

RD6 : Analyse des correspondances multiples

Tableaux étudiés : L'AFCM est utilisée pour l'étude de volumineux tableaux constitués de plusieurs variables qualitatives, principalement sous forme de questionnaires. Le tableau se présente sous forme codé, disjonctif codé, ou de tableau de Burt. Il est mathématiquement traité sous forme d'un tableau disjonctif complet analysé comme un unique tableau de contingence.

Les références suivantes peuvent compléter ce chapitre :

- <http://www.sthda.com/french/articles/38-methodes-des-composantes-principales-dans-r-guide-pratique/75-acm-analyse-des-correspondances-multiples-avec-r-l-essentiel/>
- Principal component analysis (article) (Abdi and Williams 2010). <https://goo.gl/1Vtwq1>.
- <https://eric.univ-lyon2.fr/~ricco/cours/slides/ACM.pdf>

Notations :

- n individus notés i , q variables X^k , ayant chacune p_k modalités avec $p = p_1 + \dots + p_q$ modalités au total.
- $Z = (Z_1 | \dots | Z_q)$ représente le tableau disjonctif complet. $z_{ij} = 1$ si l'individu i a la modalité j , 0 sinon. On a $\sum_i z_{ij} = z_{.j}$, $\sum_j z_{ij} = q$, $\sum_{i,j} z_{ij} = nq$.
- $z_{.j}$ définit l'effectif marginal de la modalité j variant de 1 à p .
- On note $D = \text{diag}(z_{.j})$.
- On appelle tableau de Burt le tableau $Z^T Z$. Il correspond aux tableaux de contingence de tous les couples de variables.

Exemple : Prenons le tableau suivant comprenant $n = 5$ individus notés i , i de 1 à n , $q = 3$ variables X^k , k de 1 à q , ayant $p_1 = 3$, $p_2 = 2$ et $p_3 = 2$ modalités m_l^k soit $p = p_1 + p_2 + p_3$ modalités au total.

A partir du tableau codé, construire le tableau disjonctif noté Z et de Burt $Z^T Z$.

$$\begin{array}{c}
 X^1 \quad X^2 \quad X^3 \\
 \left(\begin{array}{ccc} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 2 & 2 \\ 3 & 2 & 2 \\ 2 & 1 & 1 \end{array} \right)
 \end{array}
 \quad
 \begin{array}{c}
 \left(\begin{array}{ccc|cc|cc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \end{array} \right)
 \end{array}
 \quad
 \begin{array}{c}
 \left(\begin{array}{ccc|cc|cc} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 1 & 0 & 2 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 3 & 1 & 2 \\ 1 & 1 & 0 & 1 & 1 & 2 & 0 \\ 1 & 1 & 1 & 1 & 2 & 0 & 3 \end{array} \right)
 \end{array}$$

Objectifs :

- Comme l'ACP, on peut envisager d'étudier la typologie des individus. Les individus sont considérés comme proches si ils ont de nombreuses modalités communes. Le plus souvent le nombre d'individus est très grand et anonyme. On se limite alors à l'étude de certaines classes (la ménagère de moins de 50 ans, les teenagers...)
- Comme l'AFC, on peut étudier la liaison entre les modalités :
 - Pour deux modalités d'une même variable (exclusives), elles se ressemblent si les modalités pour les autres variables sont similaires.
 - Pour des modalités de deux variables, leur proximité indique une plus forte association que sous l'hypothèse d'indépendance et des modalités communes.
- L'ACM permet aussi de construire des variables quantitatives (composantes principales F_L à partir de variables qualitatives exploitables dans d'autres méthodes comme la méthode DISQUAL en analyse discriminante.

1 Nuage des profils et ajustement

L'idée de départ est d'utiliser une approche analogue à l'AFC pour l'étude des liaisons entre modalités. L'utilisation des programmes de l'AFC sur les tableaux disjonctifs complets (ou les tableaux de Burt) permet une étude de ces tableaux avec des caractéristiques spécifiques ici :

- le tableau de contingence N de l'AFC est remplacé ici par le tableau disjonctif Z , le tableau ne contient donc que des 0 et 1,
- l'effectif total est nq , la somme d'une ligne est toujours q ,
- la matrice des fréquence relative F de l'AFC devient $F = \frac{1}{nq}Z$ en ACM,
- les colonnes peuvent être regroupées en variables,
- le profil marginal ligne est $\frac{q}{nq} = \mathbf{1}_n$, le profil marginal colonne est $(\frac{z_{.j}}{nq})$,
- par analogie avec l'AFC, on en déduit :

$$D = (z_{.j}) \quad D_I = \frac{1}{n}I_n \quad D_J = \left(\frac{z_{.j}}{nq}\right) = \frac{1}{nq}D \quad F = \frac{1}{nq}Z.$$

Définition 1 On appelle ACM d'un tableau l'AFC du tableau disjonctif complet.

Remarque : On obtient une représentation des modalités identiques en réalisant l'AFC du tableau de Burt.

Proposition 1 On en déduit par analogie :

- $L = D_I^{-1}F = \frac{1}{q}Z$ de poids $D_I = \frac{1}{n}I_n$ et de métrique :

$$d_{\chi^2}^2(i, i') = \sum_{j=1}^p \frac{nq}{z_{.j}} \left(\frac{z_{ij}}{q} - \frac{z_{i'j}}{q} \right)^2$$

Deux individus sont donc proches si ils ont de nombreuses modalités communes : $\left(\frac{z_{ij}}{q} - \frac{z_{i'j}}{q} \right) = 0$.

