

# AFC – R version > 2.12.0



Lancer Rstudio à partir de la console : `/usr/local/rstudio/bin/rstudio`. (Plus d'information sur R et Rstudio dans le Tutoriel « ACP NORMEE »)

## ENQUETE DE LA REGIE FRANCAISE DES TABACS

Cette étude de cas est extraite de l'ouvrage de J.P. Benzecri "Analyse des données. Tome 2: l'analyse des correspondances". Une enquête a été effectuée auprès de cent fumeurs afin de choisir les noms de deux nouvelles marques de cigarettes. La première marque est destinée à une clientèle à une clientèle masculine: l'homme ciblé est un connaisseur distingué, raffiné mais viril, de niveau socio-économique élevé. La seconde vise un public féminin, élégant, assuré, dynamique.

Douze marques possibles ont été retenues:  $I = \{\text{Orly (ORLY), Alezan (ALEZ), Corsaire (CORS), Directoire (DIRE), Ducat (DUCA), Fontenoy (FONT), Icare(ICAR), Zodiac(ZODI), Pavois (PAVI), Cocker(COCK), Escala (ESCA), Hôtesse (HOTE)}\}$ . Pour évaluer leur image auprès du public, onze attributs ont été proposés aux fumeurs:  $J = \{\text{Vieillot-désuet (viel), Nouveau-Riche(Nouv), Sobre-élégant(Sobr), Cocasse-Ridicule(Coca), Racé (racé), Mièvre(miév), Distingué(dist), Vulgaire-Commun (vulg), Pour un Homme (hom), Pour une femme (fem), Pour une petite nature (petit)}\}$ . Chaque personne interrogée devait successivement pour chaque attribut, désigner une ou plusieurs marques lui correspondant. Le tableau de données est le suivant:

libellé	ORLY	ALEZ	CORS	DIRE	DUCA	FONT	ICAR	ZODI	PAVI	COCK	ESCA	HOTE
viel	1.000	2.000	14.000	38.000	18.000	10.000	9.000	5.000	9.000	4.000	0.000	1.000
nouv	20.000	9.000	1.000	11.000	10.000	9.000	1.000	1.000	20.000	9.000	7.000	12.000
sobr	9.000	23.000	1.000	15.000	7.000	11.000	6.000	2.000	7.000	12.00	3.000	17.000
coca	1.000	3.000	15.000	15.000	6.000	5.000	12.000	18.000	4.000	25.00	2.000	2.000
racé	4.000	33.000	7.000	8.000	3.000	6.000	6.000	4.000	5.000	15.00	5.000	3.000
miev	3.000	9.000	1.000	7.000	7.000	5.000	12.000	9.000	6.000	9.000	6.000	13.000
dist	11.000	9.000	1.000	17.000	4.000	21.000	6.000	1.000	5.000	4.000	5.000	27.000
vulg	4.000	4.000	32.000	2.000	6.000	0.000	9.000	7.000	3.000	10.00	12.00	7.000
hom	9.000	12.000	23.000	4.000	7.000	13.000	5.000	5.000	10.000	5.000	13.00	9.000
fem	9.000	3.000	9.000	8.000	4.000	2.000	6.000	8.000	1.000	6.000	23.00	33.000
petit	7.000	5.000	2.000	7.000	11.000	2.000	6.000	11.000	9.000	24.00	10.00	5.000

Les valeurs dans le tableau sont des nombres entiers. Chaque valeur du tableau correspond au nombre de personnes ayant choisi simultanément les deux modalités (lignes et colonnes) considérées. Par exemple la valeur 1 de la première et de la première colonne signifie qu'une seule personne parmi l'échantillon associe les modalités VIEL et ORLLY. Alors 20 personnes interrogées associent NOUV à ORLY etc. On veut mettre en œuvre l'Analyse Factorielle des Correspondances (AFC) en utilisant R. La démarche est la même qu'en ACP.

## 1<sup>ère</sup> PARTIE : IMPORT DES DONNEES

```
> mydata<-read.table("fumeur.txt",sep="," ,dec=".",header=TRUE)
> str(mydata)
'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 ...
```

```
> head(mydata)
  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
```

NB : Si la première colonne avait un nom dans le fichier importé, alors R ajoute un identifiant de ligne (1,2,3,...). Il est important de reformater le dataframe pour que le pays soit traité en tant qu'identifiant des lignes. Le script ci-dessous permet de faire la transformation.

```
> rownames(mydata)<-mydata$libellé
```

```
> mydata<-mydata[,-1]
```

```
> head(mydata)
```

	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

## 2<sup>ème</sup> PARTIE : REALISATION DE L'AFC

Une méthode d'AFC est incluse dans le package "FactoMineR".

