

Analyse factorielle des correspondances multiples avec R

Ricco.Rakotomalala
<http://eric.univ-lyon2.fr/~ricco/cours>

Références :

1. M. Tenehaus, « Statistique – Méthodes pour décrire, expliquer et prévoir », Dunod, 2006 ; pages 253 à 264, partie théorique ; pages 266 à 276, pour l'exemple que nous traiterons.
2. Tutoriels Tanagra, « AFCM – Races canines », <http://tutoriels-data-mining.blogspot.com/2008/03/afcm-races-canines.html> ; description des mêmes calculs sous le logiciel Tanagra. Les résultats sont complètement identiques.
3. Husson, Le, Josse, Mazet, « FactoMineR », <http://factominer.free.fr/> ; package que nous privilégierons pour l'ACM (méthode MCA).
4. D'autres packages réalisent l'ACM (procédure mca avec MASS, dudi.acm avec ADE4)

Objectif de l'étude

L'exemple des races canines

Objectifs de l'étude

Ce tutoriel reproduit sous le logiciel R, l'analyse menée dans l'ouvrage de Tenenhaus, pages 266 à 276. Les justifications théoriques et les formules sont disponibles dans le même ouvrage, pages 253 à 264.

Les mêmes calculs ont été reproduits dans Tanagra. D'autres packages de R peuvent réaliser également une AFCM (ou ACM – Analyse des correspondances multiples), nous avons choisi FactorMineR pour sa simplicité et son adéquation avec les sorties usuelles des logiciels qui font référence (et que l'on retrouve dans notre ouvrage ci-dessus).

Traitements réalisés

- Réaliser une AFCM (ACM) sur un fichier de données.
- Afficher les valeurs propres. Construire le graphiques éboulis des valeurs propres.
- Calculer les coordonnées factorielles, les \cos^2 et les contributions des individus
- Calculer les coordonnées factorielles, les \cos^2 et les contributions des points modalités
- Projeter les points dans le premier plan factoriel (représentation pseudo-barycentrique)
- Projeter les points dans le premier plan factoriel (représentation barycentrique)
- Positionner les modalités d'une variable illustrative qualitative
- Commenter les autres sorties/graphiques produits par la procédure MCA de FactoMineR

Tableau de données

Chien	Taille	Poids	Velocite	Intelligence	Affection	Agressivite	Fonction
Beauceron	Taille++	Poids+	Veloc++	Intell+	Affec+	Agress+	utilite
Basset	Taille-	Poids-	Veloc-	Intell-	Affec-	Agress+	chasse
Berger All	Taille++	Poids+	Veloc++	Intell++	Affec+	Agress+	utilite
Boxer	Taille+	Poids+	Veloc+	Intell+	Affec+	Agress+	compagnie
Bull-Dog	Taille-	Poids-	Veloc-	Intell+	Affec+	Agress-	compagnie
Bull-Mastif	Taille++	Poids++	Veloc-	Intell++	Affec-	Agress+	utilite
Caniche	Taille-	Poids-	Veloc+	Intell++	Affec+	Agress-	compagnie
Chihuahua	Taille-	Poids-	Veloc-	Intell-	Affec+	Agress-	compagnie
Cocker	Taille+	Poids-	Veloc-	Intell+	Affec+	Agress+	compagnie
Colley	Taille++	Poids+	Veloc++	Intell+	Affec+	Agress-	compagnie
Dalmatien	Taille+	Poids+	Veloc+	Intell+	Affec+	Agress-	compagnie
Doberman	Taille++	Poids+	Veloc++	Intell++	Affec-	Agress+	utilite
Dogue All	Taille++	Poids++	Veloc++	Intell-	Affec-	Agress+	utilite
Epag. Breton	Taille+	Poids+	Veloc+	Intell++	Affec+	Agress-	chasse
Epag. Français	Taille++	Poids+	Veloc+	Intell+	Affec-	Agress-	chasse
Fox-Hound	Taille++	Poids+	Veloc++	Intell-	Affec-	Agress+	chasse
Fox-Terrier	Taille-	Poids-	Veloc+	Intell+	Affec+	Agress+	compagnie
Gd Bleu Gasc	Taille++	Poids+	Veloc+	Intell-	Affec-	Agress+	chasse
Labrador	Taille+	Poids+	Veloc+	Intell+	Affec+	Agress-	chasse
Levrier	Taille++	Poids+	Veloc++	Intell-	Affec-	Agress-	chasse
Mastiff	Taille++	Poids++	Veloc-	Intell-	Affec-	Agress+	utilite
Pekinois	Taille-	Poids-	Veloc-	Intell-	Affec+	Agress-	compagnie
Pointer	Taille++	Poids+	Veloc++	Intell++	Affec-	Agress-	chasse
St-Bernard	Taille++	Poids++	Veloc-	Intell+	Affec-	Agress+	utilite
Setter	Taille++	Poids+	Veloc++	Intell+	Affec-	Agress-	chasse
Teckel	Taille-	Poids-	Veloc-	Intell+	Affec+	Agress-	compagnie
Terre-Neuve	Taille++	Poids++	Veloc-	Intell+	Affec-	Agress-	utilite

