Chapitre 5 : Analyse des Correspondances Multiples (ACM)

M1 du Master MMAS

James Ledoux

Dépt de mathématiques, Univ. Poitiers

12 juillet 2009

- n individus décrits par deux variables qualitatives X_1, X_2 à m_1 et m_2 modalités :

AFC sur la table de contingence de (X_1, X_2)

« $\Updownarrow$ »

Analyser le tableau binaire à n lignes et $(m_1 + m_2)$ colonnes décrivant les réponses

	X_1			X_2		
	1	$\cdots$	m_1	1	$\cdots$	m_2
1						
$\vdots$						
i	1	$0\cdots$	0	$0\cdots$	010	$\cdots 0$
$\vdots$						
$\boldsymbol{n}$						

➜ plus coûteux mais se généralise à **$\boldsymbol{K} > 2$ variables observées**

- n individus décrits par K variables **qualitatives** à $m_1, \ldots, m_K$ modalités
- Étude de tableaux de données issus d'enquêtes avec réponses à choix multiples des catégories (rajouter les conditions une et une seule réponse)

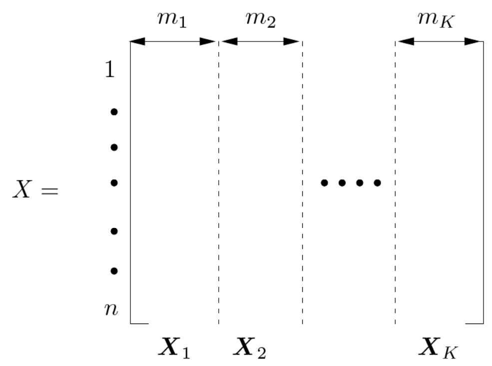
N°	X_1	X_2	X_3
1	1	2	3
2	2	1	1
3	2	2	2
4	3	2	1
5	3	1	2

$m_1 = 3$			$m_2 = 2$		$m_3 = 3$		
1	0	0	0	1	0	0	1
0	1	0	1	0	1	0	0
0	1	0	0	1	0	1	0
0	0	1	0	1	1	0	0
0	0	1	1	0	0	1	0

$$X = [\boldsymbol{X}_1 \mid \boldsymbol{X}_2 \mid \boldsymbol{X}_3]$$

TABLE 17: « Tableau sous forme de codage condensé » en Tableau disjonctif : $n = 5$ individus auxquels on a posé trois questions X_1, X_2, X_3 avec respectivement $3, 2, 3$ modalités de réponse.

- ACM $\equiv$ formellement appliquer l'AFC au Tableau dit Disjonctif Complet des $m := m_1 + \cdots + m_K$ variables indicatrices



$$\boldsymbol{x}_{k;i,j} := \begin{cases} 1 & \text{si l'ind. N° i présente la modalité } j \text{ de la variable N° } k \\ 0 & \text{autrement} \end{cases}$$

$$x_{i,j} := \begin{cases} 1 & \text{si l'ind. N° i présente la modalité } j \\ 0 & \text{autrement} \end{cases}$$

- L'ensemble des individus $I := \{1, \ldots, n\}$
- L'ensemble des modalités
  $J := \{1, \ldots, m_1, m_1 + 1, \ldots, m_1 + m_2, \ldots m_1 + \cdots m_k = m\}$
- Chaque ligne de $\boldsymbol{X}_k$ somme à 1 (une réponse unique à la question N° k)
- Chaque ligne de X somme à K, le nombre de variables (ou questions)
- La colonne j de $\boldsymbol{X}_k$ représente un vecteur binaire indiquant les individus à avoir choisi la modalité j pour répondre à la question N° k
  Cette colonne somme à $n_{k;.j}$, le nombre d'individus à répondre j à la question N° k
- La colonne j de X représente un vecteur binaire indiquant les individus à avoir choisie la modalité j (associée à une certaine variable) : on parle de variable indicatrice de la modalité.
  La somme de cette colonne est noté $n_{.j}$ et représente l'effectif d'individus ayant choisi j.

N°	$\boldsymbol{X}_1$			$\boldsymbol{X}_2$		$\boldsymbol{X}_3$			Marge	
1	1	0	0	0	1	0	0	1	$K=3$	$n_{1.}$
2	0	1	0	1	0	1	0	0	$K=3$	$n_{2.}$
3	0	1	0	0	1	0	1	0	$K=3$	$n_{3.}$
4	0	0	1	0	1	1	0	0	$K=3$	$n_{4.}$
5	0	0	1	1	0	0	1	0	$K=3$	$n_{5.}$
Marge	1	2	2	2	3	2	2	1	15	
	$n_{1;.1}$	$n_{1;.2}$	$n_{1;.3}$	$n_{2;.1}$	$n_{2;.2}$	$n_{3;.1}$	$n_{3;.2}$	$n_{3;.3}$	nK	
	$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{.4}$	$n_{.5}$	$n_{.6}$	$n_{.7}$	$n_{.8}$	nK	

TABLE 18: Tableau disjonctif $n \times m$ avec $m = \sum_k m_k$ le nombre total de modalités

$$n_{i.} := \sum_{j \in J} x_{i,j} = \sum_{k=1}^{K} \sum_{j=1}^{m_k} \boldsymbol{x}_{k;ij} = K \quad X\mathbf{1}_n = K\mathbf{1}_n$$

$$k = 1, \ldots, K \quad \sum_{j=1}^{m_k} n_{k;.j} = n$$

$$\text{Effectif total} = \sum_{i \in I} n_{i.} = \sum_{j \in J} n_{.j} = \sum_{k=1}^{K} \sum_{j=1}^{m_k} n_{k;.j} = nK$$

	$\boldsymbol{F}_1 = \frac{1}{nK}\boldsymbol{X}_1$			$\boldsymbol{F}_2 = \frac{1}{nK}\boldsymbol{X}_2$		$\boldsymbol{F}_3 = \frac{1}{nK}\boldsymbol{X}_3$			Marge col.
	$\frac{1}{nK}$	0	0	0	$\frac{1}{nK}$	0	0	$\frac{1}{nK}$	$f_{1.} = \frac{1}{n}$
	0	$\frac{1}{nK}$	0	$\frac{1}{nK}$	0	$\frac{1}{nK}$	0	0	$f_{2.} = \frac{1}{n}$
	0	$\frac{1}{nK}$	0	0	$\frac{1}{nK}$	0	$\frac{1}{nK}$	0	$f_{3.} = \frac{1}{n}$
	0	0	$\frac{1}{nK}$	0	$\frac{1}{nK}$	$\frac{1}{nK}$	0	0	$f_{4.} = \frac{1}{n}$
	0	0	$\frac{1}{nK}$	$\frac{1}{nK}$	0	0	$\frac{1}{nK}$	0	$f_{5.} = \frac{1}{n}$
Marge lig.	$\frac{n_{1;.1}}{nK}$	$\frac{n_{1;.2}}{nK}$	$\frac{n_{1;.3}}{nK}$	$\frac{n_{2;.1}}{nK}$	$\frac{n_{2;.2}}{nK}$	$\frac{n_{3;.1}}{nK}$	$\frac{n_{3;.2}}{nK}$	$\frac{n_{3;.3}}{nK}$	1
	$\frac{n_{.1}}{nK}$	$\frac{n_{.2}}{nK}$	$\frac{n_{.3}}{nK}$	$\frac{n_{.4}}{nK}$	$\frac{n_{.5}}{nK}$	$\frac{n_{.6}}{nK}$	$\frac{n_{.7}}{nK}$	$\frac{n_{.8}}{nK}$	1