FactoMineR is an R package dedicated to multivariate Exploratory Data Analysis. It is developed and maintained by F. Husson\*, J. Josse\*, S. Lê\*, from Agrocampus Rennes, and J. Mazet.

Source : <http://factominer.free.fr/index.html>

```
> install.packages("FactoMineR")
trying URL
...
The downloaded binary packages are in
  D:\Users\Gregory\AppData\Local\Temp\RtmpCq0FwY\downloaded_packages
> library(FactoMineR)
> mydata.ca = mydata.ca = CA(mydata, ncp=4, graph=FALSE)
```

L'afc est effectuée sur le jeu de donnée « mydata ». Les 4 premiers axes sont retenus (ncp=4), les graphiques ne sont pas générés à cette étape (graph=FALSE). Les résultats sont stockés dans "mydata.ca" : valeurs propres, coordonnées, contributions, etc.

```
> str(mydata.ca)
List of 5
 $ eig :'data.frame': 10 obs. of  3 variables:
  ..$ eigenvalue      : num [1:10] 0.1698 0.134 0.1018 0.0759 0.0572 ...
  ..$ percentage of variance : num [1:10] 29.43 23.22 17.65 13.15 9.91 ...
  ..$ cumulative percentage of variance: num [1:10] 29.4 52.6 70.3 83.4 93.4 ...
 $ call:List of 9
  ..$ X      :'data.frame':  11 obs. of  12 variables:
  .. ..$ ORLY: num [1:11] 1 20 9 1 4 3 11 4 9 9 ...
  .. ..$ ALEZ: num [1:11] 2 9 23 3 33 9 9 4 12 3 ...
  .. ..$ CORS: num [1:11] 14 1 1 15 7 1 1 32 23 9 ...
  .. ..$ DIRE: num [1:11] 38 11 15 15 8 7 17 2 4 8 ...
  .. ..$ DUCA: num [1:11] 18 10 7 6 3 7 4 6 7 4 ...
  .. ..$ FONT: num [1:11] 10 9 11 5 6 5 21 0 13 2 ...
  .. ..$ ICAR: num [1:11] 9 1 6 12 6 12 6 9 5 6 ...
  .. ..$ ZODI: num [1:11] 5 1 2 18 4 9 1 7 5 8 ...
  .. ..$ PAVI: num [1:11] 9 20 7 4 5 6 5 3 10 1 ...
  .. ..$ COCK: num [1:11] 4 9 12 25 15 9 4 10 5 6 ...
  .. ..$ ESCA: num [1:11] 0 7 3 2 5 6 5 12 13 23 ...
  .. ..$ HOTE: num [1:11] 1 12 17 2 3 13 27 7 9 33 ...
  ..$ marge.col: Named num [1:12] 0.0672 0.0965 0.0913 0.1137 0.0715 ...
  .. ..- attr(*, "names")= chr [1:12] "ORLY" "ALEZ" "CORS" "DIRE" ...
  ..$ marge.row: Named num [1:11] 0.0956 0.0947 0.0973 0.093 0.0853 ...
  .. ..- attr(*, "names")= chr [1:11] "viel" "nouu" "sabr" "coca" ...
  ..$ ncp      : num 4
  ..$ row.w    : num [1:11] 1 1 1 1 1 1 1 1 1 1 ...
  ..$ excl     : NULL
  ..$ call     : language CA(X = mydata, ncp = 4, graph = FALSE)
  ..$ xtot     :'data.frame':  11 obs. of  12 variables:
  .. ..$ ORLY: num [1:11] 1 20 9 1 4 3 11 4 9 9 ...
  .. ..$ ALEZ: num [1:11] 2 9 23 3 33 9 9 4 12 3 ...
  .. ..$ CORS: num [1:11] 14 1 1 15 7 1 1 32 23 9 ...
  .. ..$ DIRE: num [1:11] 38 11 15 15 8 7 17 2 4 8 ...
  .. ..$ DUCA: num [1:11] 18 10 7 6 3 7 4 6 7 4 ...
  .. ..$ FONT: num [1:11] 10 9 11 5 6 5 21 0 13 2 ...
  .. ..$ ICAR: num [1:11] 9 1 6 12 6 12 6 9 5 6 ...
  .. ..$ ZODI: num [1:11] 5 1 2 18 4 9 1 7 5 8 ...
```

```

.. ..$ PAVI: num [1:11] 9 20 7 4 5 6 5 3 10 1 ...
.. ..$ COCK: num [1:11] 4 9 12 25 15 9 4 10 5 6 ...
.. ..$ ESCA: num [1:11] 0 7 3 2 5 6 5 12 13 23 ...
.. ..$ HOTE: num [1:11] 1 12 17 2 3 13 27 7 9 33 ...
..$ N      : num 1161
$ row :List of 4
..$ coord  : num [1:11, 1:4] 0.204 -0.48 -0.428 0.616 -0.111 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:11] "viel" "nouv" "sobr" "coca" ...
.. .. ..$ : chr [1:4] "Dim 1" "Dim 2" "Dim 3" "Dim 4"
..$ contrib: num [1:11, 1:4] 2.339 12.831 10.494 20.789 0.624 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:11] "viel" "nouv" "sobr" "coca" ...
.. .. ..$ : chr [1:4] "Dim 1" "Dim 2" "Dim 3" "Dim 4"
..$ cos2   : num [1:11, 1:4] 0.0447 0.4023 0.637 0.5981 0.017 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:11] "viel" "nouv" "sobr" "coca" ...
.. .. ..$ : chr [1:4] "Dim 1" "Dim 2" "Dim 3" "Dim 4"
..$ inertia: num [1:11] 0.0889 0.0542 0.028 0.059 0.0623 ...
$ col :List of 4
..$ coord  : num [1:12, 1:4] -0.4783 -0.3543 0.836 -0.0692 0.1046 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:12] "ORLY" "ALEZ" "CORS" "DIRE" ...
.. .. ..$ : chr [1:4] "Dim 1" "Dim 2" "Dim 3" "Dim 4"
..$ contrib: num [1:12, 1:4] 9.05 7.129 37.569 0.321 0.461 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:12] "ORLY" "ALEZ" "CORS" "DIRE" ...
.. .. ..$ : chr [1:4] "Dim 1" "Dim 2" "Dim 3" "Dim 4"
..$ cos2   : num [1:12, 1:4] 0.46552 0.14828 0.63007 0.00873 0.03995 ...
.. ..- attr(*, "dimnames")=List of 2
.. .. ..$ : chr [1:12] "ORLY" "ALEZ" "CORS" "DIRE" ...
.. .. ..$ : chr [1:4] "Dim 1" "Dim 2" "Dim 3" "Dim 4"
..$ inertia: num [1:12] 0.033 0.0817 0.1013 0.0624 0.0196 ...
$ svd :List of 3
..$ vs: num [1:10] 0.412 0.366 0.319 0.275 0.239 ...
..$ U : num [1:11, 1:4] 0.495 -1.164 -1.038 1.495 -0.27 ...
..$ V : num [1:12, 1:4] -1.161 -0.86 2.029 -0.168 0.254 ...
- attr(*, "class")= chr [1:2] "CA" "list"

