

Initiation à l'analyse des correspondances multiples

A.B. Dufour & D. Clot

Présentation de l'analyse des correspondances multiples à travers un exemple simple, suivi de cas réels.

Table des matières

| | | |
|----------|---------------------------------------------------------|-----------|
| 1 | Un exemple simple pour commencer ... | 2 |
| 1.1 | Les données | 2 |
| 1.2 | L'analyse des correspondances multiples (ACM) | 2 |
| 1.2.1 | La mise en oeuvre | 2 |
| 1.2.2 | Les valeurs retournées par l'ACM | 3 |
| 1.2.3 | Les graphiques | 6 |
| 2 | Un exemple réel pour poursuivre ... | 9 |
| 2.1 | Les données | 9 |
| 2.2 | Quelques questions autour de ces données | 12 |
| 2.3 | L'analyse des correspondances multiples | 15 |
| 3 | Une autre étude pour finir ... | 18 |

1 Un exemple simple pour commencer ...

1.1 Les données

Les données sont extraites du data frame `banque` de la librairie `ade4`. 26 clients ont été sélectionnés et trois variables retenues parmi l'ensemble des possibles :

1. l'âge avec deux modalités : 45 et 75 ans ($m_1 = 2$),
2. l'épargne sur livret avec trois modalités : nulle, faible et forte ($m_2 = 3$),
3. le prélèvement par le trésor public avec trois modalités : nul, faible, moyen ($m_3 = 3$).

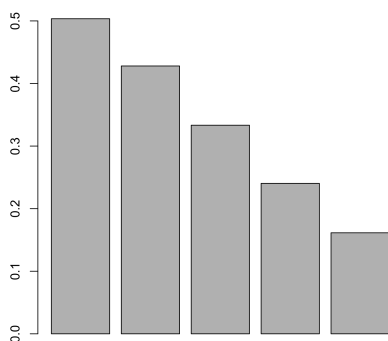
```
bank26 <- read.table("http://pbil.univ-lyon1.fr/R/donnees/bank26.txt",h=T)
summary(bank26)
  age      livret      impot
a45:14  faible: 6  faible: 2
a75:12  forte : 2  moyen : 2
        nulle :18  nul   :22
```

1.2 L'analyse des correspondances multiples (ACM)

1.2.1 La mise en oeuvre

Elle nécessite uniquement un data frame comprenant les variables en colonnes et les individus en lignes. Comme toutes les méthodes basées sur les valeurs et les vecteurs propres, le premier élément examiné est la représentation en bâtons des valeurs propres afin de sélectionner le nombre de facteurs à conserver pour l'interprétation de l'analyse.

```
library(ade4)
acmbank <- dudi.acm(bank26,scannf=FALSE)
barplot(acmbank$eig)
```



Le data frame `bank26` contient $n = 26$ individus et $m = 8$ modalités ($m = m_1 + m_2 + m_3$).

La représentation graphique montre cinq valeurs propres correspondant à la somme des modalités de toutes les variables intervenant dans l'analyse moins le nombre de variables soit $m_1 + m_2 + m_3 - 3$ ou écrit autrement $(m_1 - 1) + (m_2 - 1) + (m_3 - 1)$.

```
acmbank$eig/sum(acmbank$eig)
[1] 0.30209961 0.25679903 0.20000000 0.14424602 0.09685534
```

On retiendra les deux premiers facteurs représentant près de 56% de l'information totale contenue dans les données.

1.2.2 Les valeurs retournées par l'ACM

```
names(acmbank)
[1] "tab" "cw" "lw" "eig" "rank" "nf" "c1" "li" "co" "l1" "call"
[12] "cr"
```

acmbank\$tab

Ce data frame contient le tableau analysé c'est-à-dire $Y = XD_m^{-1} - \mathbf{1}_{nm}$ où X est le tableau disjonctif complet et D_m la matrice des pondérations associées à chaque modalité.

```
round(acmbank$tab,4)
  age.a45 age.a75 livret.faible livret.forte livret.nulle impot.faible impot.moyen
1  0.8571 -1.0000      3.3333      -1      -1.0000      -1      -1
2  0.8571 -1.0000      3.3333      -1      -1.0000      -1      -1
3  0.8571 -1.0000      3.3333      -1      -1.0000      12      -1
4  0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
5  0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
6  0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
7  0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
8  0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
9  0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
10 0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
11 0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
12 0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
13 0.8571 -1.0000     -1.0000      -1      0.4444     -1      -1
14 0.8571 -1.0000     -1.0000      -1      0.4444     12      -1
15 -1.0000  1.1667     -1.0000      12     -1.0000     -1      -1
16 -1.0000  1.1667     -1.0000      12     -1.0000     -1      -1
17 -1.0000  1.1667      3.3333     -1     -1.0000     -1      -1
18 -1.0000  1.1667      3.3333     -1     -1.0000     -1      -1
19 -1.0000  1.1667      3.3333     -1     -1.0000     -1      12
20 -1.0000  1.1667     -1.0000     -1      0.4444     -1      -1
21 -1.0000  1.1667     -1.0000     -1      0.4444     -1      -1
22 -1.0000  1.1667     -1.0000     -1      0.4444     -1      -1
23 -1.0000  1.1667     -1.0000     -1      0.4444     -1      -1
24 -1.0000  1.1667     -1.0000     -1      0.4444     -1      -1
25 -1.0000  1.1667     -1.0000     -1      0.4444     -1      -1
26 -1.0000  1.1667     -1.0000     -1      0.4444     -1      12
  impot.nul
1  0.1818
2  0.1818
3  -1.0000
4  0.1818
5  0.1818
6  0.1818
7  0.1818
8  0.1818
9  0.1818
10 0.1818
11 0.1818
12 0.1818
13 0.1818
14 -1.0000
15 0.1818
16 0.1818
17 0.1818
18 0.1818
19 -1.0000
20 0.1818
21 0.1818
22 0.1818
23 0.1818
24 0.1818
25 0.1818
26 -1.0000
```

