

Réalisation d'une ACM

Rémi Taniel, Clara Maignan, Vincent Rosset, Hugo Deleque

24/09/2019

Contents

1ère partie : Importation des données	2
2e partie : Mise en oeuvre de l'ACM	3
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	7
Interprétation du premier plan factoriel	8

1ère partie : Importation des données

On commence par importer les données grâce à :

```
data <- read.table("/home/remi/Documents/Cours/AD/data/classe.csv", sep = ",", dec = ".", header = TRUE)
```

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

```
str(data)

## 'data.frame': 14 obs. of 4 variables:
## $ libelle: Factor w/ 14 levels "EL1","EL10","EL11",...: 1 7 8 9 10 11 12 13 14 2 ...
## $ Q1      : Factor w/ 2 levels "F","H": 1 1 1 1 1 1 2 2 2 2 ...
## $ Q2      : Factor w/ 4 levels "BLO","BRU","CHA",...: 1 1 2 2 4 3 2 4 3 2 ...
## $ Q3      : Factor w/ 5 levels "AUTRE","HDF",...: 5 3 4 2 1 2 3 1 5 2 ...
```

Notre jeu de données comporte 14 observations (14 réponses) et 4 variables (les 3 questions et un libellé), on remarque également que toutes les variables sont des variables qualitatives, c'est à dire, le sexe :

- H: Homme
- F: Femme

Les réponses à la seconde question (quelle est la couleur de tes cheveux) sont :

- BLO: Blond
- BRU: Brun
- NOIR: Noir
- CHA: Chatain

Puis pour la région de naissance :

- HDF: Haut De France:
- IDF: Ile de France
- PDLNORM: Pays de la Loire et Normandie
- SUD: Sud
- Autre: à l'étranger

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

```
head(data)

##   libelle Q1  Q2   Q3
## 1    EL1  F  BLO   SUD
## 2    EL2  F  BLO   IDF
## 3    EL3  F  BRU PDLNORM
## 4    EL4  F  BRU   HDF
## 5    EL5  F NOIR  AUTRE
## 6    EL6  F  CHA   HDF
```

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

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

```
##   Q1  Q2   Q3
## EL1  F  BLO   SUD
## EL2  F  BLO   IDF
## EL3  F  BRU PDLNORM
## EL4  F  BRU   HDF
```

```
## EL5 F NOIR AUTRE
## EL6 F CHA HDF
```

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 data.ca, 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 = 5, 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 6.666667e-01      2.500000e+01
## dim 2 5.581547e-01      2.093080e+01
## dim 3 4.972628e-01      1.864736e+01
## dim 4 3.598069e-01      1.349276e+01
## dim 5 2.475796e-01      9.284233e+00
## dim 6 2.019731e-01      7.573993e+00
## dim 7 1.352229e-01      5.070858e+00
## dim 8 7.192388e-33      2.697145e-31
##          cumulative percentage of variance
## dim 1          25.00000
## dim 2          45.93080
## dim 3          64.57816
## dim 4          78.07092
## dim 5          87.35515
## dim 6          94.92914
## dim 7         100.00000
## dim 8         100.00000
```

Part d'inertie supérieure à la moyenne

La moyenne des part d'inertie expliquée par chaque axe peut être obtenue par :

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