Ils seront d'autant plus éloignés qu'ils présentent des modalités rares différentes : $\frac{nq}{z_{.j}}$ élevé.

- $C = D^{-1}Z^t$ de poids $D_J = \frac{1}{nq}D$ et de métrique

$$d_{\chi^2}^2(j, j') = \sum_{i=1}^n n \left(\frac{z_{ij}}{z_{.j}} - \frac{z_{ij'}}{z_{.j'}} \right)^2$$

Deux modalités sont proches si les individus les possédant ont des modalités communes.

- On réalise ainsi les dvs des triplets $(L = \frac{1}{q}Z, nqD^{-1}, \frac{1}{n}I_n)$ et $(C = D^{-1}Z^t, nI_n, \frac{1}{nq}D)$
- Les matrices d'inertie sont :

$$Z^t Z D^{-1} \text{ pour } L \quad Z D^{-1} Z^t \text{ pour } C$$

- Les formules de transition sont :

$$F_L = \frac{1}{q} Z F^C \Lambda^{-1} \quad F^C = D^{-1} Z^t F_L \Lambda^{-1}$$

$$F_L^s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{j=1}^q \frac{z_{ij}}{s} F_s^C(j) = \frac{1}{s\sqrt{\lambda_s}} \sum_{j \in I(i)} F_s^C(j)$$

avec $I(i)$ l'ensemble des modalités j de l'individu e_i

$$F_s^C(j) = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^n \frac{z_{ij}}{z_{.j}} F_s^C(j) = \frac{1}{z_{.j}\sqrt{\lambda_s}} \sum_{i \in I(j)} F_L^s(i)$$

avec $I(j)$ l'ensemble des individus i qui possèdent la modalité j .

Preuve : Bon exercice en reprenant les résultats de RD5 sur l'AFC avec $F = \frac{1}{nq}Z$.

2 Paramètres d'interprétation de l'ACM

2.1 Inertie

Proposition 2 *L'inertie totale se décompose par modalité et par variable :*

- *L'inertie d'une modalité est : $I(j) = \frac{1}{q}(1 - \frac{z_{.j}}{n})$*
- *L'inertie d'une variable est : $I(X^k) = \frac{p_k}{q} - \frac{1}{q}$*
- *L'inertie totale est : $I_T = \frac{p}{q} - 1$*

avec q le nombre de variables, p_k le nombre de modalités de la variable k , p le nombre total de modalités, $z_{.j}$ l'effectif des individus possédant la modalité j et n le nombre d'individus.

Preuve : En remarquant que $F_J = (\frac{1}{n})$, le poids d'un individu $\frac{q}{nq}$ et d'une modalité $\frac{z_{.j}}{nq}$ de poser le calcul, la distance entre j et j' est :

$$d_{\chi^2}^2(j, j') = \sum_i \frac{nq}{q} (\frac{z_{ij'}}{z_{.j'}} - \frac{z_{ij}}{z_{.j}})^2 \text{ donc } d_{\chi^2}^2(j, g) = \sum_i n (\frac{z_{ij}}{z_{.j'}} - \frac{1}{n})^2$$

Soit en remarquant que z_{ij} ne prend la valeur 1 que pour $z_{.j}$ individus, l'inertie de j :

$$I(j) = \frac{z_{.j}}{nq} \sum_i n (\frac{z_{ij}}{z_{.j}} - \frac{1}{n})^2 = \frac{z_{.j}}{q} \sum_i (\frac{z_{ij}^2}{z_{.j}^2} - 2\frac{z_{ij}}{nz_{.j}} + \frac{1}{n^2}) = \frac{z_{.j}}{q} - \frac{1}{q}.$$

De même en sommant sur les modalités de la variable k et en sommant sur les variables pour l'inertie totale.

Remarque :

- Le poids d'une modalité est d'autant plus grand qu'elle est peu représentée ($z_{.j}$ faible). On regroupera éventuellement certaines modalités peu représentées.
- Le poids d'une variable est d'autant plus grand qu'elle a beaucoup de modalité (p_k grand). On limitera éventuellement le nombre de modalités.
- L'inertie totale est une grandeur abstraite, sans interprétation statistique. Elle est liée à la structure du tableau et une partie de cette inertie ne représente pas une information statistique. L'interprétation de cette grandeur et de celles en découlant (λ_s , \cos^2 ...) est donc délicate et négative.

Avant toute ACM, il est donc indispensable de réaliser une analyse préliminaire de chaque variable, afin de voir si toutes les classes sont aussi bien représentées ou s'il existe un déséquilibre. L'ACM est sensible aux effectifs faibles, aussi il est préférable de regrouper les classes peu représentées le cas échéant.

2.2 Représentations graphiques

Les relations quasi-barycentriques justifient la représentation simultanée mais le nombre d'individus est souvent très important.

- La proximité de deux individus indique qu'ils ont de nombreuses modalités en commun.
- La proximité entre deux modalités de deux variables différentes indique que ces modalités apparaissent souvent ensemble (association)
- La proximité entre deux modalités d'une même variable indique que les autres variables présentent des modalités similaires pour les individus concernés.

Proposition 3 *Pour les profils lignes, la somme pour chaque variable vaut $1/q$ (q équations), soit q équations. Le nuage est donc représenté dans un sous espace de dimension $p - q$. En conséquence, l'ACM ne possède au plus que $p - q$ valeurs propres non nulles, soit $p - q$ axes pour l'interprétation.*

En particulier, les modalités d'une même variables ont l'origine comme centre de gravité dans les cartes factorielles

Preuve : Bon exercice en calculant le centre de gravité des modalités d'un même variable. On en déduit une conséquence sur la dimension de l'image.