$$D_m = \text{diag}(X^T D \mathbf{1}_n)$$

```

matX <- as.matrix(acm.disjonctif(bank26))
matD <- diag(1/26,26,26)
mat1n <- rep(1,26)
t(matX)%*%mat1n
  [,1]
age.a45      14
age.a75      12
livret.faible 6
livret.forte  2
livret.nulle 18
impot.faible  2
impot.moyen   2
impot.nul     22

t(matX)%*%matD)%*%mat1n
  [,1]
age.a45      0.53846154
age.a75      0.46153846
livret.faible 0.23076923
livret.forte  0.07692308
livret.nulle  0.69230769
impot.faible  0.07692308
impot.moyen   0.07692308
impot.nul     0.84615385

frequences <- t(matX)%*%matD)%*%mat1n
matDm <- diag(as.numeric(frequences),8,8)
round(matDm,4)
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,] 0.5385 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
[2,] 0.0000 0.4615 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
[3,] 0.0000 0.0000 0.2308 0.0000 0.0000 0.0000 0.0000 0.0000
[4,] 0.0000 0.0000 0.0000 0.0769 0.0000 0.0000 0.0000 0.0000
[5,] 0.0000 0.0000 0.0000 0.0000 0.6923 0.0000 0.0000 0.0000
[6,] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0769 0.0000 0.0000
[7,] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0769 0.0000
[8,] 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.8462

```

Ce sont donc les fréquences relatives associées à chaque variable.

```

sapply(1:3, function(x) round(summary(bank26[,x])/26,4))
[[1]]
  a45  a75
0.5385 0.4615
[[2]]
faible forte nulle
0.2308 0.0769 0.6923
[[3]]
faible moyen nul
0.0769 0.0769 0.8462

```

Une fois, D_m définie et reconnue, on obtient Y

```

matDmm1 <- diag(1/as.numeric(frequences),8,8)
mat1nm <- matrix(1,nrow=26,ncol=8)
matY <- matX)%*%matDmm1-mat1nm
round(matY,4)
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
1  0.8571 -1.0000 3.3333 -1 -1.0000 -1 -1 0.1818
2  0.8571 -1.0000 3.3333 -1 -1.0000 -1 -1 0.1818
3  0.8571 -1.0000 3.3333 -1 -1.0000 12 -1 -1.0000
4  0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
5  0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
6  0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
7  0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
8  0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
9  0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
10 0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
11 0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
12 0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
13 0.8571 -1.0000 -1.0000 -1 0.4444 -1 -1 0.1818
14 0.8571 -1.0000 -1.0000 -1 0.4444 12 -1 -1.0000
15 -1.0000 1.1667 -1.0000 12 -1.0000 -1 -1 0.1818
16 -1.0000 1.1667 -1.0000 12 -1.0000 -1 -1 0.1818
17 -1.0000 1.1667 3.3333 -1 -1.0000 -1 -1 0.1818

```

```

18 -1.0000 1.1667 3.3333 -1 -1.0000 -1 -1 0.1818
19 -1.0000 1.1667 3.3333 -1 -1.0000 -1 12 -1.0000
20 -1.0000 1.1667 -1.0000 -1 0.4444 -1 -1 0.1818
21 -1.0000 1.1667 -1.0000 -1 0.4444 -1 -1 0.1818
22 -1.0000 1.1667 -1.0000 -1 0.4444 -1 -1 0.1818
23 -1.0000 1.1667 -1.0000 -1 0.4444 -1 -1 0.1818
24 -1.0000 1.1667 -1.0000 -1 0.4444 -1 -1 0.1818
25 -1.0000 1.1667 -1.0000 -1 0.4444 -1 -1 0.1818
26 -1.0000 1.1667 -1.0000 -1 0.4444 -1 12 -1.0000

```

ce qui est bien la matrice `acmbank$tab`. Que peut-on dire des moyennes et des variances de ce tableau ?

```

round(apply(acmbank$tab,2,mean),4)
  age.a45      age.a75 livret.faible livret.forte livret.nulle impot.faible
0          0          0              0           0              0
  impot.moyen  impot.nul
0              0

round(apply(acmbank$tab,2,var)*25/26,4)
  age.a45      age.a75 livret.faible livret.forte livret.nulle impot.faible
0.8571      1.1667      3.3333      12.0000      0.4444      12.0000
  impot.moyen  impot.nul
12.0000      0.1818

```

`acmbank$cw` représentant $\frac{1}{\nu} \mathbf{D}_m$.

```

acmbank$cw
  age.a45      age.a75 livret.faible livret.forte livret.nulle impot.faible
0.17948718 0.15384615 0.07692308 0.02564103 0.23076923 0.02564103
  impot.moyen  impot.nul
0.02564103 0.28205128

frequences/3
      [,1]
age.a45 0.17948718
age.a75 0.15384615
livret.faible 0.07692308
livret.forte 0.02564103
livret.nulle 0.23076923
impot.faible 0.02564103
impot.moyen 0.02564103
impot.nul 0.28205128

```

`acmbank$lw`

```

round(acmbank$lw,4)
[1] 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385
[12] 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385 0.0385
[23] 0.0385 0.0385 0.0385 0.0385

round(1/26,4)
[1] 0.0385

```

`acmbank$eig`, `acmbank$rank`, `acmbank$nf`, `acmbank$c1`, `acmbank$li`, `acmbank$co`, `acmbank$l11` sont des informations déjà rencontrées en analyse en composantes principales et en analyse des correspondances.

