

Réalisation d'une ACM

Rémi Taniel

30/09/2019

Contents

Introduction	2
1ère partie : Importation des données	2
2e partie : Mise en oeuvre de l'ACM	2
3e partie : Analyse des résultats	3
Nombre d'axe à retenir	3
Analyse des quatres premiers axes en fonction des modalités	4
Analyse des quatres premiers axes en fonction des individus	6
Interprétation du premier plan factoriel	9

Introduction

1ère partie : Importation des données

On commence par importer les données grâce à :

```
data <- read.table("/home/remi/Documents/Cours/AD/data/race.csv", sep = ",", dec = ".", colClasses = "f
```

Puis on visualise les données grâce à la fonction `str(...)` :

```
str(data)

## 'data.frame': 27 obs. of 8 variables:
## $ Race : Factor w/ 27 levels "B.AL","BASS",...: 3 2 1 4 5 6 7 8 9 10 ...
## $ Taille : Factor w/ 3 levels "1","2","3": 3 1 3 2 1 3 1 1 2 3 ...
## $ Poids : Factor w/ 3 levels "1","2","3": 2 1 2 2 1 3 1 1 1 2 ...
## $ Velocite : Factor w/ 3 levels "1","2","3": 3 1 3 2 1 1 2 1 1 3 ...
## $ Intelligence: Factor w/ 3 levels "1","2","3": 3 1 3 2 2 3 3 1 2 2 ...
## $ Affection : Factor w/ 2 levels "1","2": 2 1 2 2 2 1 2 2 2 2 ...
## $ Agressivite : Factor w/ 2 levels "1","2": 2 2 2 2 1 2 1 1 2 1 ...
## $ Fonction : Factor w/ 3 levels "1","2","3": 3 2 3 1 1 3 1 1 1 1 ...
```

Notre jeu de données comporte 27 observations (27 races de chiens) et 8 variables (les 7 modalités et un libellé), on remarque également que toutes les variables sont des variables qualitatives, et ont 3 niveaux chacune :

- Faible : 1
- Moyen : 2
- Fort : 3

On se décide de visualiser les 6 premières lignes de nos données ;

```
head(data)

## Race Taille Poids Velocite Intelligence Affection Agressivite Fonction
## 1 BEAU 3 2 3 3 2 2 3
## 2 BASS 1 1 1 1 1 2 2
## 3 B.AL 3 2 3 3 2 2 3
## 4 BOXE 2 2 2 2 2 2 1
## 5 BULD 1 1 1 2 2 1 1
## 6 BULM 3 3 1 3 1 2 3
```

On remarque que nous devons enlever la colonne `race`, on décide donc de formater nos données pour donner un identifiant aux différentes lignes, dans notre cas, ce sera la variable `race` :

```
rownames(data) <- data$Race
data <- data[,-1]
head(data)

## Taille Poids Velocite Intelligence Affection Agressivite Fonction
## BEAU 3 2 3 3 2 2 3
## BASS 1 1 1 1 1 2 2
## B.AL 3 2 3 3 2 2 3
## BOXE 2 2 2 2 2 2 1
## BULD 1 1 1 2 2 1 1
## BULM 3 3 1 3 1 2 3
```

2e partie : Mise en oeuvre de l'ACM

Pour réaliser l'ACM, nous aurons besoin du package `FactoMineR` :

```
library(FactoMineR)
```

Puis on range les résultats de l'ACM (valeurs propres, coordonnées, contribution) dans la variable `data.mca`, dans notre cas, nous ne retenons que les 4 premiers axes et on souhaite que les graphiques ne soient pas générés lors de l'appel de la fonction :

```
data.mca <- FactoMineR::MCA(data, ncp = 4, quali.sup = c(7), graph = FALSE)
```

(Explication de où se trouve les variables ?)