TABLE 19: Tableau $n \times m$ F des fréquences relatives

où $\quad f_{i.} := 1/n \quad f_{.j} := n_{.j}/nK$ ou $f_{k;.j} := n_{k;.j}/nK$

$$\boldsymbol{X}_k \mathbf{1}_{m_k} = (\frac{1}{nK})\mathbf{1}_n \quad X\mathbf{1}_m = (1/n)\mathbf{1}_n$$

$$\boxed{F = [\boldsymbol{F}_1 \mid \boldsymbol{F}_2 \mid \boldsymbol{F}_3] = \frac{1}{nK}X}$$

	$\boldsymbol{Z}_{I,1} := D_I^{-1}\boldsymbol{F}_1 = \frac{1}{K}\boldsymbol{F}_1$			$\boldsymbol{Z}_{I,2} := \frac{1}{K}\boldsymbol{F}_2$		$\boldsymbol{Z}_{I,3} := \frac{1}{K}\boldsymbol{F}_3$			
	$\frac{1}{K}$	0	0	0	$\frac{1}{K}$	0	0	$\frac{1}{K}$	1
	0	$\frac{1}{K}$	0	$\frac{1}{K}$	0	$\frac{1}{K}$	0	0	1
	0	$\frac{1}{K}$	0	0	$\frac{1}{K}$	0	$\frac{1}{K}$	0	1
	0	0	$\frac{1}{K}$	0	$\frac{1}{K}$	$\frac{1}{K}$	0	0	1
	0	0	$\frac{1}{K}$	$\frac{1}{K}$	0	0	$\frac{1}{K}$	0	1
Prof. Moy.	$\frac{n_{1;.1}}{nK}$	$\frac{n_{1;.2}}{nK}$	$\frac{n_{1;.3}}{nK}$	$\frac{n_{2;.1}}{nK}$	$\frac{n_{2;.2}}{nK}$	$\frac{n_{3;.1}}{nK}$	$\frac{n_{3;.2}}{nK}$	$\frac{n_{3;.3}}{nK}$	1
	$\frac{n_{.1}}{nK}$	$\frac{n_{.2}}{nK}$	$\frac{n_{.3}}{nK}$	$\frac{n_{.4}}{nK}$	$\frac{n_{.5}}{nK}$	$\frac{n_{.6}}{nK}$	$\frac{n_{.7}}{nK}$	$\frac{n_{.8}}{nK}$	1

TABLE 20: Tableau $n \times m$ des profils-lignes : $Z_I = D_I^{-1}F$

où $\quad D_I := \frac{1}{n}I_n \qquad Z_I = [\boldsymbol{Z}_{I,1} \mid \boldsymbol{Z}_{I,2} \mid \boldsymbol{Z}_{I,3}] = nF = \frac{1}{K}X$

$\boldsymbol{Z}_{J,1} := \boldsymbol{F}_1\boldsymbol{D}_1^{-1}$			$\boldsymbol{Z}_{J,2} := \boldsymbol{F}_2\boldsymbol{D}_2^{-1}$		$\boldsymbol{Z}_{J,3} := \boldsymbol{F}_3\boldsymbol{D}_3^{-1}$			Prof. Moy.
$\frac{1}{n_{1;.1}}$	0	0	0	$\frac{1}{n_{2;.2}}$	0	0	$\frac{1}{n_{3;.3}}$	$\frac{1}{n}$
0	$\frac{1}{n_{1;.2}}$	0	$\frac{1}{n_{2;.1}}$	0	$\frac{1}{n_{3;.1}}$	0	0	$\frac{1}{n}$
0	$\frac{1}{n_{1;.2}}$	0	0	$\frac{1}{n_{2;.2}}$	0	$\frac{1}{n_{3;.2}}$	0	$\frac{1}{n}$
0	0	$\frac{1}{n_{1;.3}}$	0	$\frac{1}{n_{2;.2}}$	$\frac{1}{n_{3;.1}}$	0	0	$\frac{1}{n}$
0	0	$\frac{1}{n_{1;.3}}$	$\frac{1}{n_{2;.1}}$	0	0	$\frac{1}{n_{3;.2}}$	0	$\frac{1}{n}$
1	1	1	1	1	1	1	1	1

TABLE 21: Tableau de profils-colonnes : $Z_J = FD_J^{-1}$

où D_J est la matrice $m \times m$ diagonale par blocs :

$$D_J = \begin{cases} \mathrm{Diag}\left(\boldsymbol{D}_k\right) \text{ avec } \boldsymbol{D}_k := \mathrm{Diag}\left(\frac{n_{k;.j}}{nK}\right) \ m_k \times m_k \\ \mathrm{Diag}\left(\frac{n_{.j}}{nK}\right) = \frac{1}{nK}D \text{ avec } D := \mathrm{Diag}\left(n_{.j}\right) \end{cases}$$

Les différentes matrices en fonction du tableau de données

Fréquences relatives : $F = \dfrac{1}{nK}X$

Poids individus : $D_I = \dfrac{1}{n}I_n \quad D_I^{-1} = nI_n$

Poids modalités : $D_J = \dfrac{1}{nK}D \quad D_J^{-1} = nKD^{-1}$

avec $D := \mathrm{diag}(n_{.j})$

Produit scalaire individus : $\langle u, v\rangle_{\chi^2} = u^\top D_J^{-1}v = nK\, u^\top D^{-1}v$

Produit scalaire modalités : $\langle u, v\rangle_{\chi^2} = u^\top D_I^{-1}v = n\, u^\top v$