`acmbank$cr`

L'objectif de l'ACM est d'obtenir des scores numériques des individus maximisant la somme des rapports de corrélation¹ entre les scores et les variables qualitatives. Le data frame `acmbank$cr` contient les rapports de corrélations entre les variables qualitatives de départ et les axes principaux.

¹variance interclasse sur variance totale, au sens descriptif des variances

```
acmbank$cr
      RS1      RS2
age    0.7557272 0.0005539102
livret 0.3058728 0.6387474774
impot  0.4488981 0.6446937421

#
vartot <- function(x) sum((x-mean(x))^2)
varinter <- function(x,gpe) {
  moyennes <- tapply(x,gpe,mean)
  effectifs <- tapply(x,gpe,length)
  sum(effectifs*(moyennes-mean(x))^2)
}
eta2 <- function(x,gpe) varinter(x,gpe)/vartot(x)
#
rapcor <- function(x,nf) eta2(acmbank$li[1:26,nf],bank26[1:26,x])
cbind(sapply(1:3,rapcor,nf=1),sapply(1:3,rapcor,nf=2))

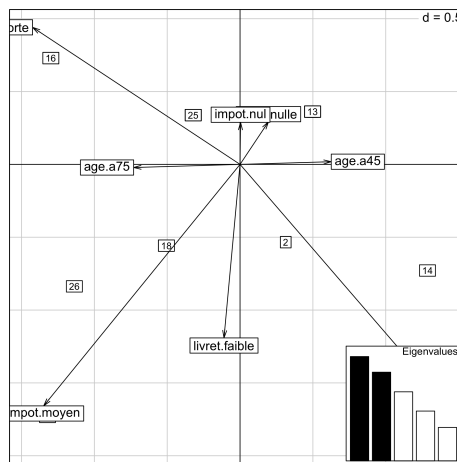
      [,1]      [,2]
[1,] 0.7557272 0.0005539102
[2,] 0.3058728 0.6387474774
[3,] 0.4488981 0.6446937421
```

1.2.3 Les graphiques

a) La représentation simultanée des lignes et des colonnes

Comme pour l'ACP et l'AFC, il existe une représentation simultanée des lignes et des colonnes.

```
scatter.dudi(acmbank, posieig="bottomright")
```



On ne voit que 9 individus car le jeu de données, réduit à 26 individus sur les 810 du fichier initial, ne contient pas l'ensemble des configurations possibles ($2 \times 3 \times 3 = 18$).

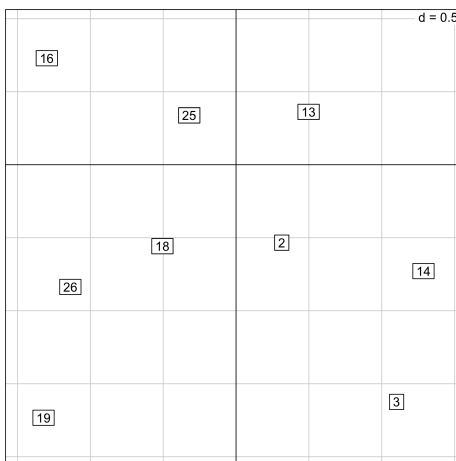
```
unique(bank26, fromLast=TRUE)
  age livret impot
2  a45 faible nul
3  a45 faible faible
13 a45 nulle nul
14 a45 nulle faible
16 a75 forte nul
18 a75 faible nul
19 a75 faible moyen
25 a75 nulle nul
26 a75 nulle moyen
```

Le nombre de répétitions pour chacune de ces configurations est :

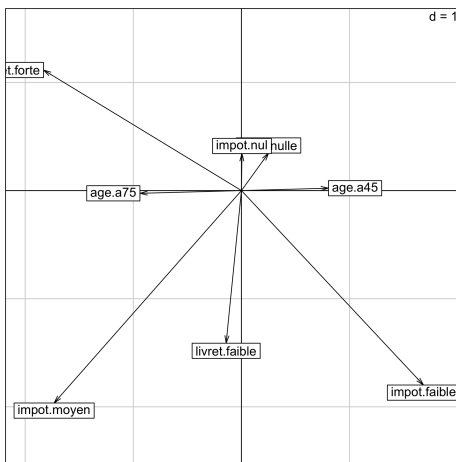
| | | | | | | | | |
|---|---|----|----|----|----|----|----|----|
| 2 | 3 | 13 | 14 | 16 | 18 | 19 | 25 | 26 |
| 2 | 1 | 10 | 1 | 2 | 2 | 1 | 6 | 1 |

b) La représentation séparée des lignes et des colonnes

```
s.label(acmbank$li)
```