```
## [1] 12.5
```

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 12.5.

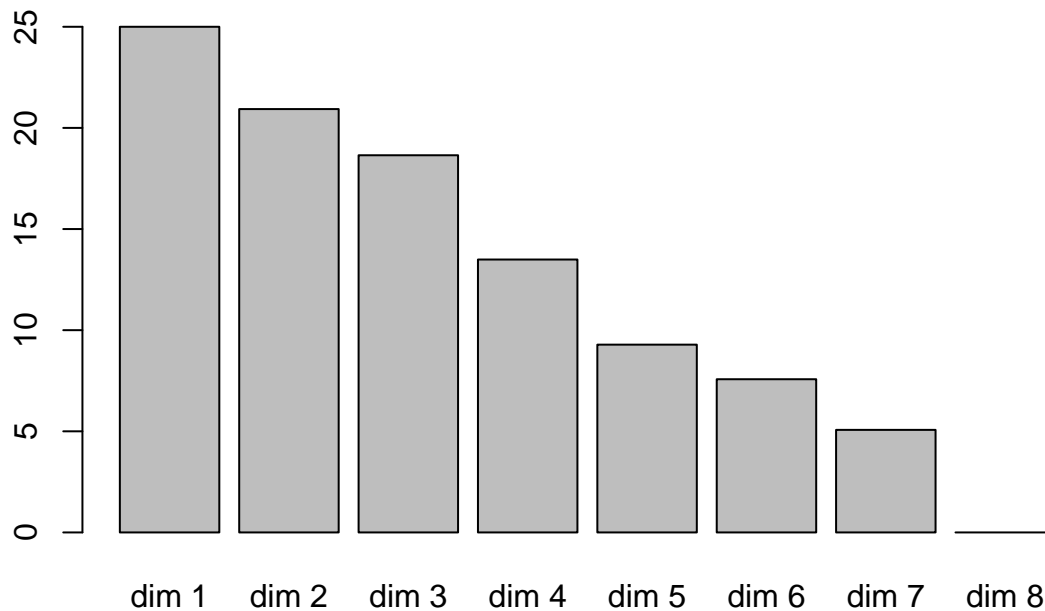
Part d'inertie cumulée supérieure à 80%

Tout comme le précédent critère, nous retenons les 5 premiers axes, en effet, ces 5 dimensions expliquent 87,36% de l'inertie totale portée par nos données.

Critère du coude

Afin d'appliquer ce critère, nous devons dans un premier temps, tracer le graphique suivant :

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



Le coude apparaît entre la 5e et 6e dimension, donc en utilisant ce critère nous devons également retenir 5 dimensions.

Conclusion sur le nombre d'axe à retenir

Selon les 3 critères, nous devons seulement retenir les 4 premiers axes, on retrouve donc la valeur du paramètre nommé `ncp` lors de l'appel à la fonction qui calcule l'ACM (FactoMineR::MCA).

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
## F      5.407468e-16  7.632649e-01 -2.704447e-01 -2.278301e-01
## H      5.181609e-16 -7.632649e-01  2.704447e-01  2.278301e-01
## BLO    -4.082483e-01  9.722857e-01 -2.863380e-01  5.837282e-01
## BRU    -4.082483e-01 -7.612794e-01 -4.601367e-01 -2.382210e-01
## CHA    -4.082483e-01  3.392669e-01  1.953086e+00 -4.527932e-01
## NOIR    2.449490e+00  1.713583e-16  7.256221e-16 -2.137118e-16
## AUTRE    2.449490e+00  2.184917e-16  1.272502e-15 -5.464572e-16
## HDF    -4.082483e-01 -5.146459e-01  2.264318e-01 -8.962340e-01
## IDF    -4.082483e-01 -6.491763e-01 -6.345322e-01  1.406500e+00
## PDLNORM -4.082483e-01  1.288087e+00 -1.308860e+00 -6.934772e-01
## SUD    -4.082483e-01  9.722925e-01  1.694578e+00  8.243116e-01
##          Dim 5
## F      1.109516e-02
## H     -1.109516e-02
## BLO    -9.195327e-01
## BRU     3.608875e-01
## CHA     7.564029e-01
## NOIR    -2.396069e-15
## AUTRE    -2.454328e-15
## HDF     -7.062372e-01
## IDF     2.706105e-01
## PDLNORM  1.042626e+00
## SUD     3.170507e-01
##
## $contrib
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## F      7.310179e-30  1.739581e+01  2.451431e+00  2.404370e+00  8.287067e-03
## H      6.712268e-30  1.739581e+01  2.451431e+00  2.404370e+00  8.287067e-03
## BLO    2.380952e+00  1.613035e+01  1.570301e+00  9.019085e+00  3.252597e+01
## BRU    3.571429e+00  1.483322e+01  6.082606e+00  2.253163e+00  7.515032e+00
## CHA    1.190476e+00  9.819948e-01  3.652898e+01  2.713382e+00  1.100455e+01
## NOIR    4.285714e+01  2.505165e-31  5.042150e-30  6.044611e-31  1.104242e-28
## AUTRE    4.285714e+01  4.072828e-31  1.550641e-29  3.952060e-30  1.158594e-28
## HDF     2.976190e+00  5.649150e+00  1.227466e+00  2.657627e+01  2.398320e+01
## IDF     1.785714e+00  5.393152e+00  5.783534e+00  3.927192e+01  2.112742e+00
## PDLNORM  1.190476e+00  1.415521e+01  1.640518e+01  6.364664e+00  2.090852e+01
## SUD     1.190476e+00  8.065291e+00  2.749908e+01  8.992780e+00  1.933408e+00
##
## $cos2
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## F      2.924071e-31  5.825733e-01  7.314031e-02  5.190654e-02  1.231025e-04
## H      2.684907e-31  5.825733e-01  7.314031e-02  5.190654e-02  1.231025e-04
## BLO    6.666667e-02  3.781358e-01  3.279579e-02  1.362954e-01  3.382162e-01
## BRU    1.250000e-01  4.346598e-01  1.587943e-01  4.256195e-02  9.767984e-02
## CHA    2.777778e-02  1.918367e-02  6.357576e-01  3.417028e-02  9.535756e-02
## NOIR    1.000000e+00  4.893944e-33  8.775458e-32  7.612126e-33  9.568575e-31
## AUTRE    1.000000e+00  7.956438e-33  2.698767e-31  4.976925e-32  1.003954e-30
## HDF     9.259259e-02  1.471446e-01  2.848409e-02  4.462418e-01  2.770950e-01
## IDF     4.545455e-02  1.149354e-01  1.098085e-01  5.395209e-01  1.997183e-02
## PDLNORM  2.777778e-02  2.765278e-01  2.855189e-01  8.015176e-02  1.811783e-01
## SUD     2.777778e-02  1.575588e-01  4.785994e-01  1.132483e-01  1.675353e-02
##
## $v.test