Label des observations

Variables actives

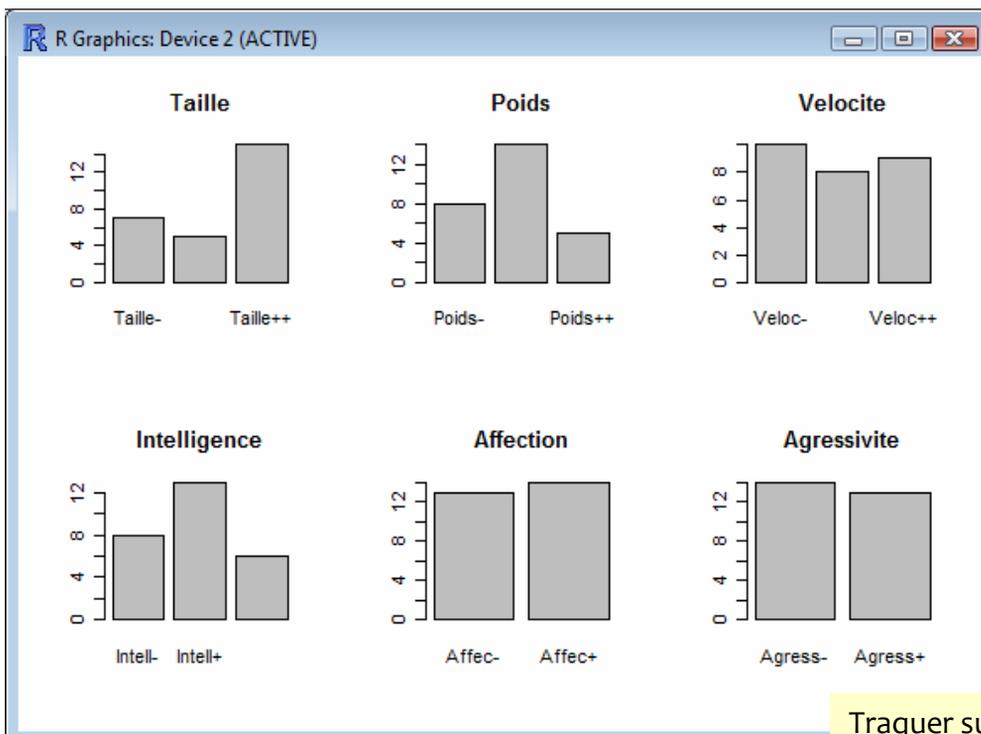
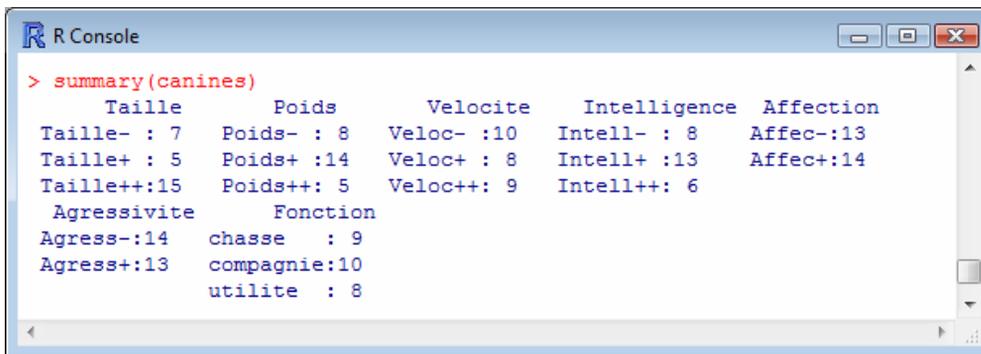
Variable illustrative qualitative

Fichier de données

Importation, statistiques descriptives et graphiques

```
#changement de répertoire
setwd("D:/_Travaux/university/Cours_Universite/Supports_de_cours/Info
rmatique/R/Tutoriels/afcm")

#chargement des données
library(xlsReadWrite)
canines <- read.xls(file="races_canines_acm.xls",rowNames=T,sheet=1)
#statistiques descriptives
summary(canines)
#distribution de fréquences (graphique) des variables actives
par(mfrow=c(2,3))
for (j in 1:6){plot(canines[,j],main=colnames(canines)[j])}
layout(1)
```



Traquer surtout les modalités très rares, elles peuvent fausser les calculs.

