

Analyse factorielle des correspondances avec R

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

Références :

1. L. Lebart, A. Morineau, M. Piron, « Statistique Exploratoire Multidimensionnelle », Dunod, 2000 ; pages 67 à 103, partie théorique ; pages 103 à 107, pour l'exemple que nous traiterons.
2. Tutoriels Tanagra, « AFC – Association Médias et Professions », <http://tutoriels-data-mining.blogspot.com/2008/03/afc-association-mdias-et-professions.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'AFC (méthode CA).
4. D'autres packages réalisent l'AFC (corresp avec MASS, dudi.coa avec ADE4)

Objectif de l'étude

Association entre types de médias et professions

Objectifs de l'étude

Ce tutoriel reproduit sous le logiciel R, l'analyse menée dans l'ouvrage de Lebart et al., pages 103 à 107. Les justifications théoriques et les formules sont disponibles dans le même ouvrage, pages 67 à 103.

Les mêmes calculs ont été reproduits dans Tanagra. D'autres packages de R peuvent réaliser également une AFC, 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 AFC 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 points lignes
- Calculer les coordonnées factorielles, les \cos^2 et les contributions des points colonnes
- Projeter les points dans le premier plan factoriel (représentation simultanée)

Types de Médias



Prof	Radio	Tel.	Quot.Nat.	Quot.Reg.	Press.Mag.	Press.TV
Agriculteur	96	118	2	71	50	17
Petit.Patr.	122	136	11	76	49	41
Prof.Cad.Sup	193	184	74	63	103	79
Prof.Int.	360	365	63	145	141	184
Employe	511	593	57	217	172	306
Ouvr.Qualif.	385	457	42	174	104	220
Ouvr.Non-Qual.	156	185	8	69	42	85
Inactif	1474	1931	181	852	642	782