```

```

##          Dim 1          Dim 2          Dim 3          Dim 4
## F      1.949690e-15  2.751991e+00 -9.751021e-01 -8.214530e-01
## H      1.868256e-15 -2.751991e+00  9.751021e-01  8.214530e-01
## BLO    -9.309493e-01  2.217152e+00 -6.529512e-01  1.331105e+00
## BRU    -1.274755e+00 -2.377094e+00 -1.436776e+00 -7.438450e-01
## CHA    -6.009252e-01  4.993874e-01  2.874865e+00 -6.664935e-01
## NOIR    3.605551e+00  2.522326e-16  1.068087e-15 -3.145753e-16
## AUTRE    3.605551e+00  3.216111e-16  1.873072e-15 -8.043632e-16
## HDF    -1.097134e+00 -1.383069e+00  6.085172e-01 -2.408556e+00
## IDF     -7.687061e-01 -1.222358e+00 -1.194785e+00  2.648353e+00
## PDLNORM -6.009252e-01  1.896012e+00 -1.926589e+00 -1.020771e+00
## SUD     -6.009252e-01  1.431176e+00  2.494352e+00  1.213354e+00
##          Dim 5
## F      4.000416e-02
## H     -4.000416e-02
## BLO    -2.096857e+00
## BRU     1.126871e+00
## CHA     1.113395e+00
## NOIR    -3.526918e-15
## AUTRE    -3.612673e-15
## HDF     -1.897955e+00
## IDF      5.095428e-01
## PDLNORM  1.534705e+00
## SUD      4.666861e-01
##
## $eta2
##          Dim 1          Dim 2          Dim 3          Dim 4          Dim 5
## Q1  2.804489e-31  0.5825733  0.07314031  0.05190654  0.0001231025
## Q2  1.000000e+00  0.5349171  0.65910025  0.15096380  0.3791350803
## Q3  1.000000e+00  0.5569736  0.75954785  0.87655046  0.3634804930

```

Pour chacune des 4 dimensions, nous allons retenir les variables dont la contribution est supérieure à la moyenne (soit 8.33), puis pour chacune des marques retenues, nous allons noter la qualité de leur représentation sous cet axe, ainsi que le signe de ses coordonnées.

