

Enquête de la régie française du tabac

Rémi Taniel

23/09/2019

Introduction

1ère partie : Importation des données

On commence par importer les données grâce à :

```
data <- read.table("../data/fumeur.csv", sep = ",", dec = ".", header = TRUE)
```

Puis on visualise les types de ces données avec la fonction `str` :

```
str(data)

## 'data.frame': 11 obs. of 13 variables:
## $ libellé: Factor w/ 11 levels "coca", "dist", ...: 10 6 9 1 8 5 2 11 4 3 ...
## $ ORLY : num 1 20 9 1 4 3 11 4 9 9 ...
## $ ALEZ : num 2 9 23 3 33 9 9 4 12 3 ...
## $ CORS : num 14 1 1 15 7 1 1 32 23 9 ...
## $ DIRE : num 38 11 15 15 8 7 17 2 4 8 ...
## $ DUCA : num 18 10 7 6 3 7 4 6 7 4 ...
## $ FONT : num 10 9 11 5 6 5 21 0 13 2 ...
## $ ICAR : num 9 1 6 12 6 12 6 9 5 6 ...
## $ ZODI : num 5 1 2 18 4 9 1 7 5 8 ...
## $ PAVI : num 9 20 7 4 5 6 5 3 10 1 ...
## $ COCK : num 4 9 12 25 15 9 4 10 5 6 ...
## $ ESCA : num 0 7 3 2 5 6 5 12 13 23 ...
## $ HOTE : num 1 12 17 2 3 13 27 7 9 33 ...
```

On remarque que nous avons 11 observations, qui représentent les marques de cigarette pour 13 variables, représentant les différents attributs associés aux marques, on remarque également que toutes nos variables sont numériques, on décide d'afficher les 6 premières lignes de nos données :

```
head(data)

## libellé ORLY ALEZ CORS DIRE DUCA FONT ICAR ZODI PAVI COCK ESCA HOTE
## 1 viel 1 2 14 38 18 10 9 5 9 4 0 1
## 2 nouv 20 9 1 11 10 9 1 1 20 9 7 12
## 3 sobr 9 23 1 15 7 11 6 2 7 12 3 17
## 4 coca 1 3 15 15 6 5 12 18 4 25 2 2
## 5 racé 4 33 7 8 3 6 6 4 5 15 5 3
## 6 miev 3 9 1 7 7 5 12 9 6 9 6 13
```

La variable `libellé` n'est pas utile, on peut donc l'enlever, pour cela on reformate les données pour donner un identifiant aux différentes lignes, dans notre cas, ce sera `libellé` qui représente les attributs :

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

## ORLY ALEZ CORS DIRE DUCA FONT ICAR ZODI PAVI COCK ESCA HOTE
## viel 1 2 14 38 18 10 9 5 9 4 0 1
## nouv 20 9 1 11 10 9 1 1 20 9 7 12
```

## sobr	9	23	1	15	7	11	6	2	7	12	3	17
## coca	1	3	15	15	6	5	12	18	4	25	2	2
## racé	4	33	7	8	3	6	6	4	5	15	5	3
## miev	3	9	1	7	7	5	12	9	6	9	6	13

2e partie : Mise en oeuvre de l'AFC

Pour réaliser l'AFC sur nos données, nous aurons besoin du package FactoMineR :

