

Rapport de l'exercice 2 du TP2 de statistique exploratoire : Analyse en Composante Principales

Rémy Gaudré

07/10/2019

Question 1 : Importer la base de données sous R

```
donnees = read.csv("cesp.csv",header = TRUE,sep=";", dec=",")
row.names(donnees)=donnees[,1]
```

Question 2 : Mettre en oeuvre l'AFC en utilisant les lignes actives pour le calcul des valeurs et vecteurs propres.

```
donnees.ca = CA(donnees[, -1], graph = FALSE, row.sup = 9:19)
```

Tableau des vecteurs propres aux lignes actives

```
donnees.ca$svd$U
```

##	[,1]	[,2]	[,3]	[,4]	[,5]
## [1,]	-1.4114350	3.6461049	-2.5014897	0.8279671	-2.3577219
## [2,]	-0.5808601	1.6857960	-2.2432411	-1.1686033	2.8247148
## [3,]	3.6526216	0.7168586	-0.1052033	-0.9644214	-1.2352883
## [4,]	0.9055831	-0.3838721	-1.0759103	1.4533042	1.3146994
## [5,]	-0.1336135	-1.1242782	-0.1771287	1.1642512	-0.3806739
## [6,]	-0.3713358	-1.1938727	-0.6728622	-2.0314725	0.1902564
## [7,]	-1.0006785	-1.1171566	-1.3833760	-0.2003914	-2.4725589
## [8,]	-0.2767969	0.3932750	0.9042375	-0.0317160	0.1028476

Tableau des vecteurs propres aux colonnes actives

```
donnees.ca$svd$V
```

##	[,1]	[,2]	[,3]	[,4]	[,5]
## [1,]	0.12691647	-0.26032809	-1.6262109	0.1527101	-0.07479529
## [2,]	-0.45263544	-0.02450667	0.5729089	-0.6435705	-1.08319073
## [3,]	4.59300804	0.07309323	0.7152980	-2.2673403	0.72770609
## [4,]	-0.92449171	1.29122535	0.1820554	-0.7481705	1.82111577
## [5,]	0.80542820	1.56032998	0.6531925	2.1977348	-0.40895964
## [6,]	-0.08285665	-1.90316731	0.9450671	0.8328431	1.00590620

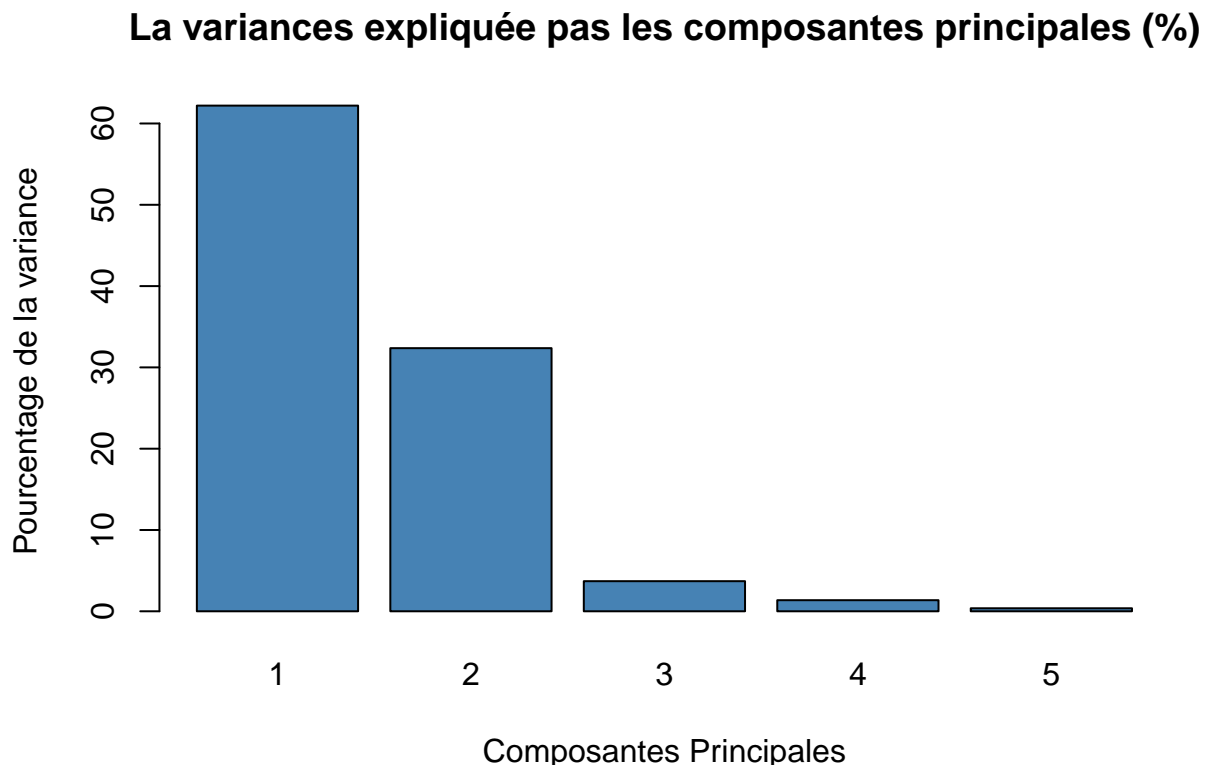
Question 3 : Combien d'axes peut-on retenir et pourquoi ?

```
donnees.ca$eig
```