2.3 Sélection des axes

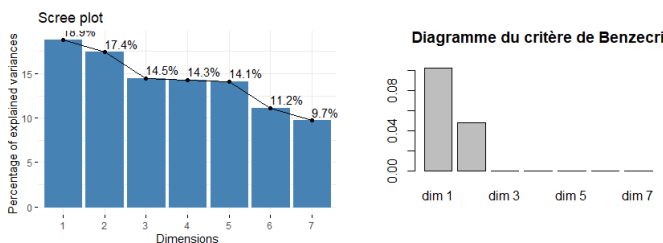
L'inertie totale comme les cos2 et contribution sont en ACM des indicateurs difficiles à interpréter car ce sont des indicateurs en général pessimistes et qui ne représentent pas une information statistique exclusivement. L'interprétation demande une bonne expérience.

Pour la sélection des axes, Benzecri a proposé de recalculer les valeurs propres pour $\lambda > \frac{1}{q}$, q le nombre de variables, par :

$$\left(\frac{q}{q-1}\right)^2\left(\lambda - \frac{1}{q}\right)^2$$

se révélant un critère de sélection des axes plus adapté. On réalise alors le diagramme de ce critère.

Valeurs propres 0.660 0.610 0.507 0.500 0.493 0.390 0.340



2.4 Interprétation des paramètres cos2 et contrib

Du fait de la remarque précédente, l'interprétation des \cos^2 est délicate. On utilise plutôt les contributions pour déterminer l'importance des modalités ainsi que de nouveaux paramètres comme la corrélation pour une variable avec l'axe ou le v-test pour les modalités.

2.5 Corrélation η^2 d'une variable avec l'axe factoriel

L'interprétation porte sur les modalités mais aussi sur les variables en définissant pour chacune d'elle un coefficient de corrélation. On doit ainsi définir une corrélation entre une variable qualitative et une variable quantitative.

Définition 2 Pour chaque variable X^k , on définit le carré d'un coefficient de corrélation avec l'axe s comme le rapport de la variance intermodalité sur la variance totale de F_L^s :

$$\eta^2(s, k) = \frac{\text{Var}(E(F_L^s | j))}{\text{Var}(F_L^s)} = \sum_{j \in I(k)} \frac{1}{\lambda_s} \frac{z_{.j}}{n} (F_L^s(B_j))^2 = \sum_{j \in I(k)} \frac{z_{.j}}{n} (F_s^C(j))^2$$

avec B_j le barycentre des individus $I(j)$ possédant la modalité j et $F_L^s(B_j) = \sqrt{\lambda_s} F_s^C(j)$ par les relations de transition.

On en déduit que :

$$\eta^2(s, k) = q \sum_{j \in I(k)} \text{inertie de } j \text{ sur } s$$

soit la contribution relative de la variable $(\sum_{j \in I(k)} \text{contrib}_s(j))$ à l'axe multipliée par l'inertie de l'axe (λ_s) et par le nombre de variables q .

Proposition 4 Sous H_0 "absence de lien entre l'axe s et la variable k , $(n-2) \frac{\eta^2}{1-\eta^2}$ suit la loi F à 1 et $n-2$ ddl. Il est ainsi possible de tester "

Proposition 5 On a ainsi la relation :

$$\lambda_s = \frac{1}{q} \sum_{k=1}^q \eta^2(s, k).$$

La valeur propre λ_s (l'inertie de l'axe s) est ainsi la moyenne des rapports de corrélation entre la variable synthétique F_s et les variables qualitatives initiales.

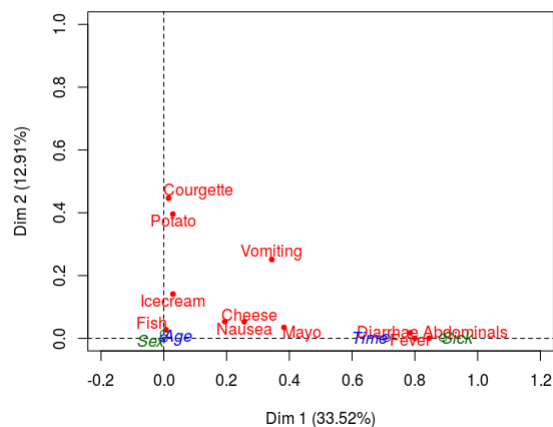
acm\$quali.sup\$eta2

Dim 1

Dim 2

Sick 0.8892035226 0.001495045

Sex 0.0007764989 0.014537117



Preuve et interprétation :

L'inertie totale de l'axe est λ_s , l'inertie inter s'obtient en remplaçant chaque individu par la moyenne des affixes des individus de sa modalité.

On remarque que :

$$F_L^s(B_j) = \frac{1}{z_{.j}} \sum_{i \in I(j)} F_L^s(i) = \sqrt{\lambda_s} \frac{1}{\sqrt{\lambda_s} \frac{1}{z_{.j}}} \sum_{i \in I(j)} \sum_{i \in I(j)} F_L^s(i) = \sqrt{\lambda_s} F_s^C(j)$$

en utilisant l'expression quasi-barycentrique de $F_s^C(j)$

De plus, l'inertie de j suivant s est $\frac{z_{.j}}{nq} F_s^C(j)^2$.

2.6 v-test : test de signification d'une modalité

Il est possible de construire un test sur la signification d'une modalité pour un axe donné.