Profils-lignes : $Z_I = nF = \dfrac{1}{K}X$

Profils-colonnes : $Z_J = XD^{-1}$

Distance du chi-deux pour les individus

- Nuage dans $\mathbb{R}^{|J|} = \mathbb{R}^m$ avec m le nombre total de modalités
- Centre de gravité $G_I = (n_{.1}/nK, n_{.2}/nK, \ldots, n_{.m}/nK)$
- $d_{\chi^2}(\mathrm{PrL}(i), \mathrm{PrL}(i'))^2$
$$\begin{aligned} d_{\chi^2}(\mathrm{PrL}(i), \mathrm{PrL}(i'))^2 &= \sum_{j\in J}\frac{1}{f_{.j}}\left(\frac{f_{ij}}{f_{i.}} - \frac{f_{i'j}}{f_{i'.}}\right)^2 \\ &= \sum_{j\in J}\frac{nK}{n_{.j}}\left(\frac{x_{ij}}{K} - \frac{x_{i'j}}{K}\right)^2 \\ &= \frac{1}{K}\sum_{j\in J}\frac{n}{n_{.j}}(x_{ij} - x_{i'j})^2 \end{aligned}$$
  - L'expression $(x_{ij} - x_{i'j})^2 \in \{0, 1\}$ et ne diffère de 0 que pour les modalités j choisies par un seul des deux individus
    ↪ la distance croît avec le nombre de choix différents de modalités
  - une modalité intervient avec le poids $n/n_{.j}$, inverse de sa fréquence : une modalité intervient donc d'autant plus dans le calcul de la distance entre deux individus qu'elle a été peu choisie

Distance du chi-deux pour les modalités

- Nuage dans $\mathbb{R}^{|I|} = \mathbb{R}^n$ avec n le nombre d'individus
- Centre de gravité $G_J = (1/n, 1/n, \ldots, 1/n)$
- $d_{\chi^2}(\mathrm{PrC}(j), \mathrm{PrC}(j'))^2$
$$= \sum_{i\in I}\frac{1}{f_{i.}}\left(\frac{f_{ij}}{f_{.j}} - \frac{f_{ij'}}{f_{.j'}}\right)^2 = \sum_{i\in I} n\left(\frac{x_{ij}}{n_{.j}} - \frac{x_{ij'}}{n_{.j'}}\right)^2 = \frac{nn_{j\Delta j'}}{n_{.j}n_{.j'}}$$
$= \dfrac{n}{n_{.j}n_{.j'}}$nbre d'individus présentant une et une seule des 2 mod.
  - Deux modalités choisies par les mêmes individus sont égales.
  - La distance croît avec le nombre de « désaccords » sur les deux modalités j et j'.
  - Deux modalités d'une même variable sont nécessairement assez éloignées l'une de l'autre.
  - La distance décroît avec l'effectif de chacune des deux modalités.
  - Une modalité est d'autant plus proche de G_J que l'effectif d'individus à l'avoir choisi est grand.
  - Les modalités rares sont éloignées de toutes les autres (et en particulier de G_J).

- **Sous-nuage associé à une modalité**

Le nuage des modalités peut être partagé en K sous-nuages :

- m_k points correspondants aux colonnes de la matrice des profils-colonnes : $\boldsymbol{X}_k \boldsymbol{D}_k^{-1}$
- Poids relatif aux m_k points : $n_{k;.j}/n$
- Le sous-nuage ont même centre de gravité G_J que le nuage global :

$$\sum_{j=1}^{m_k} \frac{n_{k;.j}}{n} (\boldsymbol{X}_k \boldsymbol{D}_k^{-1})_{.,j} = \sum_{j=1}^{m_k} \frac{n_{k;.j}}{n} \left(\frac{\boldsymbol{x}_{k;i,j}}{n_{k;.j}} \right)_i = \frac{1}{n} \sum_{j=1}^{m_k} (\boldsymbol{x}_{k;i,j})_i = \frac{1}{n} \mathbf{1}_n$$

Inertie dans le nuage des modalités

- **Inertie d'une modalité par rapport à G_J** :

$$\begin{aligned} I(\mathrm{PrC}(j)) &= f_{.j}\, d_{\chi^2}(\mathrm{PrC}(j), G_J)^2 \\ &= f_{.j} \times \frac{n(n-n_{.j})}{n_{.j} n} = \frac{n_{.j}}{nK} \frac{n-n_{.j}}{n_{.j}} \\ &= \frac{1}{K}\left(1 - \frac{n_{.j}}{n}\right) \end{aligned}$$

↬ la part d'inertie due à une modalité est d'autant plus grande qu'elle est rarement sélectionnée. Il est alors courant que les premiers axes d'une ACM soient essentiellement dus à des modalités très rares partagées par quelques individus

- **Inertie d'une variable X_k par rapport à G_J** :

$$I(X_k) = \sum_{j \in J_k} I(\mathrm{PrC}(j)) = \sum_{j \in J_k} \frac{1}{K}\left(1 - \frac{n_{.j}}{n}\right) = \frac{1}{K}(m_k - 1)$$

  - C'est une fonction croissante du nombre de modalités m_k de la variable.
  - Elle est minimale pour $m_k = 2$ modalités
    ↬ faire attention aux découpages en modalités des variables

- **Inertie totale** : $I(\mathcal{N}(J)) = \sum_{k=1}^{K} I(X_k) = \dfrac{m}{K} - 1$
  - Elle vaut 1 lorsque toutes les variables présentent deux modalités On peut alors montrer que l'ACM et l'ACP donnent des résultats équivalents
  - L'inertie totale ne dépend que des nombres de variables et de modalités : peu d'intérêt comme en ACP

- Les axes factoriels s'obtiennent à partir des résultats : (Th 10, Prop. 3 et (28,29)), c'est-à-dire comme les vecteurs propres orthonormés de la matrice :

$$F^\top D_I^{-1} F D_J^{-1} = Z_I{}^\top Z_J = \frac{1}{K} X^\top X D^{-1} \quad \text{où } D := \mathrm{diag}(n_{.j}).$$

- **Nombre d'axes**
  - Les colonnes du tableau des profils-colonnes $Z_J = XD^{-1}$ engendrent un sous-espace de dimension le $\mathrm{rang}(XD^{-1}) = \mathrm{rang}(X)$.
  - Tous les sous-espaces engendrés par les $\boldsymbol{X}_k$ ont en commun la première bissectrice $\mathbf{1}_n$.