Dimension 1

Dimension 2

Mddalité	Contribution	Qualité	Signe
F	17.4	0.583	-
H	17.4	0.583	+
BLO	16.1	0.378	-
BRU	14.8	0.435	+
PDLNORM	14.2	0.277	-
Somme	79.9		

Ces 5 modalités représentent 79.9% de l'information totale portée par la dimension 2, cet axe oppose les variables Homme (H) et Brund (BRU) aux variables Femmes (F), Blond (BLO) et Pays de la Loire Normandie (PDLNORM).

Dimension 3

Dimension 4

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      Dim 5
## EL1  -0.3333333  1.2081614  0.53783625  0.65584792 -0.39618040
## EL2  -0.3333333  0.4847088 -0.56313476  0.97937289 -0.42729152
## EL3  -0.3333333  0.5755929 -0.96404412 -0.64435518  0.94767143
## EL4  -0.3333333 -0.2287343 -0.23831157 -0.75702807 -0.22392297
## EL5   2.0000000  0.3405467 -0.12783924 -0.12660622  0.00743284
## EL6  -0.3333333  0.2622977  0.90241927 -0.87626680  0.04103976
## EL7  -0.3333333 -0.9698514 -0.38961093  0.77582437  0.41561877
## EL8   2.0000000 -0.3405467  0.12783924  0.12660622 -0.00743284
## EL9  -0.3333333  0.2446332  1.85209092  0.33306068  0.71169261
## EL10 -0.3333333 -0.9098278  0.01736690 -0.50381564 -0.23878865
## EL11 -0.3333333  1.3490596 -0.88188947 -0.18759423  0.08989547
## EL12 -0.3333333 -0.9698514 -0.38961093  0.77582437  0.41561877
## EL13 -0.3333333 -0.1363611  0.09952155 -0.04705469 -1.09656461
## EL14 -0.3333333 -0.9098278  0.01736690 -0.50381564 -0.23878865
##
## $contrib
##          Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## EL1   1.190476 18.6795888  4.155144424  8.53903374  4.528383116
## EL2   1.190476  3.0066251  4.555233685 19.04138752  5.267514251
## EL3   1.190476  4.2398279 13.349955221  8.24238857 25.910318838
## EL4   1.190476  0.6695452  0.815785991 11.37696440  1.446619995
## EL5  42.857143  1.4841264  0.234754716  0.31820903  0.001593921
## EL6   1.190476  0.8804536 11.697760183 15.24316489  0.048592218
## EL7   1.190476 12.0372632  2.180460690 11.94893710  4.983649446
## EL8  42.857143  1.4841264  0.234754716  0.31820903  0.001593921
## EL9   1.190476  0.7658580 49.273180108  2.20216146 14.613090941
## EL10  1.190476 10.5934101  0.004332422  5.03901578  1.645070573
## EL11  1.190476 23.2905453 11.171572468  0.69862062  0.233148566
## EL12  1.190476 12.0372632  2.180460690 11.94893710  4.983649446
## EL13  1.190476  0.2379569  0.142272264  0.04395499 34.691704194
## EL14  1.190476 10.5934101  0.004332422  5.03901578  1.645070573
##
## $cos2
##          Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## EL1  0.03508772  0.46094335  0.091347735  0.135832578  4.956597e-02
## EL2  0.04651163  0.09834809  0.132748225  0.401513549  7.642802e-02
## EL3  0.04000000  0.11927058  0.334577183  0.149469695  3.233092e-01
## EL4  0.08064516  0.03797374  0.041220293  0.415953504  3.639302e-02
## EL5  0.92307692  0.02676279  0.003771432  0.003699031  1.274933e-05
## EL6  0.03787879  0.02345458  0.277622913  0.261764829  5.741801e-04
## EL7  0.05555556  0.47030582  0.075898340  0.300951729  8.636948e-02
## EL8  0.92307692  0.02676279  0.003771432  0.003699031  1.274933e-05
## EL9  0.02564103  0.01381048  0.791594021  0.025599096  1.168861e-01
```