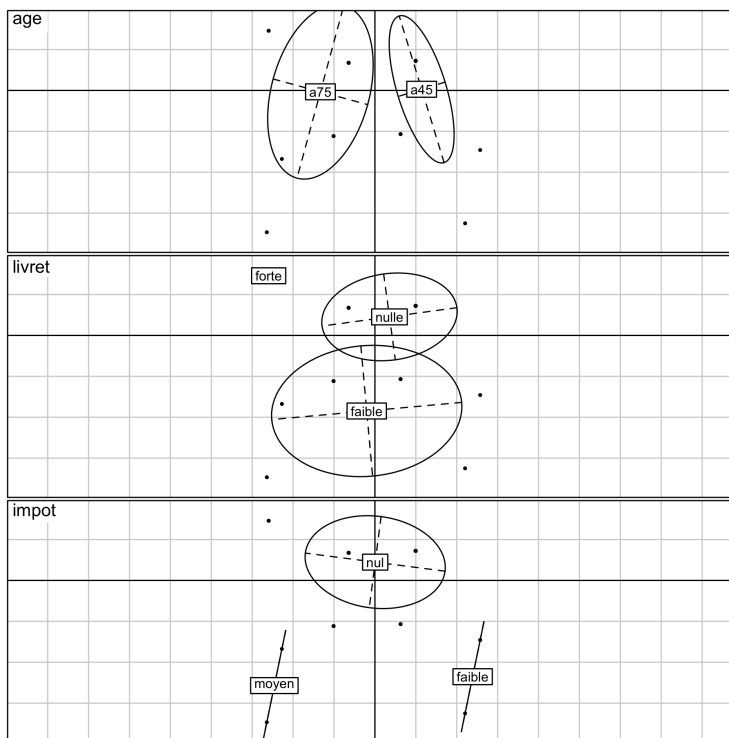
```
s.arrow(acmbank$co)
```



c) La représentation spécifique de l'ACM

La représentation simultanée des lignes et des colonnes peut devenir vite illisible si le nombre total de modalités est élevé. C'est pourquoi une autre démarche a été préférée. Dans la représentation graphique ci-dessous, le même plan factoriel est répété autant de fois qu'il y a de variables qualitatives. Sur chaque plan, il y a 26 points correspondant aux individus enquêtés même si seuls 9 apparaissent (cf explication ci-dessus). Pour faciliter l'interprétation, on représente, variable par variable, la modalité prise par chaque individu et une ellipse résumant la dispersion des points.

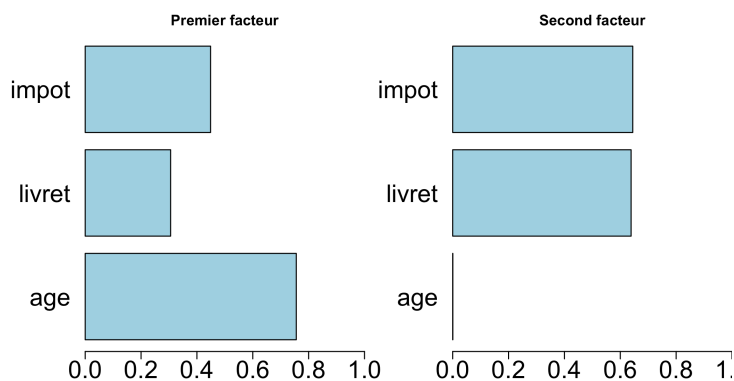
`scatter(acmbank)`



d) Le lien entre les scores et les variables qualitatives

On peut représenter les rapports de corrélations.

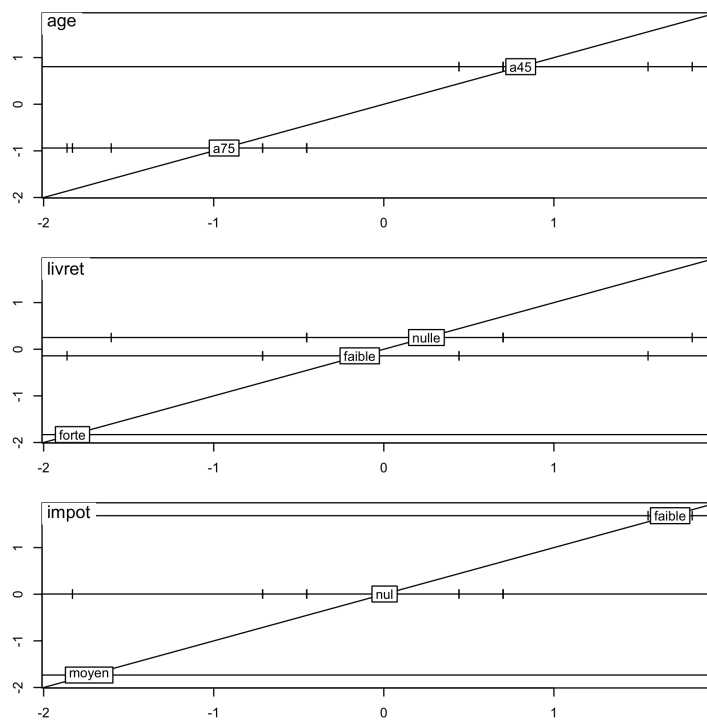
```
par(mfrow=c(1,2), mar=c(5,6,2,0), cex=0.70)
barplot(acmbank$cr[,1], horiz=TRUE, xlim = c(0,1), names.arg = colnames(bank26), las = 1,
main = "Premier facteur", col = "lightblue",
cex.names=2, cex.axis=2)
barplot(acmbank$cr[,2], horiz=TRUE, xlim = c(0,1), names.arg = colnames(bank26), las = 1,
main = "Second facteur", col = "lightblue",
cex.names=2, cex.axis=2)
```



e) La représentation d'une seule dimension

La fonction `score()` permet de visualiser les variables qualitatives avec un facteur. Pour chaque variable, les individus sont positionnés sur l'axe des abscisses par leur score sur l'axe factoriel considéré et, sur l'axe des ordonnées par le score de la modalité qu'ils portent. Le score d'une modalité est la moyenne des scores des individus portant cette modalité, ce qui est mis en évidence par la première bissectrice.

```
score(acmbank, xax=1)
```