3e partie : Analyse des résultats

Nombre d'axe à retenir

Pour connaître le nombre d'axe que nous devons retenir, nous pouvons utiliser 3 critères :

- Part d'inertie supérieure à la moyenne
- Part d'inertie cumulée supérieure à 80%
- Critère du coude

Pour rappel, les valeurs propres des différents axes sont stockés dans :

```
data.mca$eig
```

```
##          eigenvalue percentage of variance
## dim 1  0.487590671          29.2554403
## dim 2  0.385667752          23.1400651
## dim 3  0.220728360          13.2437016
## dim 4  0.164472270           9.8683362
## dim 5  0.148659735           8.9195841
## dim 6  0.101816275           6.1089765
## dim 7  0.081305114           4.8783069
## dim 8  0.044665793           2.6799476
## dim 9  0.024078978           1.4447387
## dim 10 0.007681718           0.4609031
##          cumulative percentage of variance
## dim 1          29.25544
## dim 2          52.39551
## dim 3          65.63921
## dim 4          75.50754
## dim 5          84.42713
## dim 6          90.53610
## dim 7          95.41441
## dim 8          98.09436
## dim 9          99.53910
## dim 10        100.00000
```

Nous n'utiliserons qu'un seul critère, le critère de la part d'inertie moyenne, cette moyenne des part d'inertie expliquée par chaque axe peut être obtenue par :

```
mean(data.mca$eig[,2])
```

```
## [1] 10
```

Selon ce critère, nous pouvons retenir les 4 premiers axes, qui possèdent tous une part d'inertie supérieure à la moyenne calculée qui est de 10.