```
library(FactoMineR)
```

Puis on range les résultats de l'AFC (valeurs propres, coordonnées, contribution) dans la variable `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.ca <- FactoMineR::CA(data, ncp = 4, 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

Part d'inertie supérieure à la moyenne

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

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

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

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

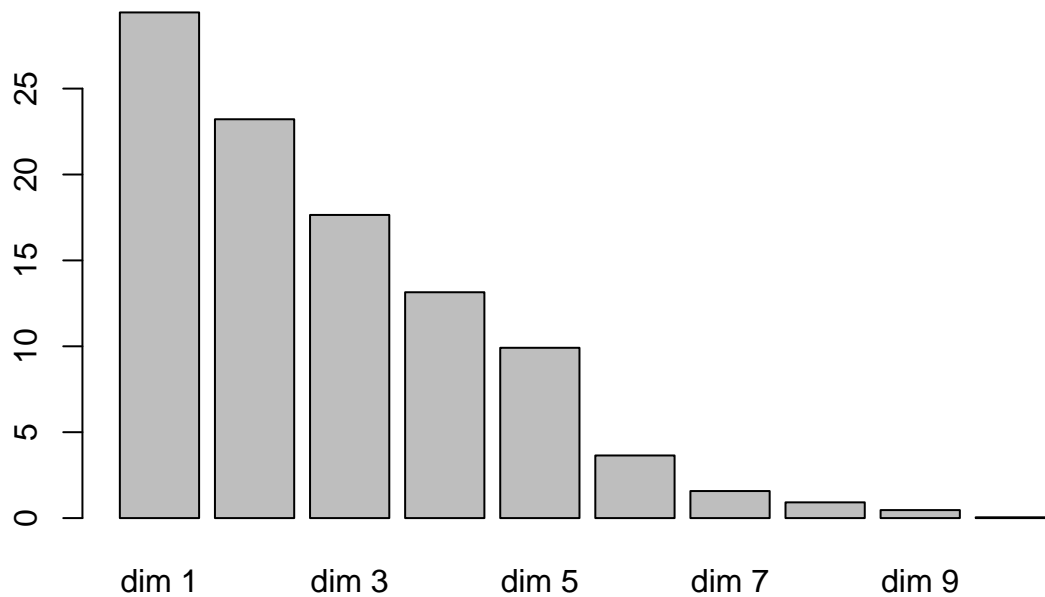
Part d'inertie cumulée supérieure à 80%

Tout comme le précédent critère, nous retenons les 4 premiers axes, en effet, ces 4 ci expliquent 83,44% 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.ca$eig[,2])
```



Le coude apparaît entre la 4e et 5e dimension, donc en utilisant ce critère nous devons également retenir 4 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'AFC (`FactoMineR::CA`).

Analyse des quatres premiers axes en fonction des marques

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

```
data.ca$col
```

```
## $coord
##          Dim 1      Dim 2      Dim 3      Dim 4
## ORLY -0.47831947  0.22165433  0.024298167 -0.013194789
## ALEZ -0.35426935 -0.15644651 -0.726248485 -0.344684396
## CORS  0.83599692  0.26835346  0.140445828 -0.556258698
## DIRE -0.06921791 -0.55105987  0.418110615  0.003482836
## DUCA  0.10461567 -0.32051558  0.246873387  0.053159128
## FONT -0.45932425 -0.19594988  0.188147869 -0.230673125
## ICAR  0.29065203 -0.08167104 -0.007082372  0.133457698
## ZODI  0.56516935 -0.01017983 -0.111298453  0.462528856
## PAVI -0.21801437 -0.22754233  0.023454541 -0.029725028
## COCK  0.28434929 -0.15395608 -0.396083330  0.370078341
## ESCA  0.08525825  0.71116711  0.016240731  0.068450686
```

```

## HOTE -0.46878764  0.54600907  0.209263994  0.161197685
##
## $contrib
##          Dim 1          Dim 2          Dim 3          Dim 4
## ORLY  9.0499380  2.463961589  0.03895623  0.015418874
## ALEZ  7.1285315  1.762530709 49.97165557 15.108259271
## CORS 37.5690403  4.908029390  1.76871878 37.240255834
## DIRE  0.3207198 25.772609744 19.52052778  0.001817997
## DUCA  0.4606663  5.482306035  4.27919305  0.266309673
## FONT  8.9873762  2.073748890  2.51543518  5.074887806
## ICAR  3.3416135  0.334516627  0.00330969  1.577375255
## ZODI 11.5008630  0.004730709  0.74399813 17.246031462
## PAVI  1.9041985  2.629894899  0.03676345  0.079254657
## COCK  5.0434099  1.874501025 16.32350725 19.126945023
## ESCA  0.3170196 27.965891963  0.01918867  0.457516990
## HOTE 14.3766234 24.727278420  4.77874623  3.805927158
##
## $cos2
##          Dim 1          Dim 2          Dim 3          Dim 4
## ORLY 0.465518364 0.0999661876 0.0012012896 3.542465e-04
## ALEZ 0.148279059 0.0289164133 0.6231364098 1.403641e-01
## CORS 0.630072208 0.0649225186 0.0177827388 2.789553e-01
## DIRE 0.008730227 0.5533325132 0.3185451566 2.210318e-05
## DUCA 0.039945571 0.3749500492 0.2224455275 1.031409e-02
## FONT 0.422943985 0.0769722323 0.0709647545 1.066690e-01
## ICAR 0.357078365 0.0281937612 0.0002120188 7.528424e-02
## ZODI 0.543689067 0.0001763902 0.0210849052 3.641422e-01
## PAVI 0.113339259 0.1234623458 0.0013117890 2.106952e-03
## COCK 0.186235189 0.0545947822 0.3613521062 3.154604e-01
## ESCA 0.012230240 0.8509527688 0.0004437857 7.883476e-03
## HOTE 0.336140787 0.4560039611 0.0669819710 3.974537e-02
##
## $inertia
## [1] 0.03301883 0.08165319 0.10127277 0.06239547 0.01958713 0.03609134
## [7] 0.01589445 0.03592797 0.02853545 0.04599551 0.04402548 0.07264213