On a défini $F_s^C(j) = \frac{1}{z_{.j} \sqrt{\lambda_s}} \sum_{i \in I(j)} F_L^s(i)$ égal au facteur près $\frac{1}{\sqrt{\lambda_s}}$ à la moyenne arithmétique des individus possédant j , $\frac{1}{z_{.j}} \sum_{i \in I(j)} PC_L^s(i)$.

On note $X_s(j)$ la variable $\frac{1}{z_{.j}} \sum_{i \in I(j)} F_L^s(i)$.

Si la variable est indépendante de l'axe s le choix des individu $I(j)$ n'a pas d'influence et

$$E(X_s(j)) = 0 \text{ et } \text{Var}(X_s(j)) = \frac{n - z_{.j}}{n - 1} \frac{\lambda_s}{z_{.j}}$$

du fait du tirage sans remise.

Proposition 6 La statistique $t_s(j) = \sqrt{z_{.j} \frac{n - 1}{n - z_{.j}}} F_s^C(j)$ tend vers une loi normale centrée.

Remarque :

- le test est approximatif et purement indicatif. En particulier les comparaisons multiples ne permettent pas d'estimer ici le risque de 1ère espèce.
- le test n'a pas de sens pour les variables actives car elles ont contribué aux calculs. Elles sont néanmoins présentes à titre indicatif.

```
acm$quali.sup$v.test
      Dim 1      Dim 2
Sickn  6.9294293 -0.2841346
Sicky -6.9294293  0.2841346
F       0.2047705  0.8860047
M      -0.2047705 -0.8860047
```

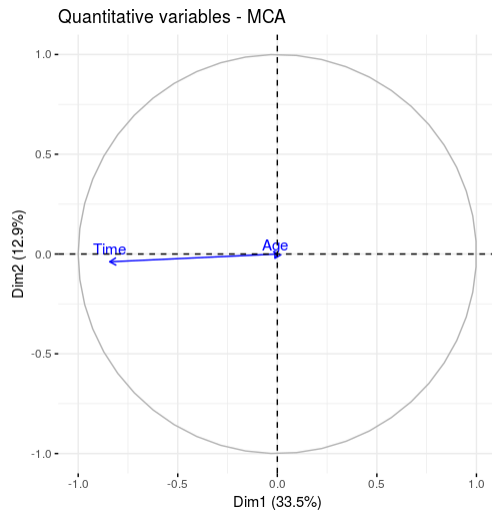
2.7 Individus et Variables supplémentaires

2.7.1 Variable quantitative

Il est possible d'utiliser une variable quantitative comme variable active en la transformant en une variable qualitative avec des classes.

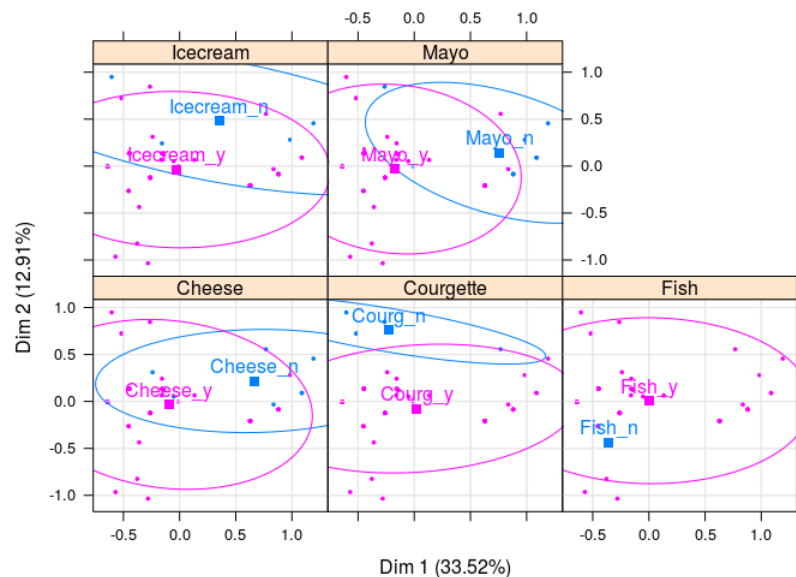
On peut aussi l'introduire comme supplémentaire en calculant ses corrélations avec les composantes principales F_L .

MCA (X, ind = NULL, quanti = NULL, quali.sup = NULL)



2.7.2 Variable qualitative

Par projection après ajustement comme en AFC. On dispose aussi du v-test pour l'interprétation.



2.7.3 Individus

Les individus supplémentaires sont projetés a posteriori.

3 Exemple d'interprétation

L'étude porte sur le tableau poison de FactoMineR : The data used here refer to a survey carried out on a sample of children of primary school who suffered from food poisoning. They were asked about their symptoms and about what they ate. A data frame with 55 rows and 15 columns.

Les variables sont : "Age" "Time" "Sick" "Sex" "Nausea" "Vomiting" "Abdominals" "Fever" "Diarrhae" "Potato" "Fish" "Mayo" "Courgette" "Cheese" "Icecream" .