Analyse des correspondances multiples

La procédure MCA de FactoMineR

```
#charger le package
library(FactoMineR)
#lancer l'ACM
#voir l'aide pour plus de détails sur les options
#notamment -> ind. supplémentaires, var. quanti. supplémentaires
canines.acm <- MCA(canines,ncp=2,quali.sup=c(7),graph=F)
print(canines.acm)
#les valeurs propres
print(canines.acm$eig)
#scree plot
plot(canines.acm$eig[,1],type="b",main="Scree plot")
```

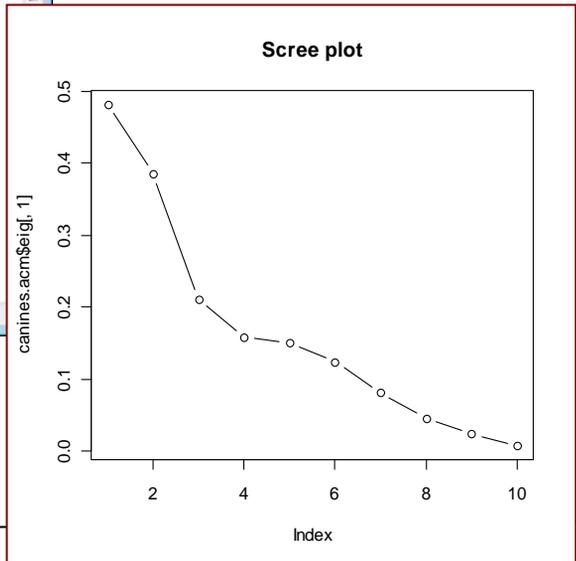
```
R Console
> print(canines.acm)
**Results of the Multiple Correspondence Analysis (MCA)**
The analysis was performed on 27 individuals, described by 7 variables
*The results are available in the following objects:

  name                description
1  "$eig"              "eigenvalues"
2  "$svar"             "results for the variables"
3  "$svar$coord"      "coord. of the categories"
4  "$svar$cos2"       "cos2 for the categories"
5  "$svar$contrib"    "contributions of the categories"
6  "$svar$v.test"     "v-test for the categories"
7  "$ind"              "results for the individuals"
8  "$ind$coord"       "coord. for the individuals"
9  "$ind$cos2"        "cos2 for the individuals"
10 "$ind$contrib"     "contributions of the individuals"
11 "$quali.sup"       "results for the supplementary qualitative vari
12 "$quali.sup$coord" "coord. for the supplementary categories"
13 "$quali.sup$cos2" "cos2 for the supplementary categories"
14 "$quali.sup$v.test" "v-test for the supplementary categories"
15 "$scall"           "intermediate results"
16 "$scall$marge.col" "weights of columns"
17 "$scall$marge.li"  "weights of rows"
> |
```

Le seul intérêt du rapport est de (1) montrer que les calculs ont été correctement réalisés ; (2) fournir la liste des champs que nous pourrions exploiter pour accéder aux résultats détaillés

```
R Console
> print(canines.acm$eig)
  eigenvalue percentage of variance cumulative percentage of variance
dim 1  0.481606165          28.8963699          28.89637
dim 2  0.384737288          23.0842373          51.98061
dim 3  0.210954049          12.6572430          64.63785
dim 4  0.157554025           9.4532415          74.09109
dim 5  0.150132670           9.0079602          83.09905
dim 6  0.123295308           7.3977185          90.49677
dim 7  0.081462460           4.8877476          95.38452
dim 8  0.045669757           2.7401854          98.12470
dim 9  0.023541911           1.4125146          99.53722
dim 10 0.007713034           0.4627820          100.00000
```