2 Un exemple réel pour poursuivre ...

2.1 Les données

Les données proviennent d'une enquête réalisée dans des supermarchés angevins et parisiens entre 1996 et 1998 dans le but de connaître l'avis de consommateurs quant aux produits biologiques et aux produits diététiques. Elles nous sont proposées par Gilles Hunault de l'université d'Angers et se trouvent originalement à l'adresse <http://www.info.univ-angers.fr/~gh/Datasets/pbio.txt> avec une copie sur le site pédagogique <http://pbil.univ-lyon1.fr/R/donnees/pbio.txt>.

419 individus ont répondu aux questions suivantes :

CONNAITRE Connaissez-vous les produits biologiques ?

- 0 non réponse
- 1 oui
- 2 non

DIFF Y a-t-il une différence entre produit biologique et produit diététique ?

- 0 non réponse
- 1 oui
- 2 non

CONSOM Avez-vous déjà consommé des produits biologiques ?

- 1 non jamais
- 2 oui une seule fois
- 3 oui rarement
- 4 oui de temps en temps
- 5 oui plusieurs fois par mois
- 6 oui plusieurs fois par semaine
- 7 ne se prononce pas

MARQUE Parmi les marques suivantes, laquelle connaissez-vous ?

- 0 non réponse
- 1 bio vivre
- 2 bjorg
- 3 carrefour bio
- 4 la vie
- 5 vrai
- 6 prosain
- 7 favrichon

CONSVIE Avez-vous déjà consommé des produits 'la vie' ?

- 0 non réponse
- 1 oui une fois
- 2 oui occasionnellement
- 3 oui régulièrement
- 4 non jamais

SEXE Sexe de la personne

- 1 homme
- 2 femme

AGE Classe d'âge

- 1 moins de 25 ans
- 2 entre 25 et 35 ans
- 3 entre 35 et 45 ans
- 4 entre 45 et 55 ans
- 5 entre 55 et 65 ans
- 6 plus de 65 ans

ETATCIVIL Etat Civil

- 0 autre
- 1 marié
- 2 célibataire
- 3 divorcé
- 4 en concubinage
- 5 veuf

NBENF Nombre d'enfants

- 1 sans enfant
- 2 1 enfant
- 3 2 enfants
- 4 3 enfants
- 5 plus de 3 enfants

SITPROF Situation Professionnelle

- 1 agriculteur
- 2 artisan
- 3 cadre supérieur
- 4 cadre moyen
- 5 employé
- 6 ouvrier
- 7 retraité
- 8 autre
- 9 non réponse

REVENU Classe de revenus mensuels

- 0 non réponse
- 1 moins de 5 kF
- 2 entre 5 et 10 kF
- 3 entre 10 et 15 kF
- 4 entre 15 et 20 kF
- 5 plus de 20 kF

6 ne se prononce pas

La première colonne CODE correspond à l'identifiant associé à la personne interrogée.

```
pbio <- read.table("http://pbil.univ-lyon1.fr/R/donnees/pbio.txt", h=T, row.names=1)
names(pbio)
[1] "CONNAITRE" "DIFF"      "CONSOM"   "MARQUE"   "CONSVIE"  "SEXE"
[7] "AGE"       "ETATCIVIL" "NBENF"    "SITPROF"  "REVENU"
don <- pbio
```

2.2 Quelques questions autour de ces données

1. Quelle est la dimension de ce data frame ?
2. Ecrire le résumé statistique du data frame `pbio`. Que constate-t-on ? Modifier-le pour le rendre conforme à la réalité des données.
3. Ecrire le nouveau résumé statistique. Donner le nombre d'enquêtés connaissant la marque `carrefour bio`.
4. On note que certains enquêtés n'ont pas répondu aux questions posées mais que la non réponse n'obéit pas toujours au même codage. On modifie le data frame (1) en remplaçant les modalités 'non réponse' codées par 0 (sauf dans un cas par 7) par des 'NA' et (2) en ne conservant qu'un data frame des données complètes.

```
int <- don
temp <- which(int==0, arr.ind = TRUE)
for (i in 1:100) int[temp[i,1],temp[i,2]] <- NA
for (i in 1:419) if(int[i,3]==7) int[i,3] <- NA
for (j in 1:11) int[,j] <- factor(int[,j])
pbio.cc <- int[complete.cases(int),]
summary(pbio.cc)
```