Analyse des quatres premiers axes en fonction des modalités

Pour obtenir les données des 4 premiers axes en fonction des modalités, on utilise l'information suivante :

```
data.mca$var
```

```
## $coord
##           Dim 1      Dim 2      Dim 3      Dim 4
## Taille_1      1.1902124  0.8814312493 -0.66619486  0.0245082
## Taille_2      0.8464733 -1.2369449391  0.98137849  0.4516478
## Taille_3     -0.8375902  0.0009803967 -0.01623523 -0.1619864
## Poids_1       1.1761704  0.7857960033 -0.43503416  0.1695709
## Poids_2     -0.3293556 -0.8124958119 -0.18601927 -0.1564473
## Poids_3     -0.9596770  1.0177146680  1.21690862  0.1667392
## Velocite_1    0.3536067  1.0417311970  0.35916042  0.0229601
## Velocite_2    0.5993249 -0.8943513681  0.39416354  0.1983599
## Velocite_3   -0.9256296 -0.3625001139 -0.74943473 -0.2018311
## Intelligence_1 -0.3298252  0.8162016046 -0.30430278 -0.1139304
## Intelligence_2  0.4660516 -0.2414724706  0.62739888 -0.4483245
## Intelligence_3 -0.4220026 -0.5188490270 -0.72776633  0.8987625
## Affection_1   -0.8082282  0.3173466033  0.17128331 -0.2363086
## Affection_2    0.7504976 -0.2946789888 -0.15904879  0.2194294
## Agressivite_1  0.4080360 -0.1994591783 -0.17920983 -0.7002575
## Agressivite_2 -0.4394234  0.2148021920  0.19299520  0.7541235
##
## $contrib
##           Dim 1      Dim 2      Dim 3      Dim 4
## Taille_1     12.553839  8.704555e+00  8.68815258  0.01578023
## Taille_2      4.535510  1.224454e+01  13.46697104  3.82791292
## Taille_3     13.322448  2.307632e-05  0.01105694  1.47720499
## Poids_1      14.010709  7.906445e+00  4.23412382  0.86334566
## Poids_2       1.922595  1.479251e+01  1.35478624  1.28604643
## Poids_3       5.829749  8.288840e+00  20.70679054  0.52172002
## Velocite_1    1.582963  1.736933e+01  3.60747890  0.01978518
## Velocite_2    3.637845  1.024184e+01  3.47591976  1.18138110
## Velocite_3    9.762175  1.892912e+00  14.13634026  1.37597681
## Intelligence_1  1.101761  8.530147e+00  2.07170863  0.38972813
## Intelligence_2  3.299734  1.119923e+00  13.20977953  9.05229126
## Intelligence_3  1.578185  3.016142e+00  10.36832982  21.22170349
## Affection_1    10.750805  2.095476e+00  1.06659687  2.72454975
## Affection_2     9.982890  1.945799e+00  0.99041137  2.52993905
## Agressivite_1   2.950901  8.914721e-01  1.25741474  25.76534276
## Agressivite_2   3.177893  9.600469e-01  1.35413895  27.74729221
##
## $cos2
##           Dim 1      Dim 2      Dim 3      Dim 4
## Taille_1      0.49581192  2.719224e-01  0.1553354552  0.0002102282
## Taille_2      0.16284478  3.477347e-01  0.2188872152  0.0463603889
## Taille_3      0.87694667  1.201472e-06  0.0003294784  0.0327995000
## Poids_1       0.58247448  2.599896e-01  0.0796861975  0.0121070633
## Poids_2       0.11681934  7.109302e-01  0.0372649528  0.0263585186
## Poids_3       0.20931365  2.353962e-01  0.3365605898  0.0063186241
## Velocite_1    0.07355161  6.383552e-01  0.0758801234  0.0003100978
## Velocite_2    0.15123804  3.367850e-01  0.0654167993  0.0165670095
## Velocite_3    0.42839511  6.570317e-02  0.2808262078  0.0203679027
```

```
## Intelligence_1 0.04580407 2.804990e-01 0.0389895501 0.0054653233
## Intelligence_2 0.17376329 4.664716e-02 0.3149034805 0.1607958967
## Intelligence_3 0.06233016 9.422151e-02 0.1853753391 0.2827209221
## Affection_1 0.60657339 9.351538e-02 0.0272424033 0.0518530622
## Affection_2 0.60657339 9.351538e-02 0.0272424033 0.0518530622
## Agressivite_1 0.17930056 4.284427e-02 0.0345866377 0.5280806741
## Agressivite_2 0.17930056 4.284427e-02 0.0345866377 0.5280806741
##
## $v.test
##           Dim 1           Dim 2           Dim 3           Dim 4
## Taille_1      3.590419    2.658943687 -2.00965714    0.07393195
## Taille_2      2.057660 -3.006842664    2.38559586    1.09789349
## Taille_3     -4.774999    0.005589121 -0.09255506   -0.92346467
## Poids_1       3.891572    2.599948123 -1.43938915    0.56105583
## Poids_2      -1.742786 -4.299323719 -0.98432148   -0.82784146
## Poids_3      -2.332843    2.473924090    2.95813714    0.40531991
## Velocite_1     1.382874    4.073970535    1.40459361    0.08979167
## Velocite_2     1.982975 -2.959123171    1.30416133    0.65630957
## Velocite_3    -3.337405 -1.307012748 -2.70212535   -0.72771249
## Intelligence_1 -1.091286    2.700550551 -1.00684075   -0.37695942
## Intelligence_2  2.125522 -1.101283907    2.86137912   -2.04467438
## Intelligence_3 -1.273022 -1.565170677 -2.19539491    2.71122555
## Affection_1    -3.971260    1.559294642    0.84160708   -1.16111137
## Affection_2     3.971260 -1.559294642   -0.84160708    1.16111137
## Agressivite_1  2.159124 -1.055438765   -0.94828929   -3.70541462
## Agressivite_2 -2.159124    1.055438765    0.94828929    3.70541462
##
## $eta2
##           Dim 1           Dim 2           Dim 3           Dim 4
## Taille      0.8897105 0.48476391 0.29356228 0.05250841
## Poids       0.6366877 0.71705966 0.34825241 0.02635943
## Velocite    0.4383338 0.68272633 0.28102789 0.02543211
## Intelligence 0.1749381 0.29309697 0.33969853 0.30259993
## Affection   0.6065734 0.09351538 0.02724240 0.05185306
## Agressivite 0.1793006 0.04284427 0.03458664 0.52808067
```

Pour chacune des 4 dimensions, nous allons retenir les modalités dont la contribution est supérieure à la moyenne (soit 6.25), puis pour chacune des modalités retenues, nous allons noter la qualité de leur représentation sous cet axe, ainsi que le signe de ses coordonnées.