  ↬ $\mathrm{rang}(X) \le m_1 + (m_2 - 1) + \cdots + (m_K - 1) = m - K + 1$
  et $\mathrm{rang}(X^\top X D^{-1}) \le m - K + 1$:
  au plus $m - K + 1$ valeurs propres non-nulles

  - Analyse par rapport à l'origine : $\mathbf{1}_n$ est vecteur propre associé à la v.p. triviale 1 donc au mieux $m - K$ valeurs propres pertinentes
  - Analyse par rapport au centre de gravité : on aura au mieux $m - K$ valeurs propres non-nulles

- Les facteurs G_l et F_l, correspondant à la projection des profils-lignes et profils-colonnes sur l'axe factoriel de rang l :

$$\begin{aligned} G_l &= Z_I D_J^{-1} u_l = \frac{1}{K} n K D^{-1} u_l = n D^{-1} u_l \\ F_l &= Z_J{}^\top D_I^{-1} v_l = \frac{1}{n} D^{-1} X^\top v_l \end{aligned}$$

- Les **relations axes-facteurs** se déduisent de la Proposition 5 :

$$\begin{aligned} G_l &= \sqrt{\lambda_l}\, D_I^{-1} v_l = n\sqrt{\lambda_l}\, v_l \\ F_l &= \sqrt{\lambda_l}\, D_J^{-1} u_l = nK\sqrt{\lambda_l}\, D^{-1} u_l \end{aligned}$$

- **Relations dites quasi-barycentriques** : à partir de la Proposition 6

$$\begin{aligned} G_l &= \frac{1}{\sqrt{\lambda_l}} \frac{1}{K} X F_l \\ F_l &= \frac{1}{\sqrt{\lambda_l}} D^{-1} X^\top G_l \end{aligned}$$

$$i \in I, \quad G_l(i) = \frac{1}{\sqrt{\lambda_l}} \sum_{j \in J} \left(\frac{f_{ij}}{f_{i.}}\right) F_l(j) = \frac{1}{\sqrt{\lambda_l}} \left[\frac{1}{K} \sum_{j \in J(i)} F_l(j)\right]$$

$$j \in J, \quad F_l(j) = \frac{1}{\sqrt{\lambda_l}} \sum_{i \in I} \left(\frac{f_{ij}}{f_{.j}}\right) G_l(i) = \frac{1}{\sqrt{\lambda_l}} \left[\frac{1}{n_{.j}} \sum_{i \in I(j)} G_l(i)\right]$$

où

- $J(i)$ désigne les modalités sélectionnées par l'individu i
- et $I(j)$ l'ensemble des individus ayant choisi la modalité j

↬ Au facteur de dilatation près $1/\sqrt{\lambda_l}$:
la coord. de la projection de i sur l'axe factoriel correspond à la moyenne arithmétique des coord. de l'ensemble des modalités qu'il a choisi ;
la coord. de la projection de la modalité j sur l'axe factoriel correspond à la moyenne arithmétique des coord. de l'ensemble des individus qui ont choisi la modalité

- L'inertie totale n'est pas pertinente (déjà vu)
- **Choix du nombre d'axes :** idem, avec une règle « à la Kaiser » fondée sur la remarque suivante :

$$I(\mathcal{N}(I)) = \frac{m}{K} - 1 = (m-K)\frac{1}{K} = \sum_{l=1}^{m-K} \lambda_l$$

D'où ne retenir qu'une valeur propre de valeur $> 1/K$.

- Les règles d'interprétation des résultats (coordonnées, contributions, cosinus carrés) concernant les éléments actifs d'une ACM sont sensiblement les mêmes que pour une AFC
- La structure spécifique des colonnes-modalités de la matrice doit être prise en compte :
  - la contribution d'une variable à l'axe l en sommant les contributions de ses modalités sur cet axe :

$$\mathrm{CTR}_l(X_k) = \sum_{j \in J_k} \mathrm{CTR}_l(j)$$

    où J_k est l'ensemble des m_k modalités de la variable X_k
  - On repère ainsi les variables qui ont participé à la construction de l'axe factoriel

À partir des interprétations sur la distance du chi-deux utilisées pour chacun des nuages :

- **La proximité entre individus en terme de ressemblance :**
deux individus se ressemblent s'ils ont choisi globalement les mêmes modalités.
- **La proximité entre modalités de variables différentes en terme d'association :**
ces modalités correspondent aux points moyens des individus qui les ont choisies et sont donc proches parce qu'elles concernent globalement les mêmes individus ou des individus semblables.
- **La proximité entre deux modalités d'une même variable en terme de ressemblance :**
par construction, les modalités d'une même variable s'excluent. Si elles sont proches, cette proximité s'interprète en termes de ressemblance entre les groupes d'individus qui les ont choisies (vis-à-vis d'autres variables actives de l'analyse)

- **Éléments supplémentaires**
  - **Variables qualitatives** :
    Les variables supplémentaires peuvent être placées sur les graphiques factoriels à l'aide de la (seconde) formule quasi-barycentrique donnée page 220 reliant la coordonnée factorielle d'une modalité au coordonnées factorielles des individus actifs.
  - **Variables quantitatives** : construire le cercle des corrélations avec les facteurs

Définition 20 (Tableau de Burt)

*On appelle **tableau de Burt**, le tableau $m \times m$ défini par $X^\top X$:*

$$[\boldsymbol{X}_1\ \boldsymbol{X}_2 \cdots \boldsymbol{X}_K]^\top [\boldsymbol{X}_1\ \boldsymbol{X}_2 \cdots \boldsymbol{X}_K] = \begin{bmatrix} \boldsymbol{X}_1{}^\top \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_1{}^\top \boldsymbol{X}_K \\ \boldsymbol{X}_2{}^\top \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_1{}^\top \boldsymbol{X}_K \\ \vdots & \vdots & \vdots \\ \boldsymbol{X}_K{}^\top \boldsymbol{X}_1 & \cdots & \boldsymbol{X}_K{}^\top \boldsymbol{X}_K \end{bmatrix}$$

Chaque tableau $\boldsymbol{X}_i{}^\top \boldsymbol{X}_j$ correspond à la table de contingence croisant les deux variables X_i et X_j.

Proposition 7 (ACM et AFC du tableau de Burt)

Si on soumet le tableau de Burt à une AFC, les modalités ont les mêmes coordonnées factorielles que dans l'ACM du tableau disjonctif complet X à une constante multiplicative près.

Une ACP d'une famille de K variables continues, construit une famille de variables synthétiques, les composantes principales, qui maximisent

$$\sum_{k=1}^{K} \rho^2_{Z,X_k}.$$

Dans le cas de l'ACM, on peut démontrer le résultat suivant :

Proposition 8

Les facteurs G_l d'une ACM définissent des variables numériques qui maximisent la somme de carrées des rapports de corrélation (voir Def 17)

$$\sum_{k=1}^{K} r^2_{Z:X_k}$$

On résume les K var. qualitatives par K var. numériques les plus plus corrélées possibles aux variables $X_1, \ldots, X_K$ selon l'indicateur du rapport de corrélation.