##	eigenvalue	percentage of variance
## dim 1	1.385729e-02	62.1981806
## dim 2	7.210685e-03	32.3650288
## dim 3	8.247314e-04	3.7017919
## dim 4	3.038516e-04	1.3638326
## dim 5	8.269302e-05	0.3711661

```
##      cumulative percentage of variance
## dim 1                      62.19818
## dim 2                      94.56321
## dim 3                      98.26500
## dim 4                      99.62883
## dim 5                      100.00000
```

```
barplot(donnees.ca$eig[, 2],
        names.arg = 1:nrow(donnees.ca$eig),
        main = "La variances expliquée pas les composantes principales (%)",
        xlab = "Composantes Principales",
        ylab = "Pourcentage de la variance",
        col = "steelblue")
```



```
mean(donnees.ca$eig[, 1])
```

```
## [1] 0.00445585
```

D'après le critère des 80% de variance expliquée : nous avons 2 composantes principales à retenir.

Le critère du coude nous donne aussi 2 composantes à retenir.

D'après le critère de Kaiser, nous avons aussi 2 composantes principales à retenir (Moyenne des valeurs propres = 0.0045).

Question 4 : Dessiner les tableaux des Poids, Distances, Coordonnées, Contributions et Cosinus carrés sur les trois premiers axes :

Poids des lignes

```
donnees.ca$call$marge.row
```

```
##      AGRI      PEPA      PRCS      PRIN      EMPL      OUQU
## 0.02857604 0.03511463 0.05618340 0.10154989 0.14982241 0.11155957
```

```
##      ONQU      INAC
## 0.04399419 0.47319987
```

Poids des colonnes

```
donnees.ca$call$marge.col
```

```
##      radio      tel      QuotN      QuotR      Pmag      PMTV
## 0.2661447 0.3203907 0.0353568 0.1345657 0.1051824 0.1383597
```

a. Des colonnes actives

```
donnees.ca$col$coord[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## radio  0.014940223 -0.022105956 -0.046701748
## tel    -0.053282876 -0.002081002  0.016452877
## QuotN  0.540675025  0.006206767  0.020542027
## QuotR -0.108828370  0.109645372  0.005228292
## Pmag   0.094812573  0.132496594  0.018758472
## PMTV  -0.009753634 -0.161608884  0.027140568
```

```
donnees.ca$col$cos2[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## radio 0.076956173 0.1684798295 0.751961487
## tel   0.850792093 0.0012977566 0.081120728
## QuotN 0.992981084 0.0001308577 0.001433359
## QuotR 0.486642767 0.4939768850 0.001123171
## Pmag  0.316773320 0.6186231393 0.012399714
## PMTV  0.003491952 0.9586627283 0.027037917
```

```
donnees.ca$col$contrib[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## radio  0.42870020  1.80368136 70.3836037
## tel     6.56412758  0.01924192 10.5160124
## QuotN 74.58771894  0.01888980  1.8090350
## QuotR 11.50112834 22.43564129  0.4460069
## Pmag   6.82333790 25.60802747  4.4877188
## PMTV   0.09498704 50.11451815 12.3576232
```