```

Pour chacune ds 4 dimensions, nous allons retenir les marques 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

Marque	Contribution	Qualité	Signe
Corsaire	37.57	0.63	+
Hotesse	14.38	0.34	+
Zodiac	11.50	0.54	+
Orly	9.05	0.47	-
Somme	72.50		

Ces 4 marques représentent plus de 70% de l'information totale expliquée par le premier axe, il permet de mettre en opposition les marques Corsaire, Hotesse, Zodiac et Orly.

Dimension 2

Marque	Contribution	Qualité	Signe
Escale	27.97	0.85	+
Directoire	25.77	0.55	-
Hotesse	24.73	0.46	+
Somme	78.47		

Les 3 marques retenus expliquent à elles seuls 78.47% de l'information portée par la dimension 2, de plus on voit apparaître une opposition entre les marques **Escale**, **Hotesse** et **Directoire**.

Dimension 3

Marque	Contribution	Qualité	Signe
Alezan	49.97	0.62	+
Directoire	19.52	0.32	-
Cocker	16.32	0.36	+
Somme	85.81		

Nous retenons ces 3 marques qui représentent un peu plus de 85% de l'information portée par l'axe 3, cette dimension met en lumière une opposition entre les marques de cigarettes **Alezan**, **Cocker** et **Directoire**.

Dimension 4

Marque	Contribution	Qualité	Signe
Corsaire	37.24	0.28	+
Cocker	19.13	0.32	-
Zodiac	17.25	0.36	-
Aleza	15.11	0.14	+
Somme	88.73		

Pour la dimension 4, on décide de retenir ces 4 marques de cigarettes, elles représentent 88.73% de l'information portée par cet axe et met en opposition **Corsaire** et **Aleza** à **Cocker** et **Zodiac**.