Exemple 14

ACM sur des caractéristiques de 2 races canines 27 races canines sont décrites via 7 variables qualitatives :

Identificateur	Modalités		
Taille	Petite Taille	Taille Moyenne	Grande Taille
Poids	Petit Poids	Poids Moyen	Poids Élevé
Vélocité	Lent	Assez Rapide	Très Rapide
Intelligence	Peu Intelligent	Intelligence Moyenne	Très Intelligent
Affection	Peu Affectueux	Affectueux	
Agressivité	Peu Agressif	Agressif	
Fonction	Compagnie	Chasse	Utilité

Les six premières sont utilisées pour réaliser l'ACM $(K = 6)$, et la dernière sera considérée comme supplémentaire.

	Identificateur	Taille	Poids	Vélocité	Intelligence	Affection	Agressivité	Fonction
1	**Beauceron**	3	2	3	3	2	2	3
	Basset	1	1	1	1	1	2	2
1	**Berger Allemand**	3	2	3	3	2	2	3
	Boxer	2	2	2	2	2	2	1
5	**Bull-Dog**	1	1	1	2	2	1	1
	Bull-Mastiff	3	3	1	3	1	2	3
	Caniche	1	1	2	3	2	1	1
8	**Chihuahua**	1	1	1	1	2	1	1
	Cocker	2	1	1	2	2	2	1
	Colley	3	2	3	2	2	1	1
11	**Dalmatien**	2	2	2	2	2	1	1
	Dobermann	3	2	3	3	1	2	3
	Dogue Allemand	3	3	3	1	1	2	3
	Épagneul Breton	2	2	2	3	2	1	2
	Épagneul Français	3	2	2	2	1	1	2
	Fox-Hound	3	2	3	1	1	2	2
	Fox-Terrier	1	1	2	2	2	2	1
	Grand Bleu de Gascogne	3	2	2	1	1	2	2
19	**Labrador**	2	2	2	2	2	1	2
	Lévrier	3	2	3	1	1	1	2
	Mastiff	3	3	1	1	1	2	3
22	**Pékinois**	1	1	1	1	2	1	1
	Pointer	3	2	3	3	1	1	2
	Saint-Bernard	3	3	1	2	1	2	3
	Setter	3	2	3	2	1	1	2
26	**Teckel**	1	1	1	2	2	1	1
	Terre-Neuve	3	3	1	2	1	1	3

TABLE 22: Table des données sous forme condensée de l'Exemple 14

Identificateur	Taille			Poids			Vélocité			Intellig.			Affect.		Agress.		Fonction		
Beauceron	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0	1
Basset	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0
Berger Allemand	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	1	0	0	1
Boxer	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	0
Bull-Dog	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0
Bull-Mastiff	0	0	1	0	0	1	1	0	0	0	0	1	1	0	0	1	0	0	1
Caniche	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1	0	1	0	0
Chihuahua	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1	0	1	0	0
Cocker	0	1	0	1	0	0	1	0	0	0	1	0	0	1	0	1	1	0	0
Colley	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	0	1	0	0
Dalmatien	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	1	0	0
Dobermann	0	0	1	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0	1
Dogue Allemand	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1
Épagneul Breton	0	1	0	0	1	0	0	1	0	0	0	1	0	1	1	0	0	1	0
Épagneul Français	0	0	1	0	1	0	0	1	0	0	1	0	1	0	1	0	0	1	0
Fox-Hound	0	0	1	0	1	0	0	0	1	1	0	0	1	0	0	1	0	1	0
Fox-Terrier	1	0	0	1	0	0	0	1	0	0	1	0	0	1	0	1	1	0	0
Grand Bleu de Gascogne	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1	0	1	0
Labrador	0	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0
Lévrier	0	0	1	0	1	0	0	0	1	1	0	0	1	0	1	0	0	1	0
Mastiff	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1
Pékinois	1	0	0	1	0	0	1	0	0	1	0	0	0	1	1	0	1	0	0
Pointer	0	0	1	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	0
Saint-Bernard	0	0	1	0	0	1	1	0	0	0	1	0	1	0	0	1	0	0	1
Setter	0	0	1	0	1	0	0	0	1	0	1	0	1	0	1	0	0	1	0
Teckel	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	0	1	0	0
Terre-Neuve	0	0	1	0	0	1	1	0	0	0	1	0	1	0	1	0	0	0	1

TABLE 23: Tableau Disjonctif Complet des données de l'Exemple 14


```
    | TA1 TA2 TA3 | PO1 PO2 PO3 | VE1 VE2 VE3 | IN1 IN2 IN3 | AF1 AF2 | AG1 AG2 |
----+-------------+-------------+-------------+-------------+---------+---------+
TA1 |   7   0   0 |
TA2 |   0   5   0 |
TA3 |   0   0  15 |
----+-------------+-------------+
PO1 |   7   1   0 |   8   0   0 |
PO2 |   0   4  10 |   0  14   0 |
PO3 |   0   0   5 |   0   0   5 |
----+-------------+-------------+-------------+
VE1 |   5   1   4 |   6   0   4 |  10   0   0 |
VE2 |   2   4   2 |   2   6   0 |   0   8   0 |
VE3 |   0   0   9 |   0   8   1 |   0   0   9 |
----+-------------+-------------+-------------+-------------+
IN1 |   3   0   5 |   3   3   2 |   4   1   3 |   8   0   0 |
IN2 |   3   4   5 |   4   6   2 |   5   5   2 |   0  12   0 |
IN3 |   1   1   5 |   1   5   1 |   1   2   4 |   0   0   7 |
----+-------------+-------------+-------------+-------------+---------+
AF1 |   1   0  12 |   1   7   5 |   5   2   6 |   6   4   3 |  13   0 |
AF2 |   6   5   3 |   7   7   0 |   5   6   3 |   2   8   4 |   0  14 |
----+-------------+-------------+-------------+-------------+---------+---------+
AG1 |   5   3   6 |   5   8   1 |   5   5   4 |   3   8   3 |   5   9 |  14   0 |
AG2 |   2   2   9 |   3   6   4 |   5   3   5 |   5   4   4 |   8   5 |   0  13 |
----+-------------+-------------+-------------+-------------+---------+---------+
    | TA1 TA2 TA3 | PO1 PO2 PO3 | VE1 VE2 VE3 | IN1 IN2 IN3 | AF1 AF2 | AG1 AG2 |
```