Professions

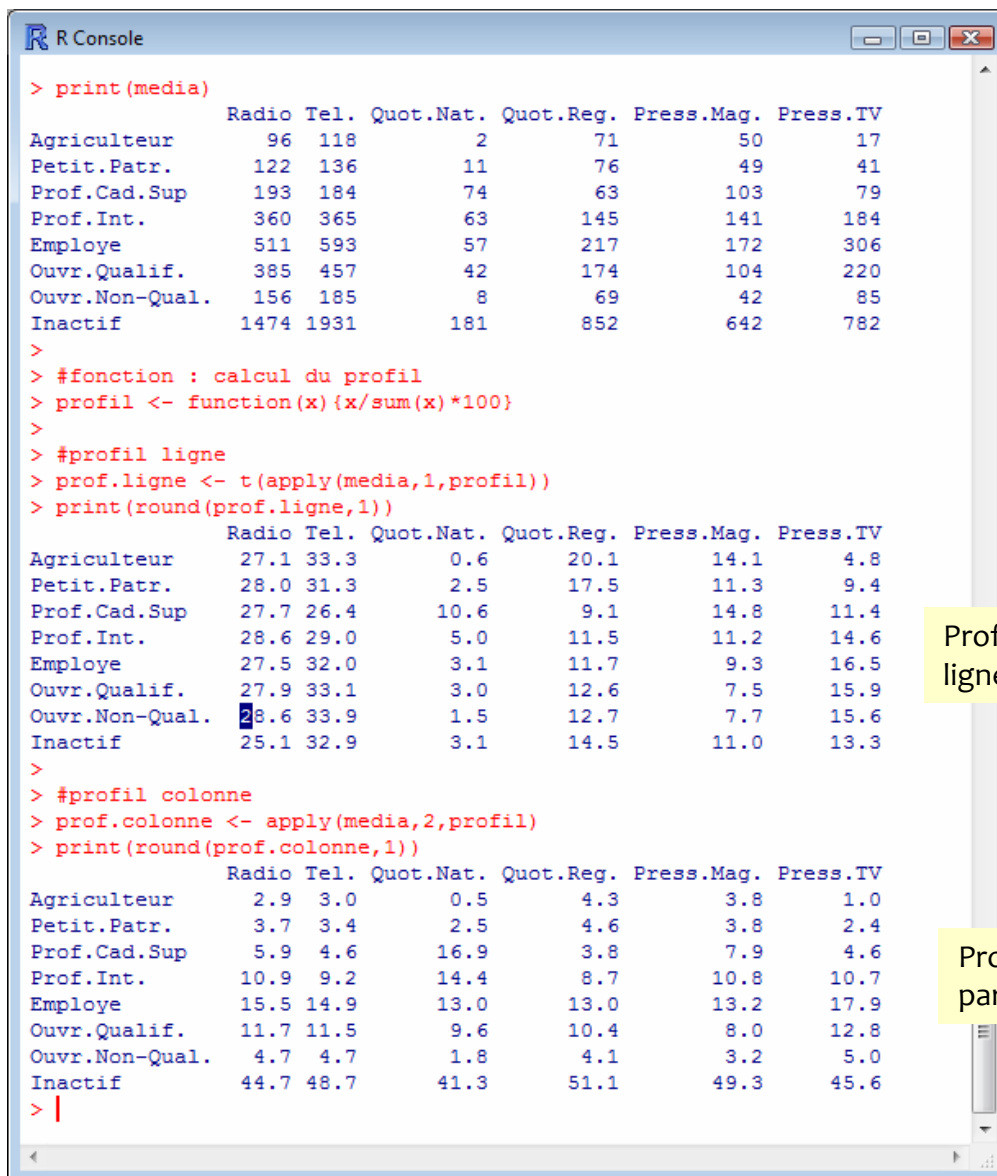
Tableau de données

Fichier de données

Importation des données, valeurs et profils

```
#changement de répertoire
setwd("D:/_Travaux/university/Cours_Universite/Supports_de_cours/Info
rmatique/R/Tutoriels/afc")
#chargement des données
library(xlsReadWrite)
media <- read.xls(file="media_prof_afc.xls",rowNames=T, sheet=1)
print(media)

#fonction : calcul du profil
profil <- fonction(x){x/sum(x)*100}
#profil ligne
prof.ligne <- t(apply(media,1,profil))
print(round(prof.ligne,1))
#profil colonne
prof.colonne <- apply(media,2,profil)
print(round(prof.colonne,1))
```



```
> print(media)
      Radio Tel. Quot.Nat. Quot.Reg. Press.Mag. Press.TV
Agriculteur    96  118         2         71         50         17
Petit.Patr.   122  136         11         76         49         41
Prof.Cad.Sup  193  184         74         63        103         79
Prof.Int.     360  365         63        145        141        184
Employe      511  593         57        217        172        306
Ouvr.Qualif.  385  457         42        174        104        220
Ouvr.Non-Qual. 156  185          8         69         42         85
Inactif     1474 1931        181        852        642        782
>
> #fonction : calcul du profil