Les deux premières sont quantitatives, les deux suivantes seront utilisées en supplémentaires, la première sexe étant a priori sans rôle et sick correspondant ici à une variable réponse.

```
library(FactoMineR)
data(poison)
?poison
summary(poison)
for (i in 5:15) {
  plot(poison[,i], main = colnames(poison)[i],
       ylab = "Count", col="steelblue", las = 2)}
```

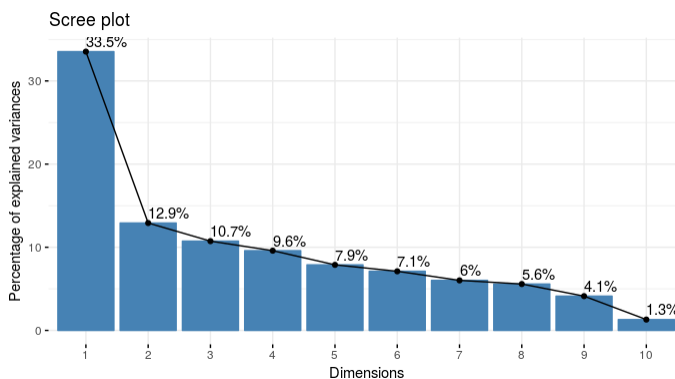
On peut construire le tableau disjonctif complet Z et le tableau de Burt :

```
Z=tab.disjonctif(poison[,c(5:15)])  $\\#$tableau disjonctif complet
t(Z)$\\%*\\%$Z  $\\#$ tableau de Burt\\
```

	Nausea_n	Nausea_y	Vomit_n	Vomit_y
Nausea_n	43	0	28	15
Nausea_y	0	12	5	7
Vomit_n	28	5	33	0
Vomit_y	15	7	0	22

3.1 sélection des axes

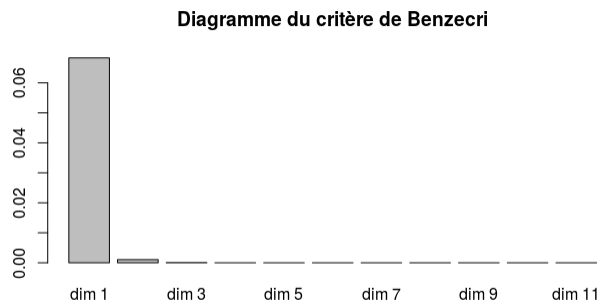
```
library(factoextra)
fviz_eig(acm, addlabels = TRUE)
```



Les 2 premiers axes de l'ACM expriment 46.4 % de l'inertie totale du jeu de données. L'interprétation de ce % est à prendre avec précaution, une partie de l'inertie est artificielle liée à la nature du tableau codé. La règle du coude indique cependant que le premier axe semble pertinent. Pour pallier les difficultés d'interprétation, on peut utiliser le critère

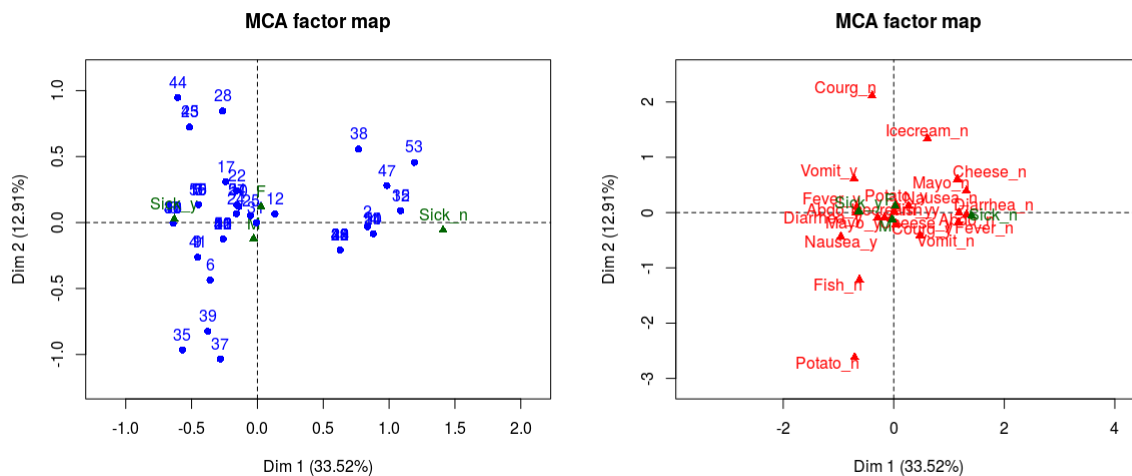
de Benzécri qui confirme clairement que seul le premier axe est pertinent. En fait, cet axe explique la plus grande partie de l'"inertie" intéressante alors que sa valeur n'est que de 33%.

```
#critère de Benzecri de sélection des axes
E=acm$eig[,1]
E[acm$eig[,1]<1/10]=1/10
E=(10/9)^2*(E-1/10)^2
barplot(E)
title("Diagramme du critère de Benzecri")
```