TABLE 24: Tableau de Burt


```
+--+--------+-------+--------+----------------------------------------------------------------------------------+
|No| VALEUR | POURC.| POURC. |                                                                                  |
|  | PROPRE |       | CUMULE |                                                                                  |
+--+--------+-------+--------+----------------------------------------------------------------------------------+
| 1| 0.4876 | 29.26 |  29.26 | ********************************************************************************** |
| 2| 0.3857 | 23.14 |  52.40 | *****************************************************************                  |
| 3| 0.2207 | 13.24 |  65.64 | *************************************                                              |
| 4| 0.1645 |  9.87 |  75.51 | ****************************                                                       |
| 5| 0.1487 |  8.92 |  84.43 | *************************                                                          |
| 6| 0.1018 |  6.11 |  90.54 | *****************                                                                  |
| 7| 0.0813 |  4.88 |  95.41 | **************                                                                     |
| 8| 0.0447 |  2.68 |  98.09 | ********                                                                           |
| 9| 0.0241 |  1.44 |  99.54 | ****                                                                               |
|10| 0.0077 |  0.46 | 100.00 | **                                                                                 |
+--+--------+-------+--------+----------------------------------------------------------------------------------+
```

TABLE 25: Histogramme des 10 premières valeurs propres

- Le nombre de modalités est $m = 16$, ce qui conduit à $m - K = 10$ axes factoriels
- L'inertie totale est $\frac{m}{K} - 1 = 5/3 \approx 1.667$: le critère à la Kaiser, $\lambda_l > 1/K \approx 0.166$ nous dit de retenir 3 axes
- L'allure de la décroissance suggère de ne retenir que les deux premières

INDIVIDUS			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDENTIFICATEUR	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Beauceron	3.70	1.43	0.53	0.48	0.58	-0.56	0.33	2.1	2.2	5.7	7.0	2.8	0.19	0.16	0.24	0.22	0.08
Basset	3.70	1.91	-0.27	-1.09	0.24	-0.26	-0.54	0.6	11.4	1.0	1.5	7.2	0.04	0.62	0.03	0.03	0.15
Berger Allemand	3.70	1.43	0.53	0.48	0.58	-0.56	0.33	2.1	2.2	5.7	7.0	2.8	0.19	0.16	0.24	0.22	0.08
Boxer	3.70	1.83	-0.45	0.88	-0.66	-0.42	-0.39	1.6	7.4	7.2	3.9	3.7	0.11	0.42	0.24	0.10	0.08
Bull-Dog	3.70	1.67	-1.04	-0.53	0.16	0.29	0.39	8.2	2.7	0.4	1.9	3.7	0.64	0.17	0.02	0.05	0.09
Bull-Mastiff	3.70	1.99	0.74	-0.56	-0.42	-0.59	0.63	4.2	3.0	3.0	7.9	10.0	0.28	0.16	0.09	0.18	0.20
Caniche	3.70	2.05	-0.88	0.06	0.63	-0.33	0.26	5.9	0.0	6.6	2.5	1.7	0.38	0.00	0.19	0.05	0.03
Chihuahua	3.70	1.86	-0.85	-0.81	0.49	0.16	-0.21	5.4	6.4	4.0	0.5	1.1	0.39	0.36	0.13	0.01	0.02
Cocker	3.70	1.96	-0.75	-0.07	-0.56	-0.48	0.06	4.3	0.1	5.2	5.2	0.1	0.29	0.00	0.16	0.12	0.00
Colley	3.70	1.14	0.11	0.51	0.24	0.60	0.29	0.1	2.5	0.9	8.0	2.1	0.01	0.23	0.05	0.31	0.07
Dalmatien	3.70	1.80	-0.65	0.99	-0.52	0.18	-0.18	3.3	9.4	4.6	0.7	0.8	0.24	0.54	0.15	0.02	0.02
Dobermann	3.70	1.46	0.90	0.31	0.47	-0.37	0.21	6.1	0.9	3.7	3.1	1.1	0.55	0.07	0.15	0.09	0.03
Dogue Allemand	3.70	1.95	1.03	-0.54	-0.18	-0.08	-0.31	8.0	2.8	0.6	0.2	2.3	0.54	0.15	0.02	0.00	0.05
Epagneul Breton	3.70	2.07	-0.44	1.06	-0.04	-0.37	0.06	1.5	10.8	0.0	3.2	0.1	0.09	0.54	0.00	0.07	0.00
Epagneul Français	3.70	1.23	0.12	0.49	-0.29	0.62	-0.11	0.1	2.3	1.4	8.6	0.3	0.01	0.20	0.07	0.31	0.01
Fox-Hound	3.70	1.38	0.88	-0.05	0.32	0.05	-0.63	5.8	0.0	1.7	0.1	10.0	0.56	0.00	0.07	0.00	0.29
Fox-Terrier	3.70	1.81	-0.89	-0.12	0.02	-0.38	-0.19	6.1	0.1	0.0	3.2	0.9	0.44	0.01	0.00	0.08	0.02
Grand Bleu de Gascogne	3.70	1.44	0.51	0.10	-0.09	-0.12	-0.91	2.0	0.1	0.1	0.3	20.6	0.18	0.01	0.01	0.01	0.58
Labrador	3.70	1.80	-0.65	0.99	-0.52	0.18	-0.18	3.3	9.4	4.6	0.7	0.8	0.24	0.54	0.15	0.02	0.02
Lévrier	3.70	1.35	0.67	0.06	0.45	0.65	-0.43	3.4	0.0	3.4	9.4	4.6	0.34	0.00	0.15	0.31	0.14
Mastiff	3.70	1.90	0.72	-0.91	-0.57	-0.18	-0.21	3.9	8.0	5.5	0.7	1.1	0.27	0.44	0.17	0.02	0.02
Pékinois	3.70	1.86	-0.85	-0.81	0.49	0.16	-0.21	5.4	6.4	4.0	0.5	1.1	0.39	0.36	0.13	0.01	0.02
Pointer	3.70	1.43	0.70	0.42	0.60	0.23	0.41	3.7	1.7	6.0	1.2	4.2	0.34	0.12	0.25	0.04	0.12
Saint-Bernard	3.70	1.72	0.53	-0.63	-0.91	-0.04	0.39	2.1	3.8	13.7	0.0	3.7	0.16	0.23	0.48	0.00	0.09
Setter	3.70	1.16	0.48	0.35	0.12	0.78	0.16	1.8	1.2	0.2	13.8	0.7	0.20	0.10	0.01	0.53	0.02
Teckel	3.70	1.67	-1.04	-0.53	0.16	0.29	0.39	8.2	2.7	0.4	1.9	3.7	0.64	0.17	0.02	0.05	0.09
Terre-Neuve	3.70	1.69	0.33	-0.52	-0.77	0.56	0.59	0.8	2.6	10.0	7.0	8.6	0.06	0.16	0.35	0.18	0.20