> profil <- fonction(x){x/sum(x)*100}
>
> #profil ligne
> prof.ligne <- t(apply(media,1,profil))
> print(round(prof.ligne,1))
      Radio Tel. Quot.Nat. Quot.Reg. Press.Mag. Press.TV
Agriculteur   27.1 33.3      0.6     20.1     14.1      4.8
Petit.Patr.   28.0 31.3      2.5     17.5     11.3      9.4
Prof.Cad.Sup  27.7 26.4     10.6      9.1     14.8     11.4
Prof.Int.     28.6 29.0      5.0     11.5     11.2     14.6
Employe      27.5 32.0      3.1     11.7      9.3     16.5
Ouvr.Qualif. 27.9 33.1      3.0     12.6      7.5     15.9
Ouvr.Non-Qual. 28.6 33.9      1.5     12.7      7.7     15.6
Inactif      25.1 32.9      3.1     14.5     11.0     13.3
>
> #profil colonne
> prof.colonne <- apply(media,2,profil)
> print(round(prof.colonne,1))
      Radio Tel. Quot.Nat. Quot.Reg. Press.Mag. Press.TV
Agriculteur   2.9  3.0      0.5      4.3      3.8      1.0
Petit.Patr.   3.7  3.4      2.5      4.6      3.8      2.4
Prof.Cad.Sup   5.9  4.6     16.9      3.8      7.9      4.6
Prof.Int.     10.9  9.2     14.4      8.7     10.8     10.7
Employe      15.5 14.9     13.0     13.0     13.2     17.9
Ouvr.Qualif.  11.7 11.5      9.6     10.4      8.0     12.8
Ouvr.Non-Qual. 4.7  4.7      1.8      4.1      3.2      5.0
Inactif      44.7 48.7     41.3     51.1     49.3     45.6
> |
```