| CONNAITRE | DIFF | CONSOM | MARQUE | CONSVIE | SEXE | AGE | ETATCIVIL | NBENF |
|-----------|---------|--------|--------|---------|-------|------|-----------|-------|
| 1:305 | 1:251 | 1:76 | 1: 1 | 1: 9 | 1: 96 | 1:46 | 1:168 | 1:176 |
| 2: 9 | 2: 63 | 2:12 | 2:135 | 2: 47 | 2:218 | 2:93 | 2: 89 | 2: 59 |
| | | 3:70 | 3: 23 | 3: 16 | | 3:51 | 3: 16 | 3: 53 |
| | | 4:94 | 4: 91 | 4:242 | | 4:77 | 4: 33 | 4: 16 |
| | | 5:20 | 5: 46 | | | 5:24 | 5: 8 | 5: 10 |
| | | 6:42 | 6: 5 | | | 6:23 | | |
| | | | 7: 13 | | | | | |
| | SITPROF | REVENU | | | | | | |
| 8 | :94 | 1:18 | | | | | | |
| 5 | :87 | 2:79 | | | | | | |
| 4 | :64 | 3:64 | | | | | | |
| 7 | :31 | 4:49 | | | | | | |
| 3 | :25 | 5:83 | | | | | | |
| 2 | : 9 | 6:21 | | | | | | |
| (Other): | 4 | | | | | | | |

On constate que, après avoir enlevé les données manquantes, la modalité `agriculteur` de la variable `SITPROF` vaut 0.

Et oui! ce sont des données réelles ...

```
summary(pbio.cc$SITPROF)
```

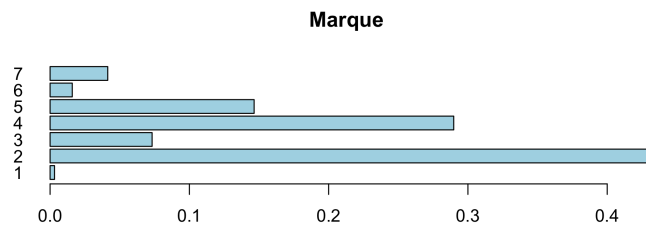
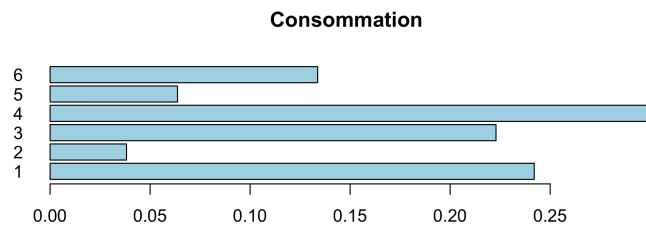
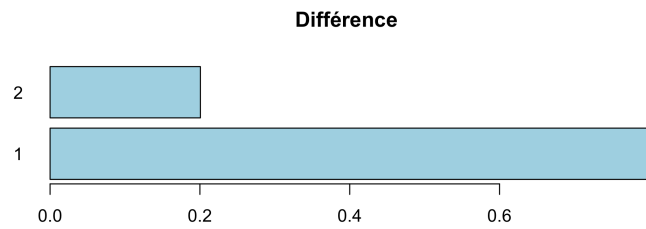
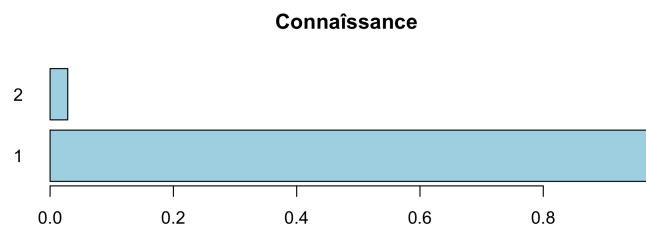
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|----|----|----|---|----|----|
| 0 | 9 | 25 | 64 | 87 | 4 | 31 | 94 |

```
levels(pbio.cc$SITPROF)
[1] "1" "2" "3" "4" "5" "6" "7" "8"
```

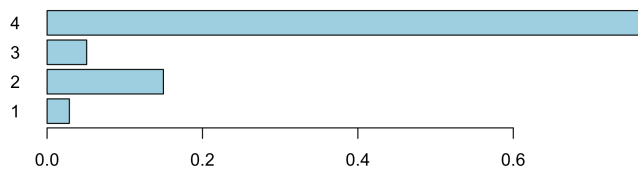
Il faut donc redéfinir les modalités de cette variable.

```
pbio.cc$SITPROF <- factor(pbio.cc$SITPROF)
levels(pbio.cc$SITPROF)
[1] "2" "3" "4" "5" "6" "7" "8"
efftot <- dim(pbio.cc)[1]
```

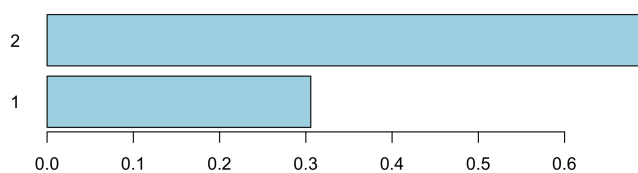
On peut visualiser chaque variable à l'aide d'une représentation en bâtons :