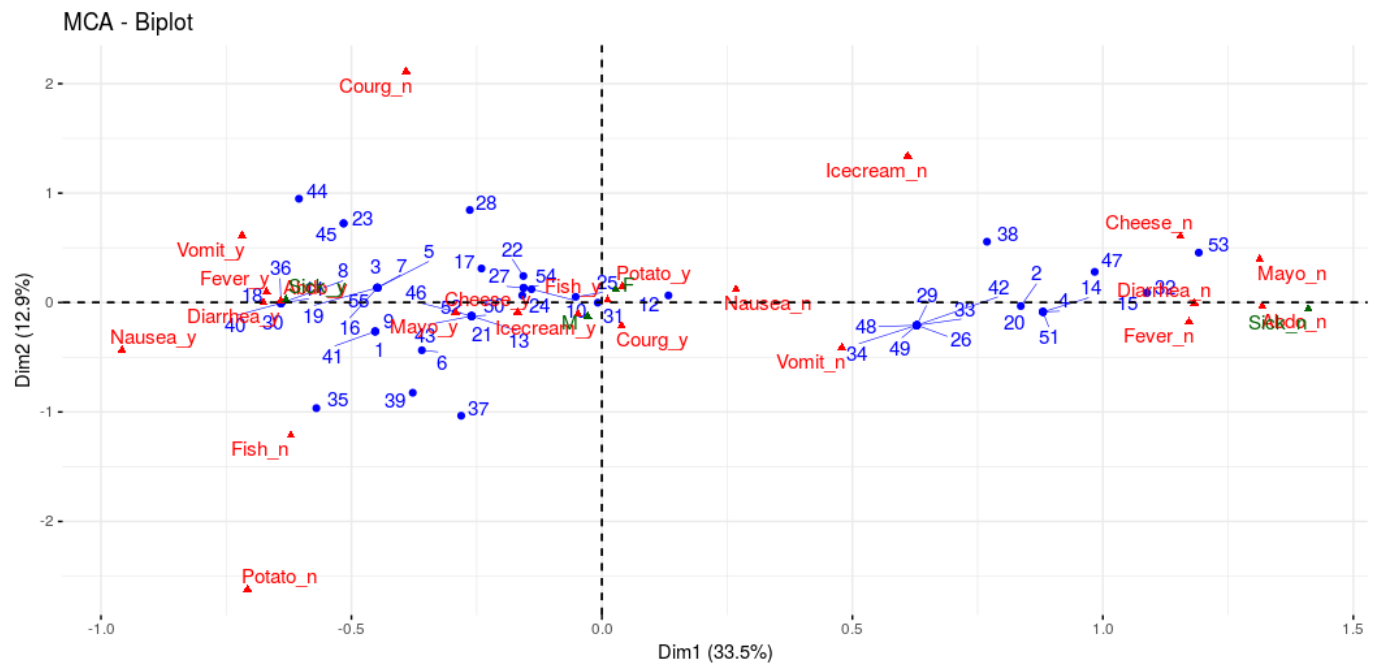


3.2 projection des individus, modalités et variables

```
plot(acm);plot(acm, invisible= "var");plot(acm, invisible = "ind")
```

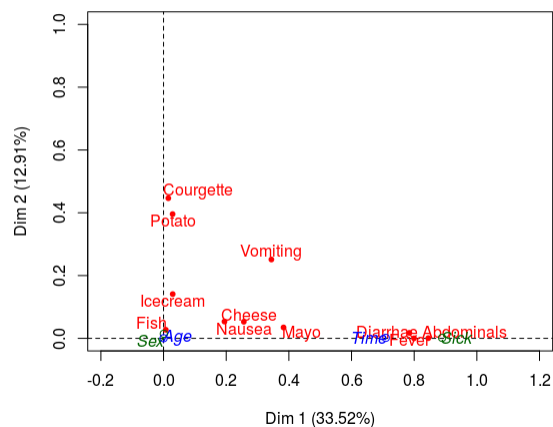


```
fviz$\\_\\$mca$\\_\\$biplot (acm, repel = TRUE,ggtheme = theme$\\_\\$minimal())
```



On peut aussi représenter le carré de la corrélation des variables avec les axes :