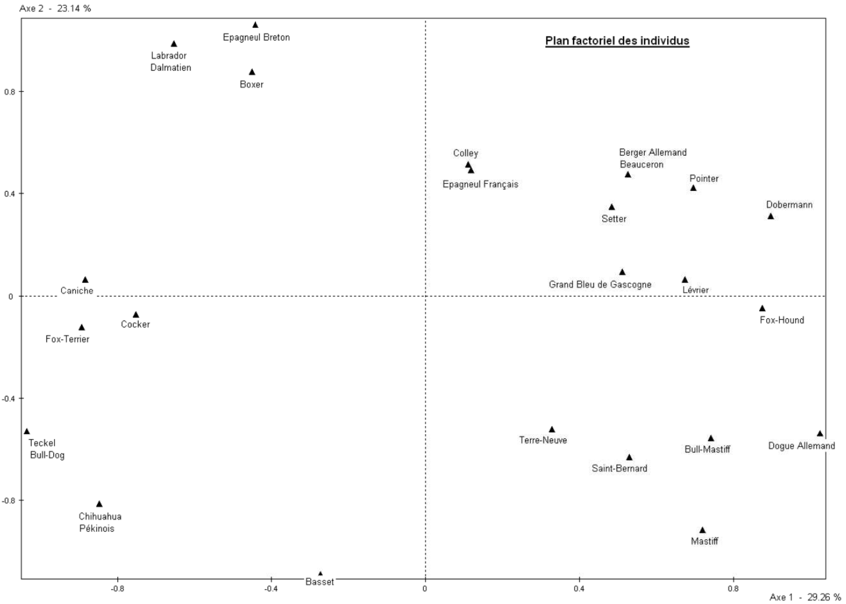
Table 26: Nuage des individus : Résultats pour les axes factoriels 1 à 5

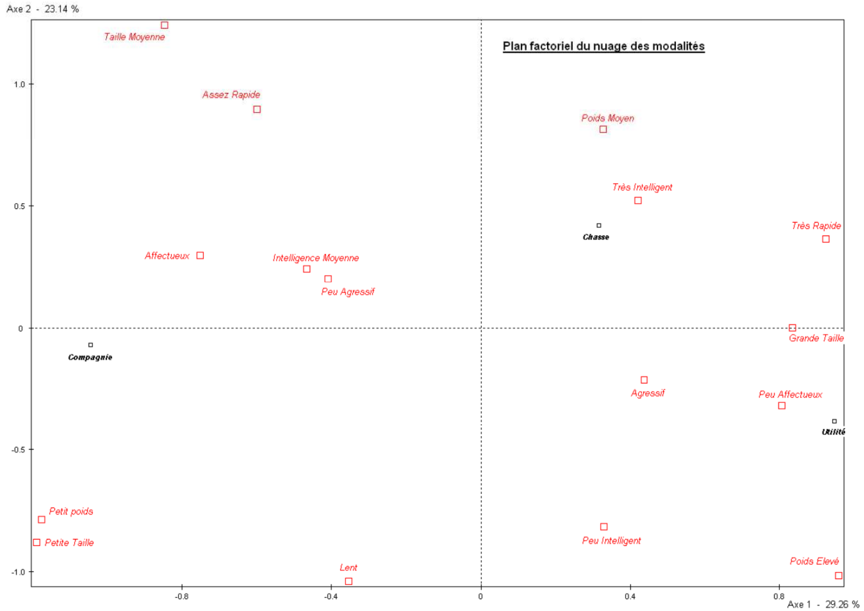
MODALITES			COORDONNEES					CONTRIBUTIONS					COSINUS CARRES				
IDEN - LIBELLE	P.REL	DISTO	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1 . Taille																	
TA01 - Petite Taille	4.32	2.86	-1.19	-0.88	0.67	-0.02	-0.04	12.6	8.7	8.7	0.0	0.1	0.50	0.27	0.16	0.00	0.00
TA02 - Taille Moyenne	3.09	4.40	-0.85	1.24	-0.98	-0.45	-0.33	4.5	12.2	13.5	3.8	2.2	0.16	0.35	0.22	0.05	0.02
TA03 - Grande Taille	9.26	0.80	0.84	0.00	0.02	0.16	0.13	13.3	0.0	0.0	1.5	1.0	0.88	0.00	0.00	0.03	0.02
							CONTRIBUTION CUMULEE =	30.4	20.9	22.2	5.3	3.3					
2 . Poids																	
PO01 - Petit poids	4.94	2.38	-1.18	-0.79	0.44	-0.17	-0.02	14.0	7.9	4.2	0.9	0.0	0.58	0.26	0.08	0.01	0.00
PO02 - Poids Moyen	8.64	0.93	0.33	0.81	0.19	0.16	-0.19	1.9	14.8	1.4	1.3	2.1	0.12	0.71	0.04	0.03	0.04
PO03 - poids Elevé	3.09	4.40	0.96	-1.02	-1.22	-0.17	0.57	5.8	8.3	20.7	0.5	6.6	0.21	0.24	0.34	0.01	0.07
							CONTRIBUTION CUMULEE =	21.8	31.0	26.3	2.7	8.8					
3 . Vélocité																	
VE01 - Lent	6.17	1.70	-0.35	-1.04	-0.36	-0.02	0.33	1.6	17.4	3.6	0.0	4.5	0.07	0.64	0.08	0.00	0.06
VE02 - Assez Rapide	4.94	2.38	-0.60	0.89	-0.39	-0.20	-0.53	3.6	10.2	3.5	1.2	9.4	0.15	0.34	0.07	0.02	0.12
VE03 - Très Rapide	5.56	2.00	0.93	0.36	0.75	0.20	0.11	9.8	1.9	14.1	1.4	0.4	0.43	0.07	0.28	0.02	0.01
							CONTRIBUTION CUMULEE =	15.0	29.5	21.2	2.6	14.3					
4 . Intelligence																	
IN01 - Peu Intelligent	4.94	2.38	0.33	-0.82	0.30	0.11	-1.12	1.1	8.5	2.1	0.4	41.6	0.05	0.28	0.04	0.01	0.53
IN02 - Intelligence Moyenne	7.41	1.25	-0.47	0.24	-0.63	0.45	0.26	3.3	1.1	13.2	9.1	3.4	0.17	0.05	0.31	0.16	0.05
IN03 - Très Intelligent	4.32	2.86	0.42	0.52	0.73	-0.90	0.83	1.6	3.0	10.4	21.2	20.1	0.06	0.09	0.19	0.28	0.24
							CONTRIBUTION CUMULEE =	6.0	12.7	25.6	30.7	65.1					
5 . Affection																	
AF01 - Peu Affectueux	8.02	1.08	0.81	-0.32	-0.17	0.24	-0.15	10.8	2.1	1.1	2.7	1.2	0.61	0.09	0.03	0.05	0.02
AF02 - Affectueux	8.64	0.93	-0.75	0.29	0.16	-0.22	0.14	10.0	1.9	1.0	2.5	1.1	0.61	0.09	0.03	0.05	0.02
							CONTRIBUTION CUMULEE =	20.7	4.0	2.1	5.3	2.3					
6 . Agressivité																	
AG01 - Peu Agressif	8.64	0.93	-0.41	0.20	0.18	0.70	0.23	3.0	0.9	1.3	25.8	2.9	0.18	0.04	0.03	0.53	0.05
AG02 - Agressif	8.02	1.08	0.44	-0.21	-0.19	-0.75	-0.24	3.2	1.0	1.4	27.7	3.2	0.18	0.04	0.03	0.53	0.05
							CONTRIBUTION CUMULEE =	6.1	1.9	2.6	53.5	6.1					