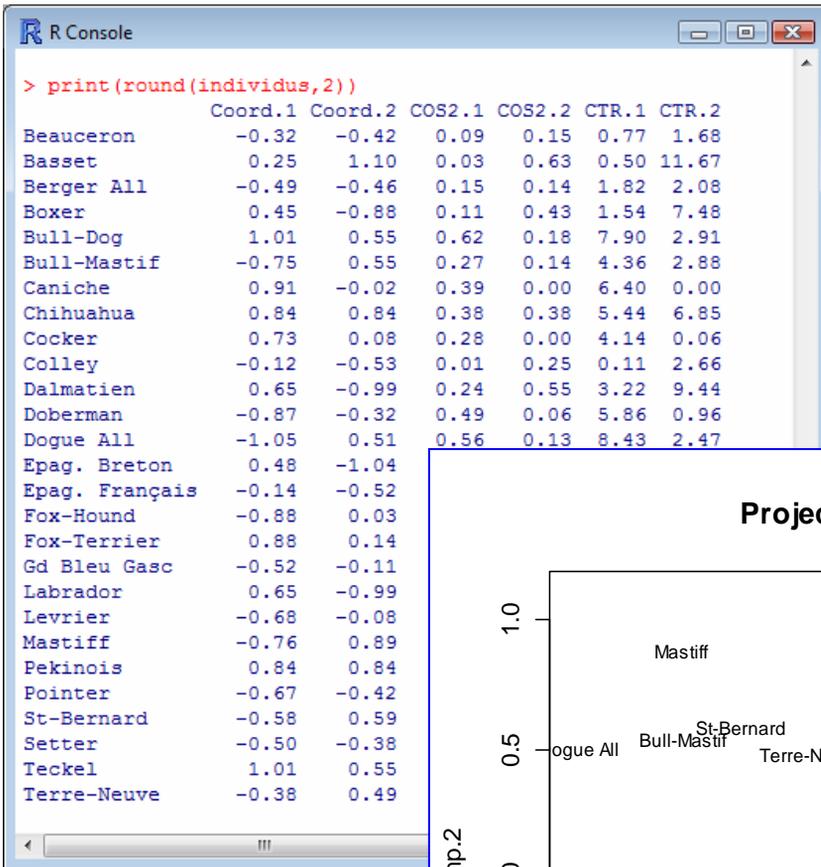
Tableau et graphique pour une meilleur appréciation