Profils lignes : somme par ligne = 100.

Profils colonnes : somme par colonne = 100.

AFC avec R via le package FactoMineR

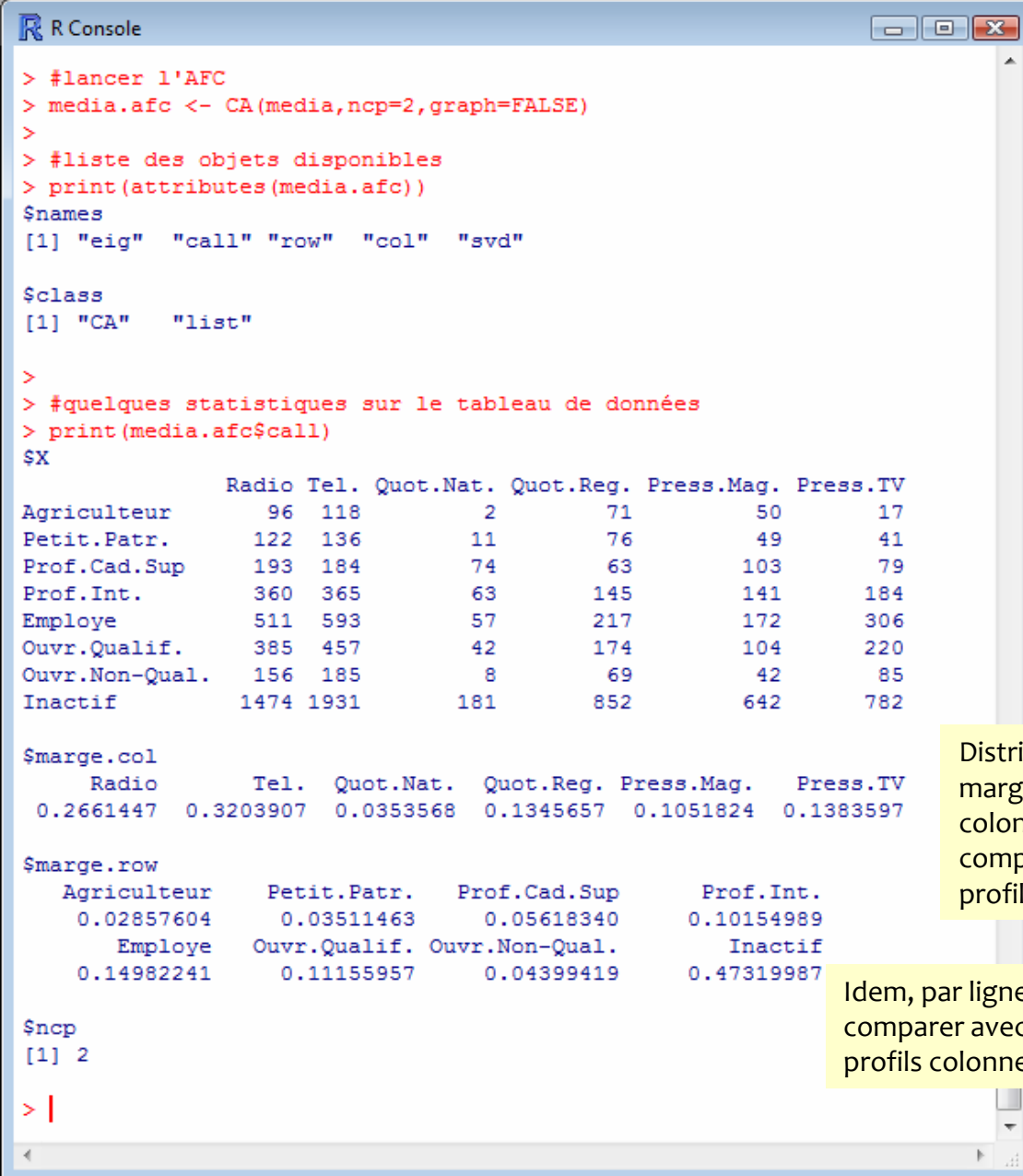
Calculs et quelques statistiques

```
#charger le package
library(FactoMineR)

#lancer l'AFC
media.afc <- CA(media,ncp=2,graph=FALSE)

#liste des objets disponibles
print(attributes(media.afc))

#quelques statistiques sur le tableau de données
print(media.afc$call)
```



```
> #lancer l'AFC
> media.afc <- CA(media,ncp=2,graph=FALSE)
>
> #liste des objets disponibles
> print(attributes(media.afc))
$names
[1] "eig" "call" "row" "col" "svd"

$class
[1] "CA" "list"

>
> #quelques statistiques sur le tableau de données
> print(media.afc$call)
$X
      Radio Tel. Quot.Nat. Quot.Reg. Press.Mag. Press.TV
Agriculteur    96  118      2      71      50      17
Petit.Patr.   122  136     11     76     49     41
Prof.Cad.Sup  193  184     74     63    103     79
Prof.Int.     360  365     63    145    141    184
Employe       511  593     57    217    172    306
Ouvr.Qualif.  385  457     42    174    104    220
Ouvr.Non-Qual. 156  185      8     69     42     85
Inactif      1474 1931    181    852    642    782

$marge.col
      Radio      Tel.  Quot.Nat.  Quot.Reg.  Press.Mag.  Press.TV
0.2661447 0.3203907 0.0353568 0.1345657 0.1051824 0.1383597

$marge.row
      Agriculteur  Petit.Patr.  Prof.Cad.Sup  Prof.Int.
0.02857604      0.03511463      0.05618340      0.10154989
      Employe  Ouvr.Qualif.  Ouvr.Non-Qual.  Inactif
0.14982241      0.11155957      0.04399419      0.47319987

$ncp
[1] 2

> |
```