b. des lignes actives

```
donnees.ca$row$coord[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## AGRI -0.16614986  0.30961174 -0.071838125
## PEPA -0.06837709  0.14315064 -0.064421704
## PRCS  0.42997558  0.06087258 -0.003021242
## PRIN  0.10660251 -0.03259679 -0.030898140
## EMPL -0.01572857 -0.09546893 -0.005086806
## OUQU -0.04371252 -0.10137860 -0.019323350
## ONQU -0.11779685 -0.09486419 -0.039727982
## INAC -0.03258370  0.03339524  0.025968017
```

```
donnees.ca$row$cos2[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## AGRI 0.21351182 0.74140722 3.991461e-02
## PEPA 0.15383515 0.67424994 1.365522e-01
## PRCS 0.97818335 0.01960542 4.829518e-05
## PRIN 0.80220729 0.07500692 6.739326e-02
## EMPL 0.02521427 0.92894899 2.637291e-03
## OUQU 0.13827479 0.74374460 2.702067e-02
## ONQU 0.55568018 0.36038123 6.320488e-02
## INAC 0.37220445 0.39097583 2.364059e-01
```

```
donnees.ca$row$contrib[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## AGRI 5.6927723 37.989221 17.88131756
## PEPA 1.1847619 9.979255 17.67013832
## PRCS 74.9579010 2.887188 0.06218227
## PRIN 8.3279108 1.496416 11.75524260
## EMPL 0.2674713 18.937575 0.47006138
## OUQU 1.5382978 15.900944 5.05078817
## ONQU 4.4053909 5.490645 8.41929638
## INAC 3.6254941 7.318755 38.69097331
```

c. des lignes illustratives (supplémentaires)

```
donnees.ca$row.sup$coord[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## HOM 0.0498239399 0.02386212 -0.0105270206
## FEM -0.0482989506 -0.02224294 0.0096108671
## AGE1 0.0163876318 -0.10007314 -0.0368847177
## AGE2 0.0301220698 -0.12076037 -0.0129330501
## AGE3 0.0306359760 -0.01252374 -0.0105350337
## AGE4 -0.0244521405 0.10024751 -0.0037569314
## AGE5 -0.0677531954 0.13652490 0.0703892109
## PRIM -0.1290407935 0.08046587 0.0235012904
## SECO -0.0006251352 -0.04102442 0.0004550801
## TECP 0.0311727618 -0.18419209 -0.0370538476
## SUPE 0.2905405017 0.01678085 -0.0141765580
```

```
donnees.ca$row.sup$cos2[,1:3]
```

```
##           Dim 1      Dim 2      Dim 3
## HOM 0.4813256748 0.110403054 2.148690e-02
## FEM 0.4862488256 0.103125885 1.925344e-02
## AGE1 0.0150399553 0.560853164 7.619160e-02
## AGE2 0.0541545301 0.870390268 9.983138e-03
## AGE3 0.6140366390 0.102612172 7.261105e-02
## AGE4 0.0477766611 0.803025309 1.127843e-03
## AGE5 0.1438329318 0.584012999 1.552426e-01
## PRIM 0.6289413984 0.244556698 2.086121e-02
## SECO 0.0001595581 0.687156914 8.455642e-05
## TECP 0.0132161317 0.461419300 1.867327e-02
## SUPE 0.9881829928 0.003296492 2.352695e-03
```

Question 5 : Interpréter les axes retenus à la question 3

Sur l'axe 1 : Les quotidiens Régionaux et Nationaux sont les colonnes contribuant le plus à la création du premier axe. La télévision et les quotidiens Nationaux sont très bien représentés sur la première composante. Les quotidiens Régionaux sont bien représentés sur la première composante. Les quotidiens Nationaux sont inversement corrélés à la télévision et aux quotidiens régionaux.

Les professions libérales, et les cadres supérieurs est la ligne qui contribue le plus la création du premier axe. Les professions libérales et cadres supérieurs et les professions intermédiaires sont très bien représentés sur le premier axe. Les ouvriers qualifiés sont bien représentés sur le premier axe. Les professions libérales et cadres supérieurs et les professions intermédiaires sont inversement corrélées aux ouvriers qualifiés.