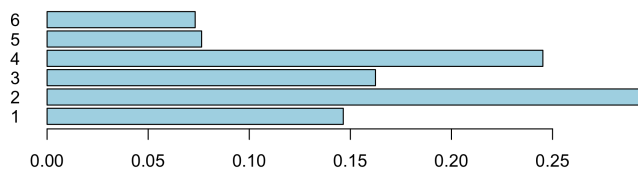
Marque La Vie



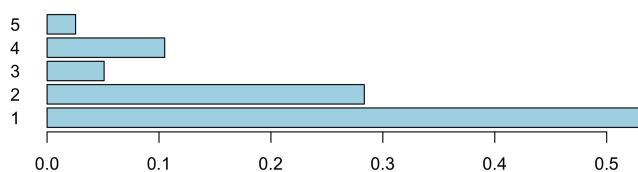
Sexe



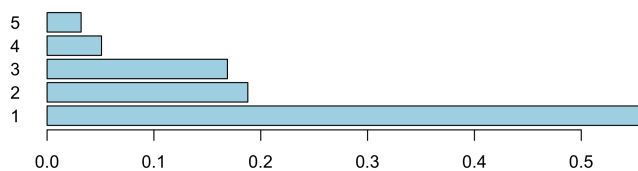
Age



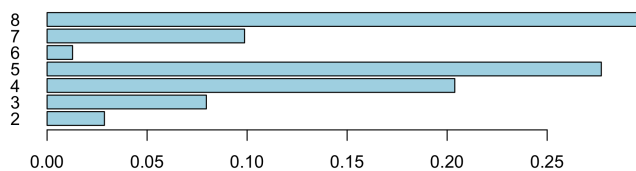
Etat Civil



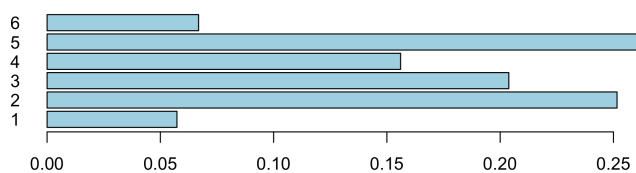
Nombre d'enfants



Situation Professionnelle

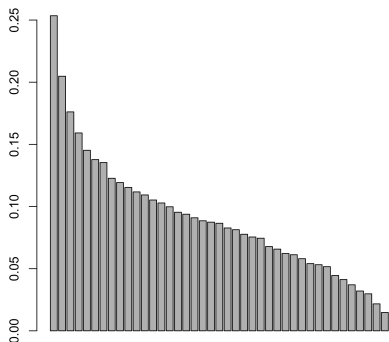


Revenu



2.3 L’analyse des correspondances multiples

```
library(ade4)
acmtot <- dudi.acm(pbio.cc,scannf=FALSE)
barplot(acmtot$eig)
```



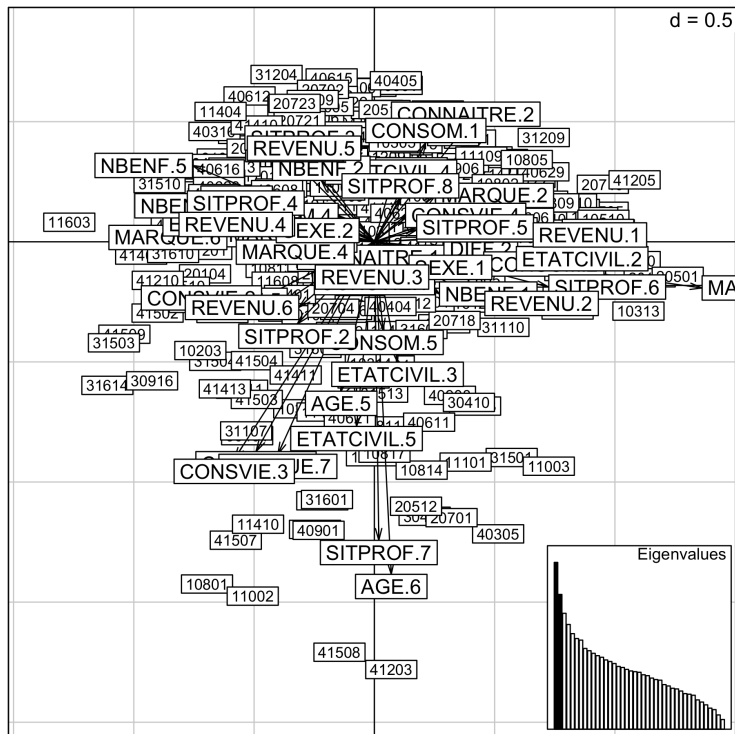
On note que le nombre important des valeurs propres (liées on le rappelle non aux variables mais aux modalités de ces variables) ne permet pas d’énoncer un critère de sélection du nombre de facteurs à conserver. On conserve 4 valeurs propres mais on ne détaillera dans la présentation que les deux premiers. A charge au lecteur de regarder les facteurs 3 et 4.

```
head(inertia.dudi(acmtot)$TOT)
  inertia      cum      ratio
1 0.2534726 0.2534726 0.06800484
2 0.2047920 0.4582646 0.12294905
```

```
3 0.1761199 0.6343845 0.17020072
4 0.1591724 0.7935569 0.21290550
5 0.1452219 0.9387788 0.25186747
6 0.1377689 1.0765476 0.28882985
```