Dimension 1

Modalité	Contribution	Qualité	Signe
Poids_1	14.01	0.58	+
Taille_3	13.32	0.87	-
Taille_1	12.56	0.49	+
Affection_1	10.75	0.60	-
Affection_2	9.98	0.60	+
Velocite_3	9.76	0.43	-
Somme	70.38		

Dimension 2

Modalité	Contribution	Qualité	Signe
Velocite_1	17.36	0.64	+
Poids_2	14.89	0.71	-
Taille_2	12.24	0.34	-
Velocite_2	10.24	0.34	-
Taille_1	8.70	0.27	+
Intelligence_1	8.53	0.28	+
Somme	71.96		

Dimension 3

Modalité	Contribution	Qualité	Signe
Poids_3	20.71	0.37	+
Velocite_3	14.14	0.28	-
Intelligence_2	13.21	0.32	-
Intelligence_3	10.37	0.18	+
Taille_1	8.68	0.16	-
Somme	67.44		

Dimension 4

Modalité	Contribution	Qualité	Signe
Agressivite_2	27.74	0.53	
Agressivite_1	25.76	0.53	
Intelligence_3	21.22	0.28	
Intelligence_2	9.01	0.16	
Somme	83.73		

Analyse des quatres premiers axes en fonction des individus

Nous allons reprendre la même méthode que précédemment mais en l'appliquant aux individus et non plus aux modalités, dans ce cas on utilisera les données suivantes :

```
data.mca$ind
```

```
## $coord
##          Dim 1      Dim 2      Dim 3      Dim 4
## BEAU -0.5259384 -0.47575949 -0.58374071  0.55564235
## BASS  0.2726981  1.08888034 -0.24197092  0.25517645
## B.AL -0.5259384 -0.47575949 -0.58374071  0.55564235
## BOXE  0.4519622 -0.87628238  0.65659133  0.41868418
## BULD  1.0369751  0.52959725 -0.16067532 -0.29265208
## BULM -0.7430945  0.55653626  0.42440108  0.59354942
## CANI  0.8836606 -0.06443988 -0.62899988  0.33303322
## CHIH  0.8470130  0.81345055 -0.49119441 -0.15522858
## COCK  0.7526566  0.07225541  0.55583657  0.48058272
## COLL -0.1117012 -0.51249596 -0.23503778 -0.59565654
## DAL   0.6542361 -0.98745978  0.52455242 -0.17901219
## DOBE -0.8979794 -0.31150711 -0.46655613  0.36835102
## DA.L -1.0264251  0.53796989  0.18135217  0.08498958
## E.BR  0.4422729 -1.06190072  0.04381058  0.37459034
```

```

## E.FR -0.1197618 -0.49097917 0.28783575 -0.61848428
## FX.H -0.8759782 0.04678712 -0.31633338 -0.04782801
## FX.T 0.8933499 0.12117846 -0.01621913 0.37712706
## GBLG -0.5119978 -0.09594847 0.08935556 0.11663557
## LABR 0.6542361 -0.98745978 0.52455242 -0.17901219
## LEVR -0.6737043 -0.06439028 -0.44837229 -0.64552437
## MAST -0.7210934 0.91483050 0.57462384 0.17737039
## PEKI 0.8470130 0.81345055 -0.49119441 -0.15522858
## POIN -0.6957054 -0.42268452 -0.59859504 -0.22934534
## ST-B -0.5311313 0.63097720 0.90514293 0.03994690
## SETT -0.4837422 -0.34824358 -0.11785320 -0.78294787
## TECK 1.0369751 0.52959725 -0.16067532 -0.29265208
## T.NE -0.3288574 0.51979980 0.77310402 -0.55774946