Analyse des quatres premiers axes en fonction des attributs

Nous allons reprendre la même méthode que précédemment mais en l'appliquant aux lignes et non plus aux colonnes, on utilise les données suivantes :

```
data.ca$row
```

```
## $coord
##          Dim 1          Dim 2          Dim 3          Dim 4
## viel      0.20384148 -0.6869093196  0.591972983 -0.15765680
## nouv     -0.47960002 -0.0555121024  0.064451788  0.01047127
## sobr     -0.42793503 -0.1580067211 -0.208636236 -0.05820138
## coca      0.61609943 -0.2731962130 -0.081241371  0.31419406
## racé     -0.11147865 -0.2132565406 -0.736754202 -0.28013809
## miev     -0.04444349 -0.0004047644 -0.087175797  0.29079169
## dist     -0.60973196  0.0634516763  0.283135564 -0.03553191
## vulg      0.76873090  0.4311869642  0.030371418 -0.33681174
## hom       0.11744125  0.2363755072  0.009188137 -0.40373313
```

```

## fem      -0.10081397  0.7442923337  0.224015652  0.23358590
## petit    0.18705039 -0.0820220311 -0.206220882  0.48456883
##
## $contrib
##          Dim 1      Dim 2      Dim 3      Dim 4
## viel      2.3389610 3.367508e+01 32.905007073  3.13257559
## nouv     12.8311607 2.179490e-01  0.386543045  0.01369445
## sobr     10.4941917 1.813913e+00  4.160956318  0.43460838
## coca     20.7893180 5.182747e+00  0.602994678 12.10522487
## racé      0.6239266 2.894855e+00 45.458560027  8.82130156
## miev      0.0871466 9.164543e-06  0.559301958  8.35288271
## dist     20.9274239 2.873400e-01  7.527435823  0.15911577
## vulg     28.7696635 1.147597e+01  0.074909451 12.36513457
## hom       0.8043667 4.131328e+00  0.008212738 21.28333964
## fem       0.5772638 3.989257e+01  4.754556232  6.93848116
## petit    1.7565776 4.282360e-01  3.561522656 26.39364129
##
## $cos2
##          Dim 1      Dim 2      Dim 3      Dim 4
## viel     0.044665918 5.072125e-01 0.3766993824 0.0267187758
## nouv     0.402265070 5.389262e-03 0.0072648037 0.0001917577
## sobr     0.636961579 8.683777e-02 0.1514038925 0.0117821367
## coca     0.598084107 1.176007e-01 0.0103995642 0.1555451941
## racé     0.017020861 6.228778e-02 0.7434356179 0.1074836571
## miev     0.009082806 7.533701e-07 0.0349458467 0.3888370922
## dist     0.639686995 6.927488e-03 0.1379362616 0.0021723345
## vulg     0.647541532 2.037278e-01 0.0010107625 0.1243063773
## hom      0.046456218 1.881947e-01 0.0002843527 0.5490239296
## fem      0.013615659 7.421375e-01 0.0672285884 0.0730954752
## petit    0.083094134 1.597770e-02 0.1009993060 0.5576530275
##
## $inertia
## [1] 0.08894051 0.05417595 0.02798264 0.05903790 0.06225939 0.01629609
## [7] 0.05556498 0.07546057 0.02940784 0.07200929 0.03590458

```

Comme précédemment, pour chaque dimension,

Dimension 1

Marque	Contribution	Qualité	Signe
vulg	28.77	0.65	+
dist	20.93	0.64	-
coca	20.79	0.60	+
nouv	12.83	0.40	-
Somme	83.32		

Ces 4 attributs expliquent plus de 80% de l'information portée par l'axe 1, celui-ci oppose vulgaire-commun (vulg) et cocasse-ridicule (coca) à distingué (dist) et nouveau-riche (nouv).

Dimension 2

Marque	Contribution	Qualité	Signe
viel	33.68	0.51	-

Marque	Contribution	Qualité	Signe
fem	39.89	0.74	+
vulg	11.48	0.20	+
Somme	85.05		

On ne retient que 3 attributs pour la 2e dimension, celles-ci expliquent 85% de l'information portée par cet axe, celui-ci met en opposition les attributs femme (**fem**) et vulgaire-commun (**vulg**) à vieillot-désuet (**viel**).

Dimension 3

Marque	Contribution	Qualité	Signe
racé	45.46	0.74	+
viel	32.91	0.38	-
Somme	78.37		

On ne retient que 2 attributs sur la dimension 3, néanmoins ces 2 attributs expliquent quasiment 80% de l'information portée par l'axe analysé, et oppose l'attribut racé (**racé**) à l'attribut vieillot-désuet (**viel**).

Dimension 4

Marque	Contribution	Qualité	Signe
petit	26.39	0.56	-
hom	21.28	0.55	+
coca	12.11	0.16	-
vulg	12.37	0.12	+
Somme	72.15		

Pour la dimension 4, les attributs retenus représentent un peu plus de 70% de l'information portée par la dimension 4, celle-ci oppose les attributs petit (**petit**) et cocasse-ridicule (**coca**) aux attributs homme (**hom**) et vulgaire-commun (**vulg**).

Interprétation du premier plan factoriel

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