Distribution marginale par colonne, on peut comparer avec les profils lignes.

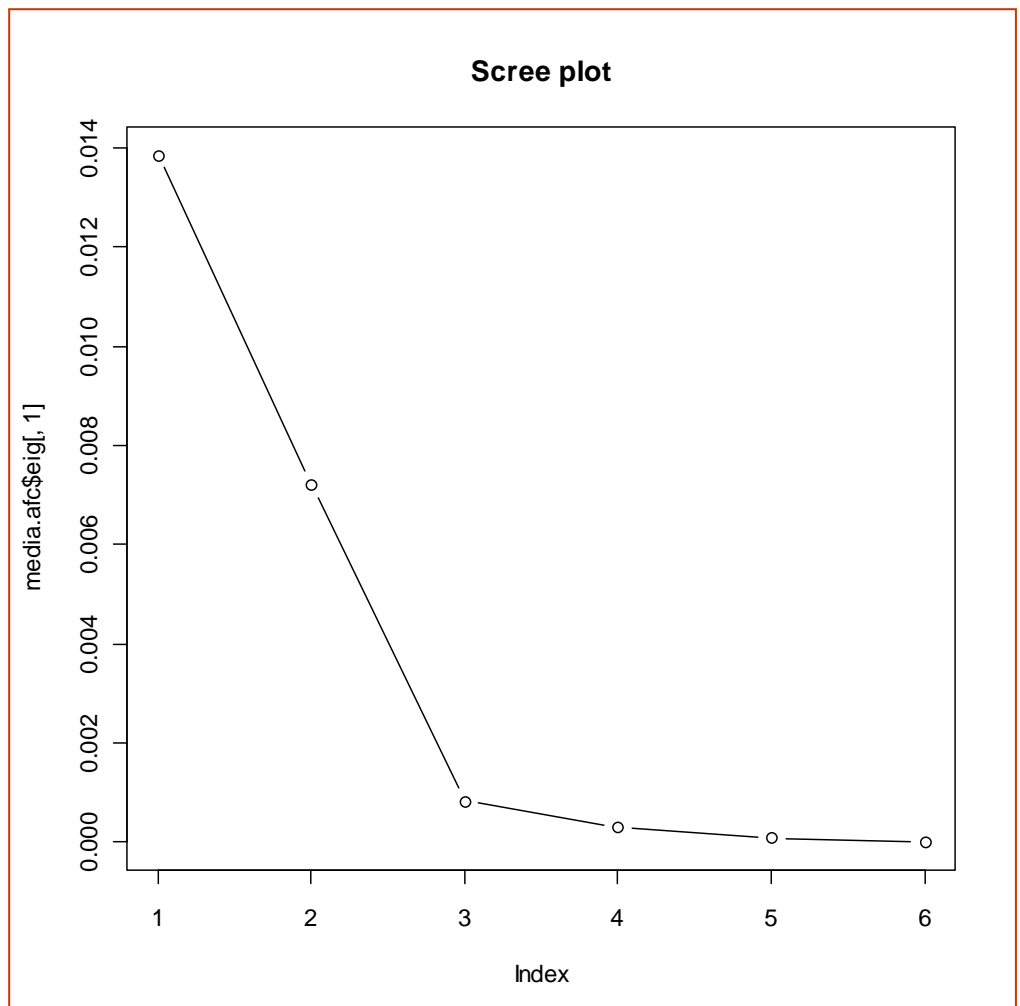
Idem, par ligne. A comparer avec les profils colonnes.

AFC avec R via le package FactoMineR

Valeurs propres et Scree plot

```
#tableau des valeurs propres et Scree plot  
print(media.afc$eig)  
plot(media.afc$eig[,1],type="b",main="Scree plot")
```

```
R Console  
> plot(media.afc$eig[,1],type="b",main="Scree plot")  
> print(media.afc$eig)  
      eigenvalue percentage of variance cumulative percentage of variance  
dim 1 1.385729e-02      6.219818e+01      62.19818  
dim 2 7.210685e-03      3.236503e+01      94.56321  
dim 3 8.247314e-04      3.701792e+00      98.26500  
dim 4 3.038516e-04      1.363833e+00      99.62883  
dim 5 8.269302e-05      3.711661e-01      100.00000  
dim 6 3.174973e-35      1.425081e-31      100.00000  
> |
```