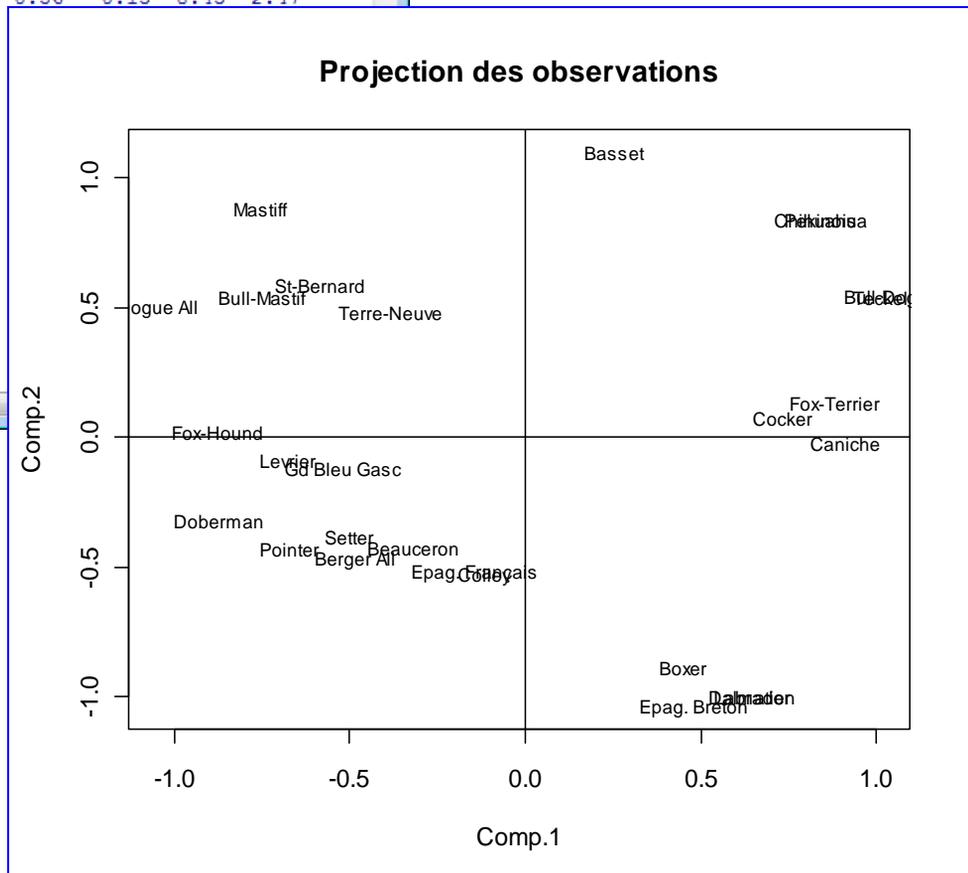
Analyse des correspondances multiples

Informations sur les individus

```
#coordonnées, cos2 et contribution des individus
individus <-
cbind(canines.acm$ind$coord,canines.acm$ind$cos2,canines.acm$ind$contrib)
colnames(individus) <- c("Coord.1","Coord.2","COS2.1","COS2.2","CTR.1","CTR.2")
print(round(individus,2))
#graphique associé
plot(individus[,1],individus[,2],main="Projection des
observations",xlab="Comp.1",ylab="Comp.2",type="n")
abline(h=0,v=0)
text(individus[,1],individus[,2],labels=rownames(individus),cex=0.75)
```



```
R Console
> print(round(individus,2))
      Coord.1 Coord.2 COS2.1 COS2.2 CTR.1 CTR.2
Beauceron  -0.32 -0.42  0.09  0.15  0.77  1.68
Basset      0.25  1.10  0.03  0.63  0.50 11.67
Berger All  -0.49 -0.46  0.15  0.14  1.82  2.08
Boxer       0.45 -0.88  0.11  0.43  1.54  7.48
Bull-Dog    1.01  0.55  0.62  0.18  7.90  2.91
Bull-Mastif -0.75  0.55  0.27  0.14  4.36  2.88
Caniche     0.91 -0.02  0.39  0.00  6.40  0.00
Chihuahua   0.84  0.84  0.38  0.38  5.44  6.85
Cocker      0.73  0.08  0.28  0.00  4.14  0.06
Colley      -0.12 -0.53  0.01  0.25  0.11  2.66
Dalmatien   0.65 -0.99  0.24  0.55  3.22  9.44
Doberman    -0.87 -0.32  0.49  0.06  5.86  0.96
Dogue All   -1.05  0.51  0.56  0.13  8.43  2.47
Epag. Breton 0.48 -1.04
Epag. Français -0.14 -0.52
Fox-Hound   -0.88  0.03
Fox-Terrier  0.88  0.14
Gd Bleu Gasc -0.52 -0.11
Labrador    0.65 -0.99
Levrier     -0.68 -0.08
Mastiff     -0.76  0.89
Pekinois    0.84  0.84
Pointer     -0.67 -0.42
St-Bernard  -0.58  0.59
Setter      -0.50 -0.38
Teckel      1.01  0.55
Terre-Neuve -0.38  0.49
```

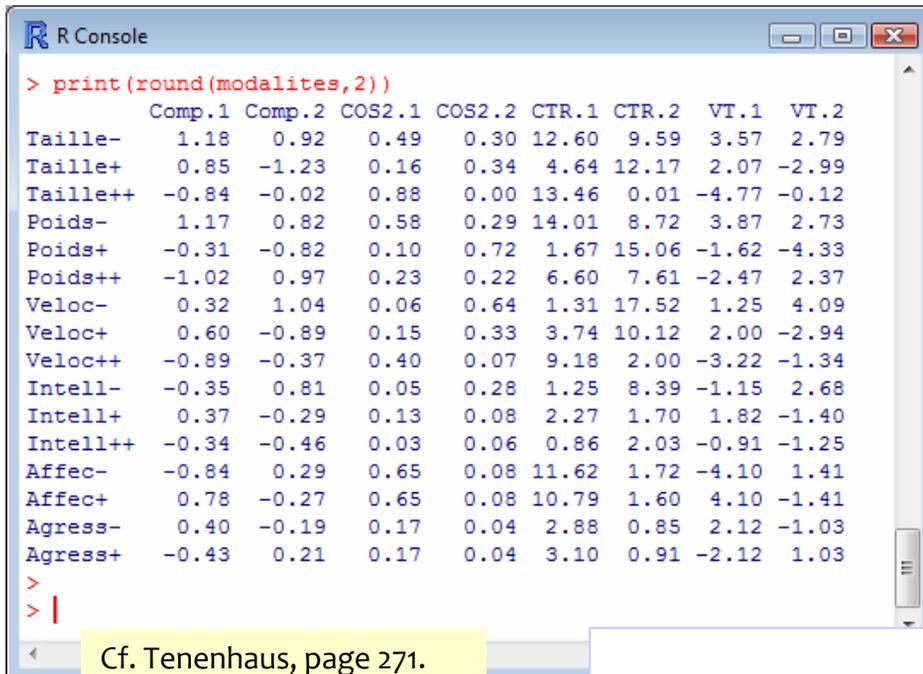


Cf. Tenenhaus, page 268.