```

```
##
```

```
## $contrib
```

```

##          Dim 1          Dim 2          Dim 3          Dim 4
## BEAU 2.10111844 2.17369113 5.71765657 6.95239166
## BASS 0.56486551 11.38631564 0.98243645 1.46630641
## B.AL 2.10111844 2.17369113 5.71765657 6.95239166
## BOXE 1.55161864 7.37413469 7.23383138 3.94745005
## BULD 8.16803372 2.69348375 0.43318803 1.92862044
## BULM 4.19439182 2.97447169 3.02225475 7.93336291
## CANI 5.93132626 0.03987790 6.63864170 2.49757571
## CHIH 5.44955505 6.35455623 4.04841413 0.54260890
## COCK 4.30303195 0.05013761 5.18408763 5.20091611
## COLL 0.09477557 2.52234098 0.92694388 7.98979010
## DAL 3.25124622 9.36400725 4.61695759 0.72162030
## DOBE 6.12510519 0.93187754 3.65246368 3.05539455
## DA.L 8.00267820 2.77932193 0.55185326 0.16265782
## E.BR 1.48580379 10.82905948 0.03220594 3.15977886
## E.FR 0.10894759 2.31498957 1.39016885 8.61392102
## FX.H 5.82864293 0.02102208 1.67906747 0.05151197
## FX.T 6.06211255 0.14101775 0.00441401 3.20271975
## GBLG 1.99121023 0.08840951 0.13397424 0.30634135
## LABR 3.25124622 9.36400725 4.61695759 0.72162030
## LEVR 3.44762033 0.03981654 3.37330515 9.38358594
## MAST 3.94969799 8.03718894 5.54045425 0.70844447
## PEKI 5.44955505 6.35455623 4.04841413 0.54260890
## POIN 3.67647489 1.71575624 6.01235105 1.18446820
## ST-B 2.14281499 3.82340445 13.74714214 0.03593434
## SETT 1.77749654 1.16463313 0.23305630 13.80413634
## TECK 8.16803372 2.69348375 0.43318803 1.92862044
## T.NE 0.82147815 2.59474760 10.02891524 7.00522151

```

```
##
```

```
## $cos2
```

```

##          Dim 1          Dim 2          Dim 3          Dim 4
## BEAU 0.19318200 0.158078170 0.2379780727 0.2156193132
## BASS 0.03893080 0.620711031 0.0306517673 0.0340886922
## B.AL 0.19318200 0.158078170 0.2379780727 0.2156193132
## BOXE 0.11183606 0.420403058 0.2360304269 0.0959733835
## BULD 0.64266568 0.167625423 0.0154293198 0.0511860542
## BULM 0.27815801 0.156023623 0.0907311457 0.1774668333
## CANI 0.38024293 0.002022086 0.1926598936 0.0540088961
## CHIH 0.38556757 0.355617092 0.1296663063 0.0129498182

```

```
## COCK 0.28975347 0.002670396 0.1580262244 0.1181329992
## COLL 0.01095173 0.230541033 0.0484889401 0.3114290851
## DAL 0.23755593 0.541172466 0.1527125209 0.0177853352
## DOBE 0.55359781 0.066618921 0.1494408973 0.0931505474
## DA.L 0.53895247 0.148051191 0.0168244731 0.0036951098
## E.BR 0.09451162 0.544844311 0.0009273902 0.0677981326
## E.FR 0.01169406 0.196541690 0.0675488576 0.3118788278
## FX.H 0.55756256 0.001590595 0.0727104867 0.0016621553
## FX.T 0.44081779 0.008110859 0.0001453019 0.0785581701
## GBLG 0.18220281 0.006398746 0.0055496040 0.0094554212
## LABR 0.23755593 0.541172466 0.1527125209 0.0177853352
## LEVR 0.33582962 0.003067758 0.1487503335 0.3083228462
## MAST 0.27298068 0.439369729 0.1733469229 0.0165162378
## PEKI 0.38556757 0.355617092 0.1296663063 0.0129498182
## POIN 0.33802415 0.124775596 0.2502437311 0.0367347283
## ST-B 0.16426904 0.231835114 0.4770745045 0.0009292189
## SETT 0.20103461 0.104185949 0.0119323415 0.5266336570
## TECK 0.64266568 0.167625423 0.0154293198 0.0511860542
## T.NE 0.06389477 0.159632893 0.3531230245 0.1837928020
```

Comme précédemment, pour chaque dimension, on ne garde que les races canines ayant une contribution supérieure à la moyenne qui est de :

```
mean(data.mca$ind$contrib[,1])
```

```
## [1] 3.703704
```