Les deux premiers facteurs traduisent près de 95% de l'information disponible

AFC avec R via le package FactoMineR

Coordonnées, cosinus² et contributions des lignes

```
#coordonnées, contributions et cos2 - lignes  
print(media.afc$row)
```

```
R Console  
> #coordonnées, contributions et cos2 - lignes  
> print(media.afc$row)  
$coord  
          Dim 1      Dim 2  
Agriculteur -0.16614986  0.30961174  
Petit.Patr. -0.06837709  0.14315064  
Prof.Cad.Sup  0.42997558  0.06087258  
Prof.Int.    0.10660251 -0.03259679  
Employe     -0.01572857 -0.09546893  
Ouvr.Qualif. -0.04371252 -0.10137860  
Ouvr.Non-Qual. -0.11779685 -0.09486419  
Inactif     -0.03258370  0.03339524  
  
$contrib  
          Dim 1      Dim 2  
Agriculteur  5.6927723 37.989221  
Petit.Patr.  1.1847619 9.979255  
Prof.Cad.Sup 74.9579010 2.887188  
Prof.Int.    8.3279108 1.496416  
Employe     0.2674713 18.937575  
Ouvr.Qualif. 1.5382978 15.900944  
Ouvr.Non-Qual. 4.4053909 5.490645  
Inactif     3.6254941 7.318755  
  
$cos2  
          Dim 1      Dim 2  
Agriculteur  0.21351182 0.74140722  
Petit.Patr.  0.15383515 0.67424994  
Prof.Cad.Sup  0.97818335 0.01960542  
Prof.Int.    0.80220729 0.07500692  
Employe     0.02521427 0.92894899  
Ouvr.Qualif. 0.13827479 0.74374460  
Ouvr.Non-Qual. 0.55568018 0.36038123  
Inactif     0.37220445 0.39097583  
  
$inertia  
Agriculteur  Petit.Patr.  Prof.Cad.Sup  Prof.Int.  Employe  
0.003694708  0.001067219  0.010618798  0.001438559  0.001469972  
Ouvr.Qualif. Ouvr.Non-Qual.  Inactif  
0.001541614  0.001098595  0.001349783  
  
> |
```

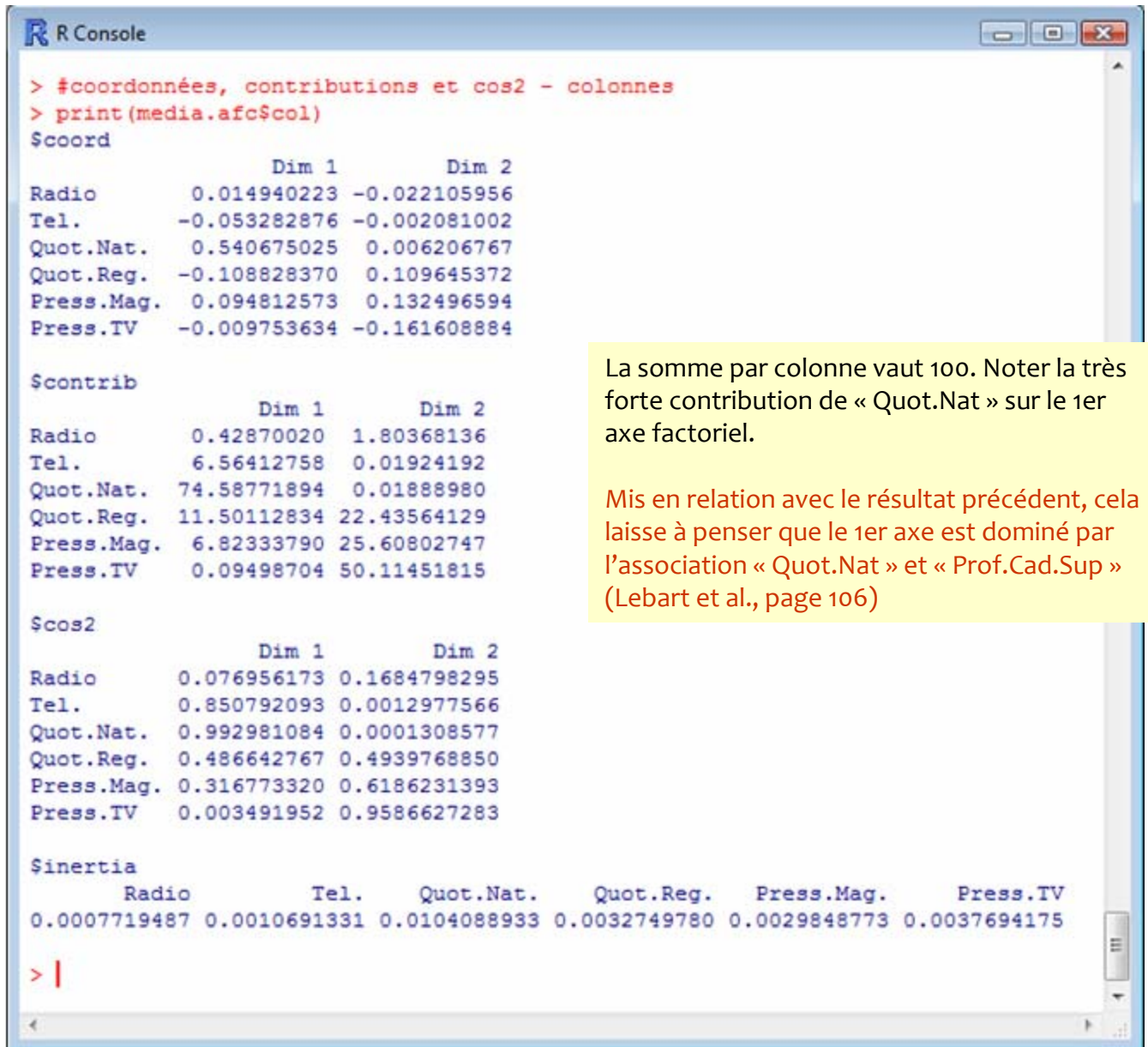
La somme par colonne vaut 100.
Noter la très forte contribution de « Prof.Cad.Sup » sur le 1er axe factoriel.