Analyse des correspondances multiples

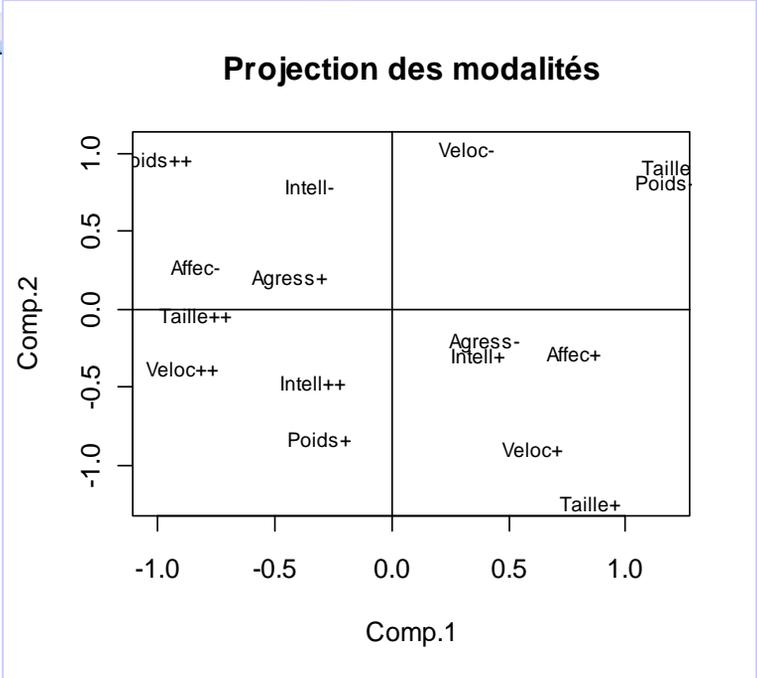
Informations sur les « points » modalités

```
#coordonnées, cos2 et contribution des modalités
modalites <-
cbind(canines.acm$var$coord,canines.acm$var$cos2,canines.acm$var$contrib,canines.acm$var$v.test)
colnames(modalites) <- c("Comp.1","Comp.2","COS2.1","COS2.2","CTR.1","CTR.2","VT.1","VT.2")
print(round(modalites,2))
#graphique associé
plot(modalites[,1],modalites[,2],main="Projection des
modalités",xlab="Comp.1",ylab="Comp.2",type="n")
abline(h=0,v=0)
text(modalites[,1],modalites[,2],labels=rownames(modalites),cex=0.75)
```



```
> print(round(modalites,2))
      Comp.1 Comp.2 COS2.1 COS2.2 CTR.1 CTR.2 VT.1 VT.2
Taille-  1.18  0.92  0.49  0.30 12.60  9.59  3.57  2.79
Taille+  0.85 -1.23  0.16  0.34  4.64 12.17  2.07 -2.99
Taille++ -0.84 -0.02  0.88  0.00 13.46  0.01 -4.77 -0.12
Poids-   1.17  0.82  0.58  0.29 14.01  8.72  3.87  2.73
Poids+  -0.31 -0.82  0.10  0.72  1.67 15.06 -1.62 -4.33
Poids++ -1.02  0.97  0.23  0.22  6.60  7.61 -2.47  2.37
Veloc-   0.32  1.04  0.06  0.64  1.31 17.52  1.25  4.09
Veloc+   0.60 -0.89  0.15  0.33  3.74 10.12  2.00 -2.94
Veloc++ -0.89 -0.37  0.40  0.07  9.18  2.00 -3.22 -1.34
Intell-  -0.35  0.81  0.05  0.28  1.25  8.39 -1.15  2.68
Intell+  0.37 -0.29  0.13  0.08  2.27  1.70  1.82 -1.40
Intell++ -0.34 -0.46  0.03  0.06  0.86  2.03 -0.91 -1.25
Affec-   -0.84  0.29  0.65  0.08 11.62  1.72 -4.10  1.41
Affec+   0.78 -0.27  0.65  0.08 10.79  1.60  4.10 -1.41
Agress-  0.40 -0.19  0.17  0.04  2.88  0.85  2.12 -1.03
Agress+ -0.43  0.21  0.17  0.04  3.10  0.91 -2.12  1.03
>
> |
```