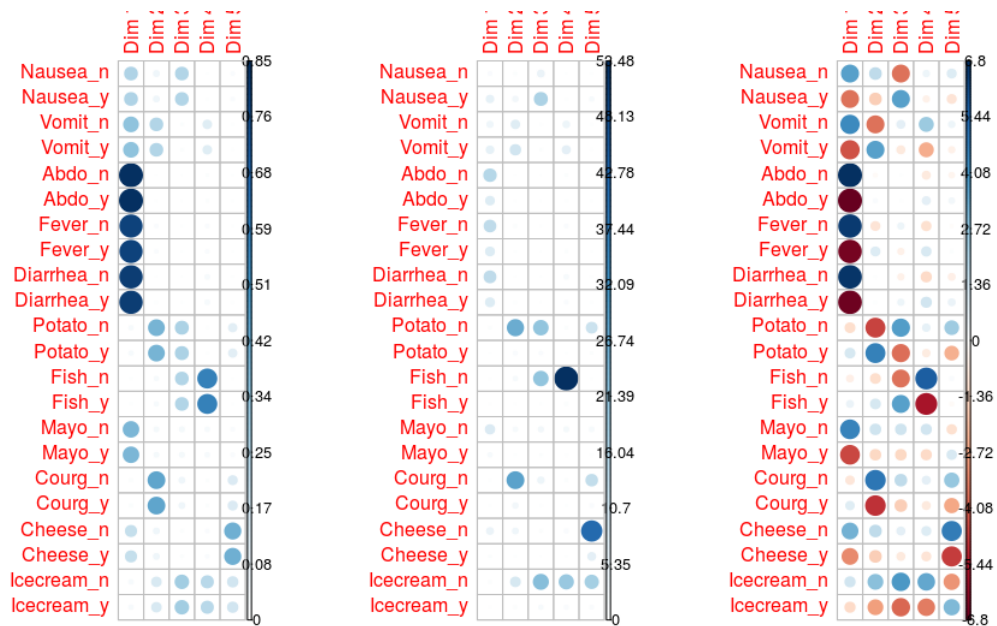
```
plot(acm, choix = "var")
```



3.3 cos2 et contrib

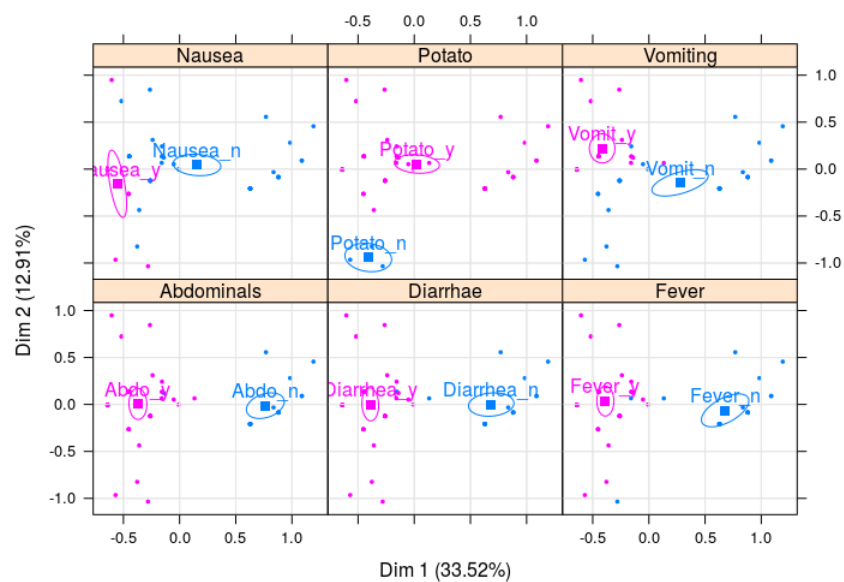
Souvent les individus sont très nombreux et anonymes. L'étude porte principalement sur les modalités et les variables. On utilise les composantes principales des individus dans d'autres méthodes (classification, disqual).

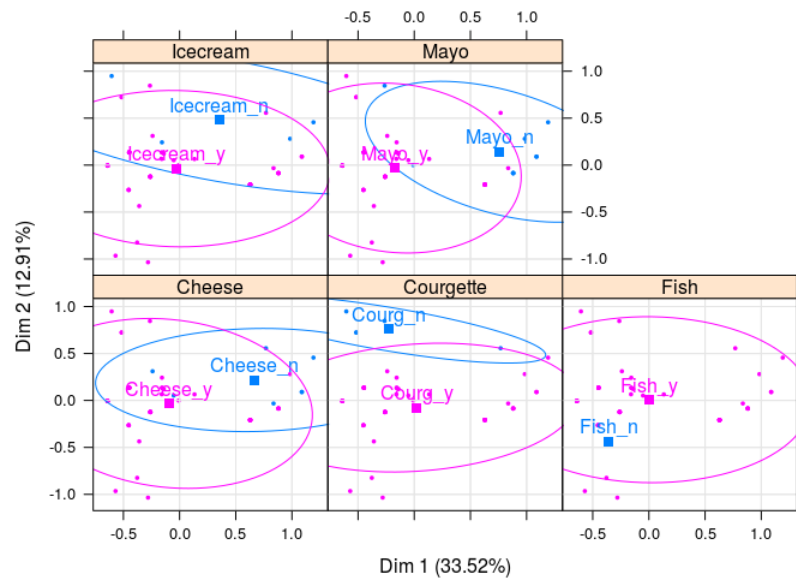
```
library(corrplot)
corrplot(acm$var$cos2, is.corr = FALSE)
corrplot(acm$var$contrib, is.corr = FALSE)
round(acm$var$v.test, 1)
corrplot(round(acm$var$v.test, 1), is.corr = FALSE)
```



On peut aussi représenter la répartition des individus par modalité d'une variable. L'ellipse de confiance du centre de gravité de chaque modalité est également représentée ou celle des observations (means=FALSE).

```
plotellipses(acm, keepvar=c(5:10), level = 0.95)
plotellipses(acm, keepvar=c(11:15), means=FALSE))
```





3.4 Interprétation du plan F1 F2

Il ressort des résultats précédents que les variables les plus pertinentes pour F1 sont Fever, Diarrhea, Abdominal principalement à travers leurs modalités _n. Les individus sans ces symptômes d'intoxication alimentaire se retrouve dans la partie positive de F1, les autres dans la partie négative. Cet axe semble indiquer la gravité de l'intoxication, du plus grave au moins grave. Ce résultat est confirmé par les tests réalisés sur les variables et modalités pour lesquels les valeurs obtenus sont fortes. On retrouve pour la variable silk la même répartition selon l'axe F1 confirmant son interprétation en terme d'intensité de l'intoxication.

L'axe F2 est conditionné principalement par les variables courgette et potato, et dans une moindre mesure, vomit, par leur modalités négatives, _n. La signification de cet axe est plus incertaine, les 2 modalités s'opposant selon ces deux axes sans relation de causalité évidente avec d'autres variables. Il faut garder en mémoire que cet axe a une signification très faible dans cette étude.