La somme par ligne, si on prend toutes les dimensions (6 dans notre exemple) vaut 100.

AFC avec R via le package FactoMineR

Coordonnées, cosinus² et contributions des colonnes

```
#coordonnées, contributions et cos2 - colonnes  
print(media.afc$col)
```



```
> #coordonnées, contributions et cos2 - colonnes  
> print(media.afc$col)  
$coord  
          Dim 1      Dim 2  
Radio      0.014940223 -0.022105956  
Tel.       -0.053282876 -0.002081002  
Quot.Nat.  0.540675025  0.006206767  
Quot.Reg.  -0.108828370  0.109645372  
Press.Mag.  0.094812573  0.132496594  
Press.TV   -0.009753634 -0.161608884  
  
$contrib  
          Dim 1      Dim 2  
Radio      0.42870020  1.80368136  
Tel.       6.56412758  0.01924192  
Quot.Nat. 74.58771894  0.01888980  
Quot.Reg. 11.50112834 22.43564129  
Press.Mag. 6.82333790 25.60802747  
Press.TV   0.09498704 50.11451815  
  
$cos2  
          Dim 1      Dim 2  
Radio      0.076956173 0.1684798295  
Tel.       0.850792093 0.0012977566  
Quot.Nat.  0.992981084 0.0001308577  
Quot.Reg.  0.486642767 0.4939768850  
Press.Mag.  0.316773320 0.6186231393  
Press.TV   0.003491952 0.9586627283  
  
$inertia  
      Radio      Tel.      Quot.Nat.      Quot.Reg.      Press.Mag.      Press.TV  
0.0007719487 0.0010691331 0.0104088933 0.0032749780 0.0029848773 0.0037694175  
> |
```