Pour les lignes supplémentaires, le supérieur est très bien représenté sur l'axe 1. Les hommes, les femmes, les personnes entre 35 et 49 ans et la primaire sont bien représentés. Les hommes, les personnes entre 35 et 49 ans et le supérieur sont inversement corrélés avec les femmes et la primaire.

Sur l'axe 2 : Les quotidiens Régionaux, la Presse Magazine et la Presse Magazine T.V. sont les colonnes qui contribuent le plus à la création de la seconde composante principale. La Presse Magazine T.V. est très bien représentée. Les quotidiens Régionaux et la Presse Magazine sont bien représentés. La Presse Magazine T.V. est inversement corrélée avec les quotidiens Régionaux et la Presse Magazine.

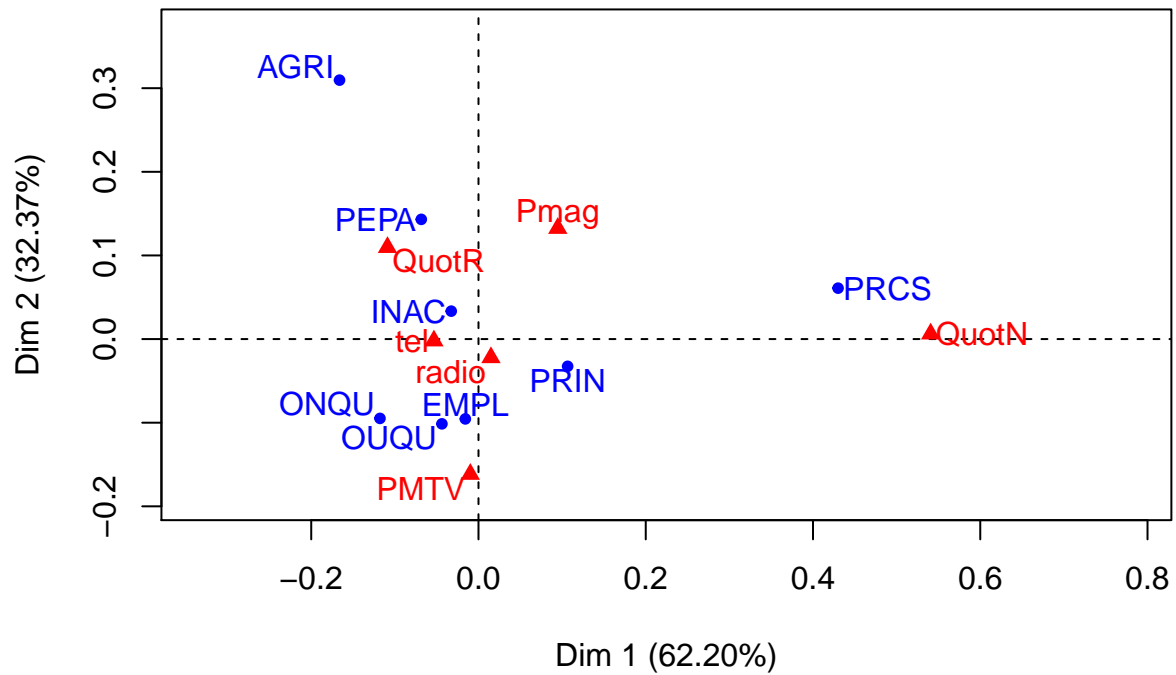
Les agriculteurs, les employés et les ouvriers qualifiés sont les professions contribuant le plus à la création de l'axe 2. Les agriculteurs, les petits patrons, les employés et les ouvriers qualifiés sont bien représentés sur l'axe 2. Les agriculteurs et les petits patrons sont inversement corrélés aux employés et aux ouvriers qualifiés.

Pour les lignes supplémentaires, les 15-24 ans, les 25-34 ans, les 50-64 ans, les 65 ans ou plus et le secondaire sont bien représentés sur le second axe. Les 15-24 ans, les 25-34 ans et le secondaire sont inversement corrélés aux 50-64 ans, et aux 65 ans ou plus.