Dimension 1

Individu	Contribution	Qualité	Signe
BULD	8.17		
TECK	8.17		
DA.L	8		
DOBE	6.13		
FX.T	6.06		
CANI	5.93		
CHIH	5.45		
PEKI	5.45		
FX.H	5.83		
COCK	4.30		
BULM	4.20		
Somme	67.69		

Dimension 2

Individu	Contribution	Qualité	Signe
BASS	11.38		
E.BR	10.83		
LABR	9.36		
DAL	9.36		
MAST	8.04		
BOXE	7.37		
PEKI	6.35		
CHIH	6.35		

Individu	Contribution	Qualité	Signe
Somme	69.04		

Dimension 3

Individu	Contribution	Qualité	Signe
ST-B	13.75		
T.NE	10.03		
BOXE	7.23		
CANI	6.64		
POIN	6.01		
BEAU	5.72		
B.AL	5.72		
MAST	5.54		
COCK	5.18		
LABR	4.62		
Somme	70.44		

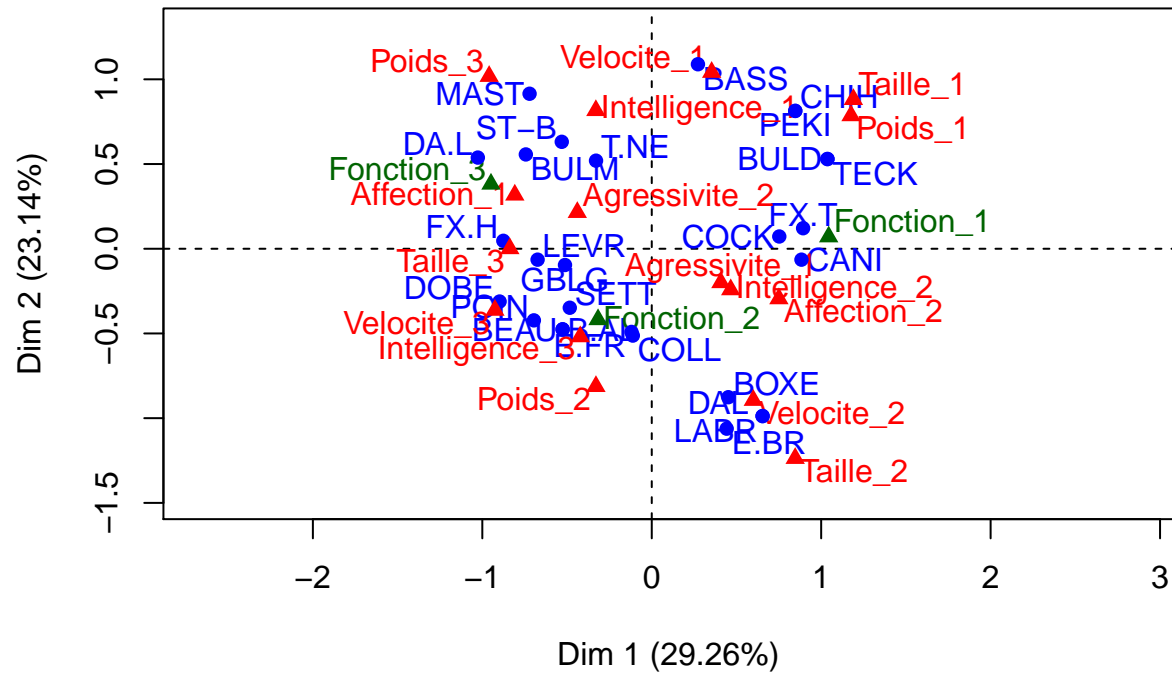
Dimension 4

Interprétation du premier plan factoriel

On peut obtenir les projections des marques et des attributs dans le premier plan factoriel grâce à :

```
plot.MCA(data.mca, axes = c(1, 2))
```

MCA factor map



l'axe2 oppose les hommes bruns représentés par les EL12 et 7 aux femmes blondes représentées par les EL11 et 1.