceci conduit à une perte d'information brute. Néanmoins, ce type de transformation peut présenter certains avantages :

- ça permet de traiter conjointement des variables continues et discrètes en correspondances multiples,
- ça permet d'observer une éventuelle contiguïté de classes voisines et de valider a posteriori les données,
- et surtout ça permet de mettre en évidence d'éventuelles liaisons non linéaires entre les variables continues.