Question 6 : Projeter dans le premier plan factoriel les lignes actives et les colonnes actives et interpréter

```
plot.CA(donnees.ca, axes=c(1,2), invisible = "row.sup")
```

CA factor map

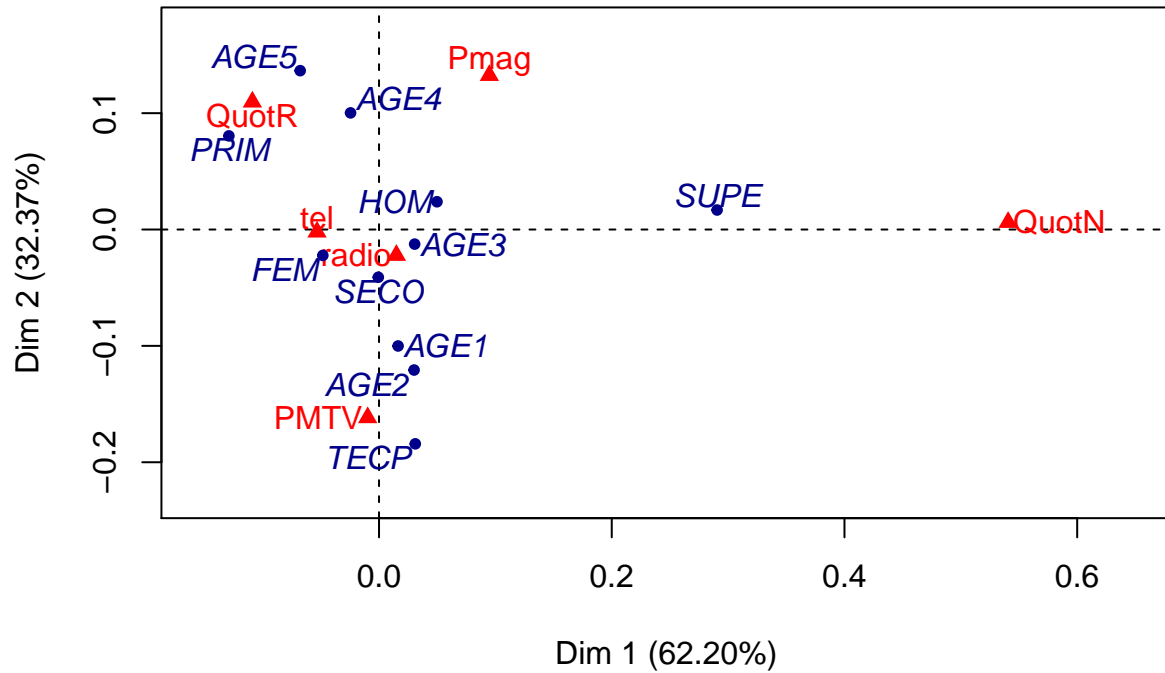


Les ouvriers non-qualifiés, les ouvriers qualifiés, les employés sont corrélés. Ils ont plus tendance à consulter la presse magazine T.V., la radio ou le téléphone. Les professions intermédiaires ont un profil qui se rapproche de la qualité d'employé, ils sont corrélés avec la radio. Les inactifs sont plus susceptibles d'utiliser le téléphone et la radio. Les professions libérales et les cadres supérieurs sont peu corrélés avec les autres types de profession. Ils sont corrélés avec les quotidiens nationaux. Les petits patrons sont bien corrélés avec les quotidiens régionaux. Les agriculteurs sont quant à eux peu corrélés avec d'autres types de profession et les médias.

Question 7 : Projeter dans le premier plan factoriel les lignes supplémentaires et interpréter

```
plot.CA(donnees.ca, axes=c(1,2), invisible = c("row"))
```

CA factor map



Le niveau primaire, et les 50 ans et + sont corrélés. Le secondaire, et les 15-49 ans sont corrélés. Le supérieur est aberrant sur l'axe 1, il n'est pas corrélé avec d'autre type de personne. Le TECP se rapproche des 25-34 ans. Les hommes et femmes sont inversement corrélés sur les 2 axes. Les hommes sont corrélés avec les 35-49 ans tandis que les femmes sont corrélés avec le secondaire. Le supérieur est à mi-chemin avec le groupe et les quotidiens nationaux (presque des cadres ?). Le TECP est très corrélé avec la presse magazine T.V.. La radio et le téléphone sont corrélés avec les femmes, les hommes, le secondaire et les 35-49 ans. Les 50-64 ans sont ceux qui se rapprochent le plus de la presse magazine. Ils se rapprochent aussi des Quotidiens régionaux avec les 65 ans ou +.