```
## EL10 0.08064516 0.60081284 0.000218910 0.184231593 4.138550e-02
## EL11 0.03508772 0.57472475 0.245598645 0.011113136 2.551956e-03
## EL12 0.05555556 0.47030582 0.075898340 0.300951729 8.636948e-02
## EL13 0.06289308 0.01052510 0.005606343 0.001253289 6.806343e-01
## EL14 0.08064516 0.60081284 0.000218910 0.184231593 4.138550e-02
```

Comme précédemment, pour chaque dimension,

Dimension 1

Dimension 2

Individu	Contribution	Qualité	Signe
EL11	23.3	0.57	+
EL1	18.7	0.46	+
EL7	12	0.47	-
EL12	12	0.47	-
Somme	66		

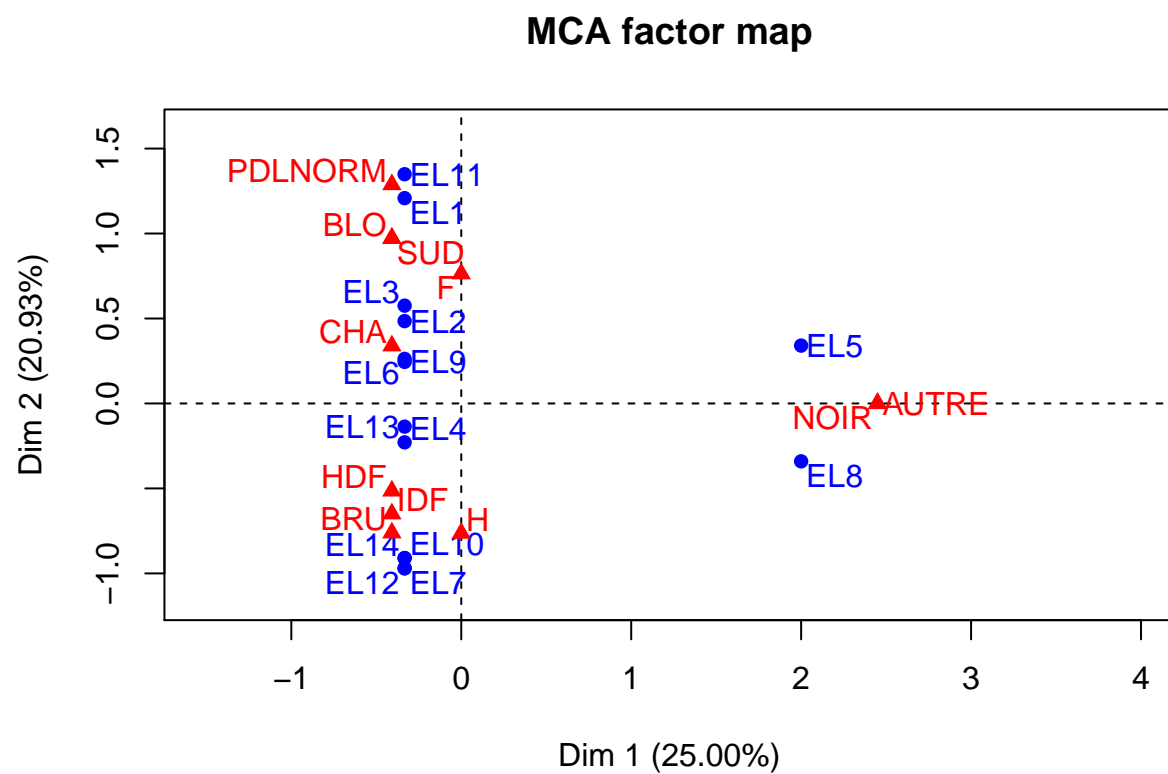
Ces 4 individus expliquent 66% de l'information portée par l'axe 2. Il oppose l'EL11 et 1 aux élèves 7 et 12.
 ### Dimension 3

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))
```

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