```

L'accès aux différents objets de mydata.ca se font par le caractère '\$' : mydata.ca\$eig pour les valeurs propres, mydata.ca\$row pour les résultats sur les lignes, mydata.ca\$col pour les résultats sur les colonnes.

## 3<sup>ème</sup> PARTIE : ANALYSE DES RESULTATS

### 1. Analyse des valeurs propres

```
> mydata.ca$eig
```

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.1698450535	29.43385786	29.43386
dim 2	0.1339617724	23.21534650	52.64920
dim 3	0.1018198712	17.64520989	70.29441
dim 4	0.0758603111	13.14646243	83.44088
dim 5	0.0572008706	9.91281325	93.35369
dim 6	0.0210353403	3.64538858	96.99908
dim 7	0.0090835971	1.57417188	98.57325
dim 8	0.0052824869	0.91544597	99.48870
dim 9	0.0026878540	0.46580051	99.95450
dim 10	0.0002625712	0.04550313	100.00000

### 2. Analyse des 4 premiers axes en fonction des marques

```
> mydata.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 0.01589445 0.03592797 0.02853545
[10] 0.04599551 0.04402548 0.07264213
```

### 3. Analyse des 4 premiers axes en fonction des attributs

```
> mydata.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
vu!g	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
vu!g	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
```

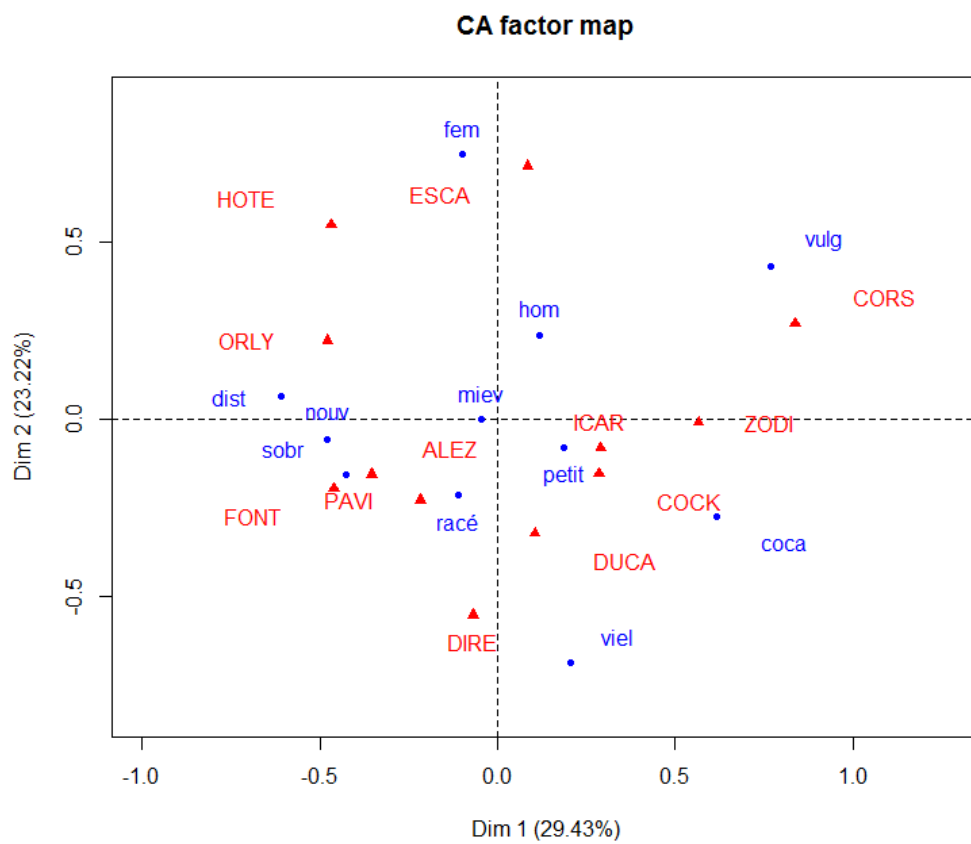
	Dim 1	Dim 2	Dim 3	Dim 4
viel	0.044665918	5.072125e-01	0.3766993824	0.0267187758
nouv	0.402265070	5.389262e-03	0.0072648037	0.0001917577
sobr	0.636961579	8.683777e-02	0.1514038925	0.0117821367
coca	0.598084107	1.176007e-01	0.0103995642	0.1555451941
racé	0.017020861	6.228778e-02	0.7434356179	0.1074836571
miev	0.009082806	7.533701e-07	0.0349458467	0.3888370922
dist	0.639686995	6.927488e-03	0.1379362616	0.0021723345
vu!g	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 0.05556498 0.07546057 0.02940784
[10] 0.07200929 0.03590458
```

#### 4. Premier plan factoriel

```
> mydata.ca = CA(mydata, ncp=2, graph=TRUE)
```





## 4<sup>ème</sup> PARTIE : AVEC LES ATTRIBUTS CORSAIRE, DISTINGUE, VULGAIRE EN ILLUSTRATIF

Mettre des colonnes en illustratif se fait grâce à l'option « col.sup ». Respectivement, mettre des lignes en illustratifs se fait grâce à l'option « row.sup »

```
> mydata.ca = CA(mydata, ncp=4, graph=TRUE, col.sup=3, row.sup=c(7,8))
```

Analyser ces nouveaux résultats

Pour aller plus loin : `> ?CA` ou `> help(CA)`