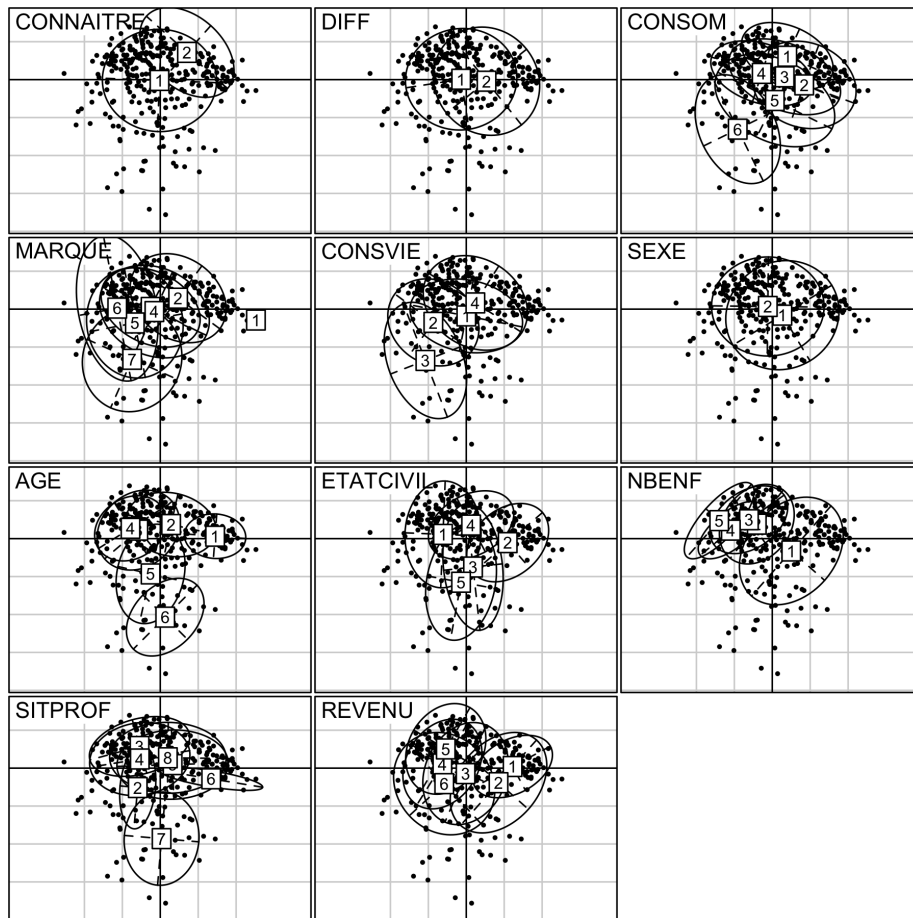
En gardant les quatre premiers facteurs, on ne conserve que 21.29% de l'inertie totale. Mais ce pourcentage est relativement courant dans ce genre d'analyse. On pourrait représenter simultanément les individus et les modalités des variables sur un même graphique, démarche classique en analyse des données.

```
scatter.dudi(acmtot, posieig="bottomright")
```



La représentation est définitivement illisible. C'est pourquoi on préfère la représentation spécifique de l'ACM.

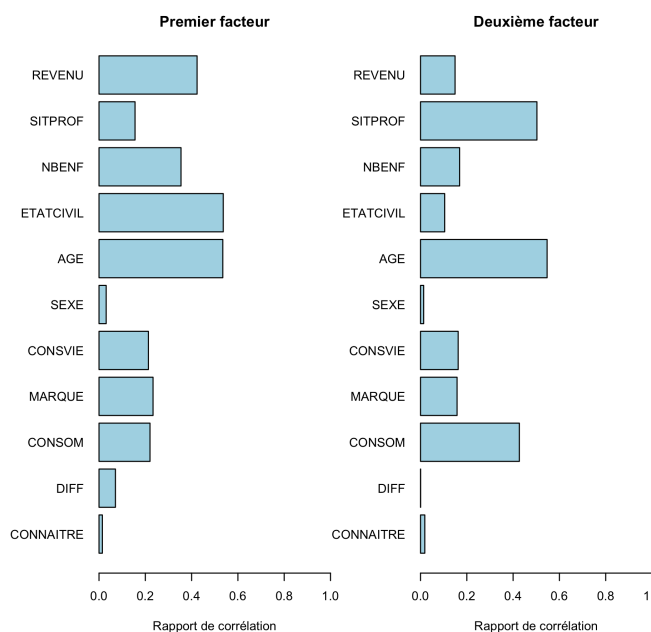
```
scatter(acmtot)
```

On voit par exemple que pour la variable **CONSO**, il y a opposition entre ceux qui consomment des produits biologiques plusieurs fois par semaine [6] et tous les autres, de ceux qui ne consomment jamais [1] à ceux qui consomment plusieurs fois par mois [5].

On peut représenter les rapports de corrélation.

```
par(mfrow = c(1,2), mar=c(5,6,2,0), cex=0.70)
barplot(acmtot$scr[,1], horiz=TRUE, xlim = c(0,1), names.arg = colnames(pbio), las = 1,
main = "Premier facteur", col = "lightblue",
xlab = "Rapport de corrélation")
barplot(acmtot$scr[,2], horiz=TRUE, xlim = c(0,1), names.arg = colnames(pbio), las = 1,
main = "Deuxième facteur", col = "lightblue",
xlab = "Rapport de corrélation")
```



3 Une autre étude pour finir ...

On considère les cinq premières variables comme des variables actives et les suivantes comme des variables illustratives.

- Réaliser une analyse des correspondances multiples sur les individus conservés dans le tableau et les 5 variables actives.
- Réaliser une analyse des correspondances multiples sur les individus conservés dans le tableau et les 6 variables illustratives.
- Calculer le coefficient de corrélation entre le premier facteur de l'analyse sur les variables actives et le premier facteur de l'analyse sur les variables illustratives. Commenter.