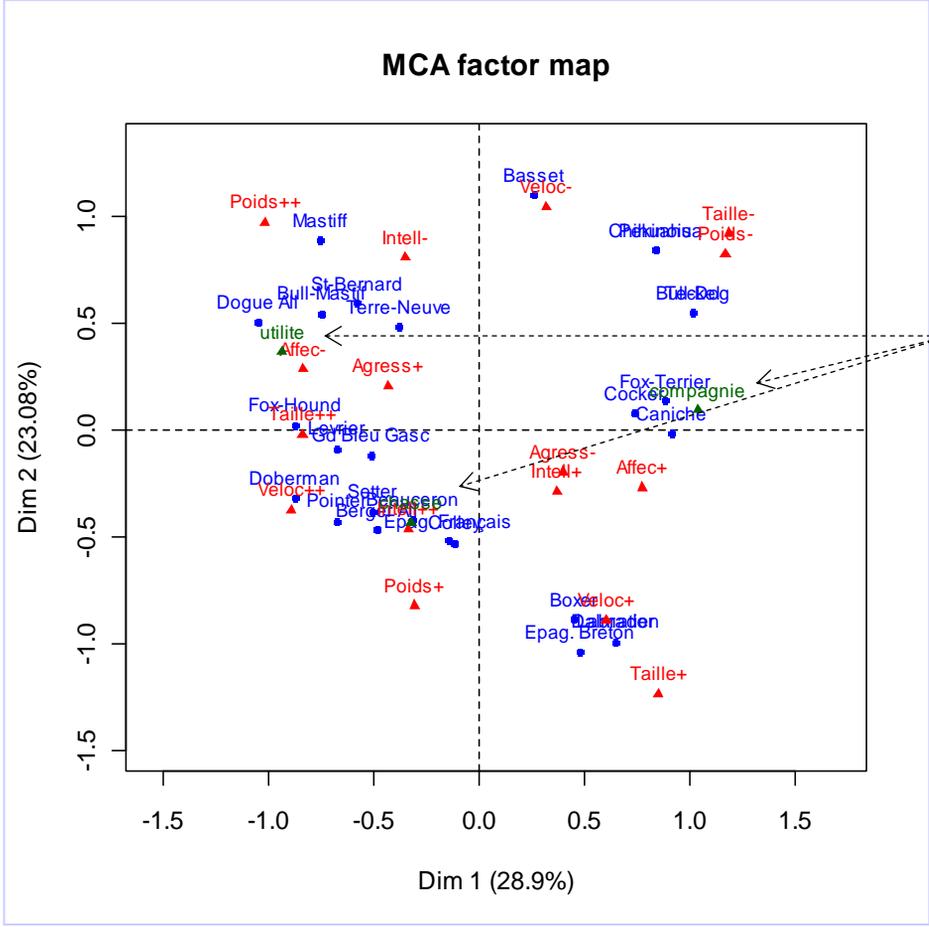
Cf. Tenenhaus, page 271.
Le carré de la corrélation correspond au COS².
La dernière colonne est la valeur-test (page 261)



Analyse des correspondances multiples

Les graphiques proposés automatiquement par la procédure MCA

```
#graphiques automatiques de MCA  
plot(canines.acm,cex=0.75,choix="ind")  
plot(canines.acm,cex=0.75,choix="var")
```

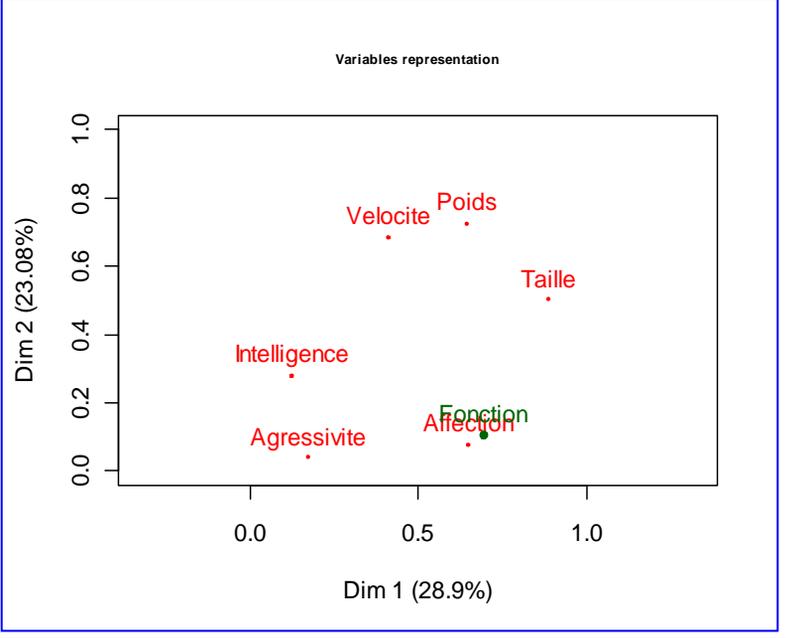


Noter le positionnement des modalités de la variable supplémentaire en noir.

Nous avons la représentation pseudo-barycentrique dans ce cas, Fig.8.1, page 270.

Il s'agit des rapports de corrélation des variables avec les facteurs, y compris la variable supplémentaire FONCTION.

Cf. formules page 260 ; valeurs dans le tableau page 271.