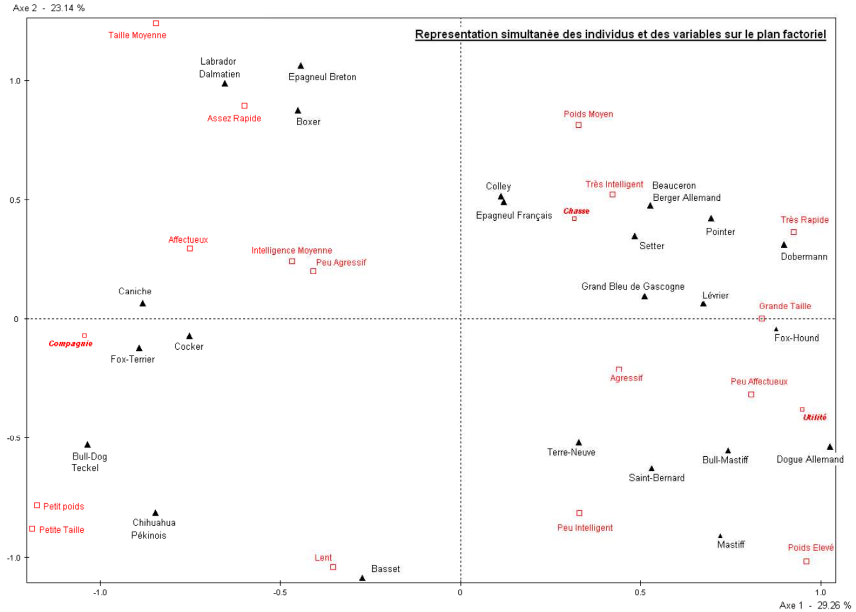
Table 27: Nuage des modalités actives : Résultats sur les axes factoriels 1 à 5

MODALITES			VALEURS-TEST					COORDONNEES					
IDEN - LIBELLE	EFF.	P.ABS	1	2	3	4	5	1	2	3	4	5	DISTO.
1 . Taille													
TA1 - Petite Taille	7	7.00	-3.6	-2.7	2.0	-0.1	-0.1	-1.19	-0.88	0.67	-0.02	-0.04	2.86
TA2 - Taille Moyenne	5	5.00	-2.1	3.0	-2.4	-1.1	-0.8	-0.85	1.24	-0.98	-0.45	-0.33	4.40
TA3 - Grande Taille	15	15.00	4.8	0.0	0.1	0.9	0.7	0.84	0.00	0.02	0.16	0.13	0.80
2 . Poids													
PO1 - Petit poids	8	8.00	-3.9	-2.6	1.4	-0.6	-0.1	-1.18	-0.79	0.44	-0.17	-0.02	2.38
PO2 - Poids Moyen	14	14.00	1.7	4.3	1.0	0.8	-1.0	0.33	0.81	0.19	0.16	-0.19	0.93
PO3 - poids Elevé	5	5.00	2.3	-2.5	-3.0	-0.4	1.4	0.96	-1.02	-1.22	-0.17	0.57	4.40
3 . Vélocité													
VE1 - Lent	10	10.00	-1.4	-4.1	-1.4	-0.1	1.3	-0.35	-1.04	-0.36	-0.02	0.33	1.70
VE2 - Assez Rapide	8	8.00	-2.0	3.0	-1.3	-0.7	-1.8	-0.60	0.89	-0.39	-0.20	-0.53	2.38
VE3 - Très Rapide	9	9.00	3.3	1.3	2.7	0.7	0.4	0.93	0.36	0.75	0.20	0.11	2.00
4 . Intelligence													
IN1 - Peu Intelligent	8	8.00	1.1	-2.7	1.0	0.4	-3.7	0.33	-0.82	0.30	0.11	-1.12	2.38
IN2 - Intelligence Moyenne	12	12.00	-2.1	1.1	-2.9	2.0	1.2	-0.47	0.24	-0.63	0.45	0.26	1.25
IN3 - Très Intelligent	7	7.00	1.3	1.6	2.2	-2.7	2.5	0.42	0.52	0.73	-0.90	0.83	2.86
5 . Affection													
AF1 - Peu Affectueux	13	13.00	4.0	-1.6	-0.8	1.2	-0.7	0.81	-0.32	-0.17	0.24	-0.15	1.08
AF2 - Affectueux	14	14.00	-4.0	1.6	0.8	-1.2	0.7	-0.75	0.29	0.16	-0.22	0.14	0.93
6 . Agressivité													
AG1 - Peu Agressif	14	14.00	-2.2	1.1	0.9	3.7	1.2	-0.41	0.20	0.18	0.70	0.23	0.93
AG2 - Agressif	13	13.00	2.2	-1.1	-0.9	-3.7	-1.2	0.44	-0.21	-0.19	-0.75	-0.24	1.08
7 . Fonction													
FO1 - Compagnie	10	10.00	-4.1	-0.3	0.4	0.1	0.2	-1.04	-0.07	0.10	0.02	0.05	1.70
FO2 - Chasse	9	9.00	1.1	1.5	0.7	1.7	-2.3	0.32	0.42	0.18	0.48	-0.63	2.00
FO3 - Utilité	8	8.00	3.1	-1.3	-1.1	-1.9	2.1	0.95	-0.38	-0.33	-0.56	0.64	2.38

Table 28: Nuage des modalités : Coordonnées et valeurs-tests pour les axes factoriels 1 à 5







Le bilan

- L'axe 1 oppose (à gauche) les chiens de petite taille, affectueux, qui coïncident avec les chiens dits de « compagnie », aux chiens de grande taille, très rapides et agressifs qui coïncident avec la fonction « utilité »
- L'axe 2 oppose (en haut) les chiens de chasse, de taille moyenne, très intelligents, à des chiens lents et peu intelligents.