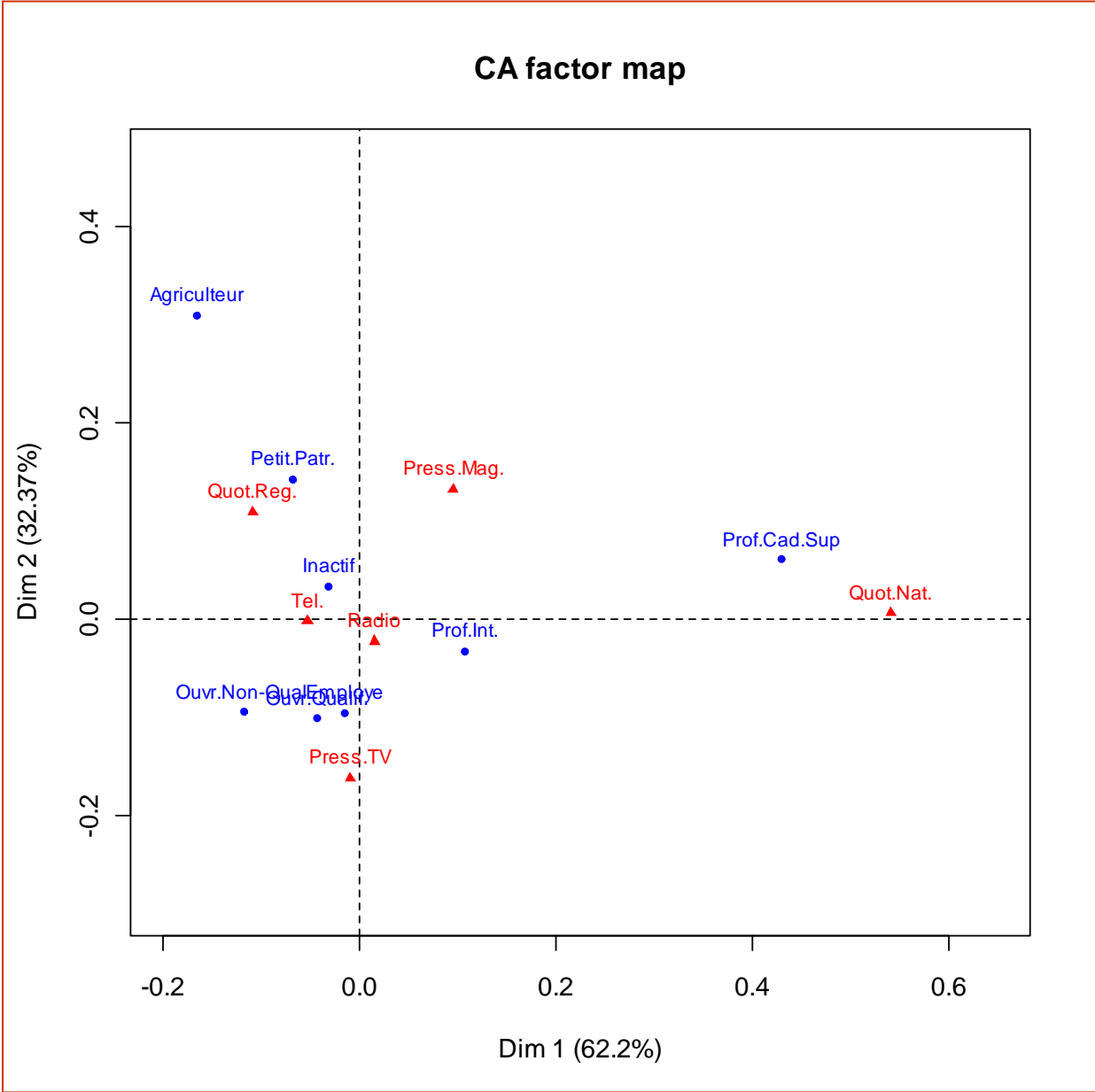
La somme par colonne vaut 100. Noter la très forte contribution de « Quot.Nat » sur le 1er axe factoriel.

Mis en relation avec le résultat précédent, cela laisse à penser que le 1er axe est dominé par l'association « Quot.Nat » et « Prof.Cad.Sup » (Lebart et al., page 106)

AFC avec R via le package FactoMineR

Représentation simultanée dans le 1er plan factoriel

```
#graphique  
plot(media.afc,cex=0.75)
```



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