Analyse des correspondances multiples

Représentation barycentrique de type (a)

c.-à-d. chaque point modalité est au barycentre des individus possédant cette modalité

```
#composantes principales réduites pour les individus
fletoile <- canines.acm$ind$coord[,1]/sqrt(canines.acm$eig[1,1])
f2etoile <- canines.acm$ind$coord[,2]/sqrt(canines.acm$eig[2,1])
#composantes principales non modifiées pour les modalités
g1 <- canines.acm$var$coord[,1]
g2 <- canines.acm$var$coord[,2]
#graphique
plot(c(fletoile,g1),c(f2etoile,g2),type="n",main="Représentation barycentrique de type (a)")
abline(h=0,v=0)
text(fletoile,f2etoile,labels=names(fletoile),cex=0.5)
text(g1,g2,labels=names(g1),cex=0.5,col="red")
```

Représentation barycentrique de type (a)

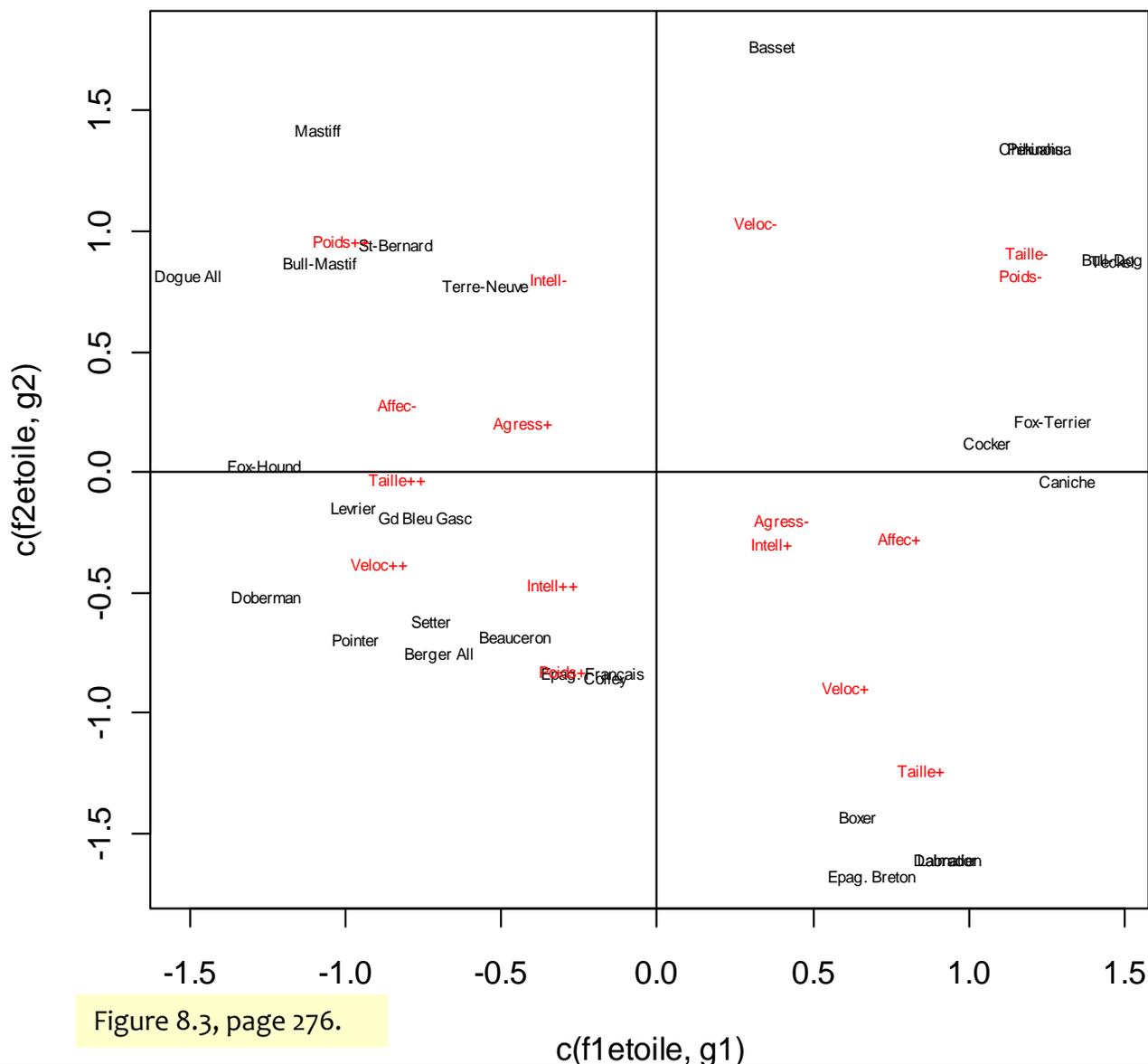


Figure 8.3, page 276.

***Et on peut faire bien
d'autres choses encore...***