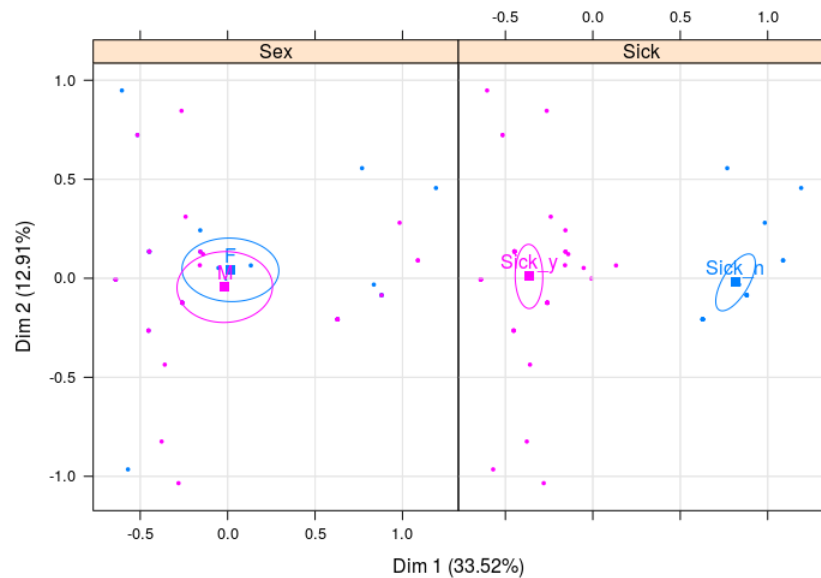
On peut remarquer un groupe d'individus à droite associés aux modalités cheese, mayo, diarrhea, fever, abdo, vomit, icecream nausea tous négatifs associés aux faibles intensités d'intoxication, opposé à un groupe à gauche présentant les modalités positives.

Les aliments cheese, mayo et icecream semblent liés à la sévérité de l'intoxication.

3.5 éléments supplémentaires

3.5.1 variable qualitative supplémentaire

```
plotellipses(acm, keepvar=c(3:4))
```

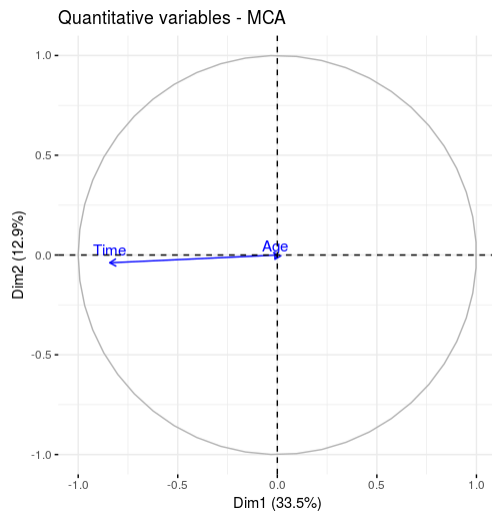


```
acm$quali.sup$v.test
      Dim 1      Dim 2
Sickn  6.9294293 -0.2841346
Sicky -6.9294293  0.2841346
F       0.2047705  0.8860047
M      -0.2047705 -0.8860047
acm$quali.sup$eta2
      Dim 1      Dim 2
Sick  0.8892035226  0.001495045
Sex   0.0007764989  0.014537117
```

Seule la variable sick est fortement liée à l'axe 1 (v-test »2, eta2 proche de 1) comme l'on pouvait s'y attendre, le sexe n'intervenant pas dans cette étude.

3.5.2 variable quantitative supplémentaire

```
fviz_mca_var(res.mca, choice = "quanti.sup", ggtheme = theme_minimal())
```



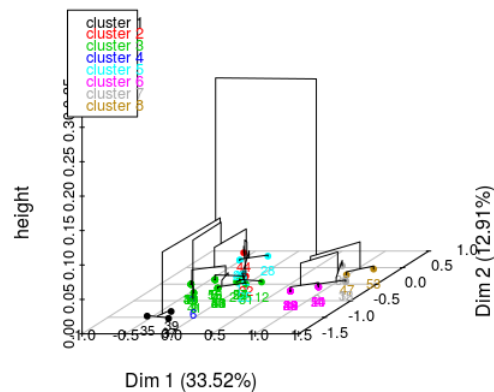
Si time signifie le temps d'indisposition, il est logique de trouver ici une corrélation positive avec la gravité de l'intoxication.

3.6 compléments

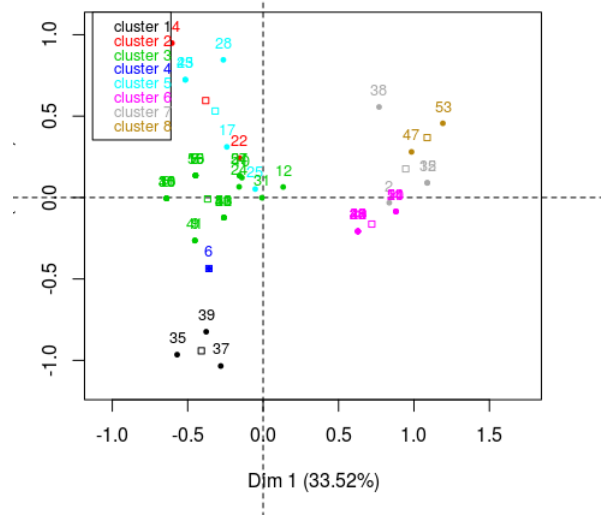
C'est tellement plus simple avec :

```
library(FactoInvestigate)
Investigate(acm)
dimdesc(acm, axes = 1:1)
HCPC(acm)
```

Hierarchical clustering on the factor map



Factor map



La CAH du tableau corrobore les observations précédentes. On observe 8 groupes.

La classe 1 est composée d'individus tels que 35, 37 et 39 : caractérisé par une forte fréquence de la modalité Potato=Potato_n.

La classe 2 contient l'individu 44, par une forte fréquence de la modalité Icecream=Icecream_n.

La classe 3 est composé d'individus partageant une forte fréquence des modalités Sick=Sick_y, Abdominals=Abdo_y, Fever=Fever_y, Diarrhae=Diarrhea_y, Vomiting=Vomit_y,

Cheese=Cheese_y, Mayo=Mayo_y et Courgette=Courg_y (du plus commun au plus rare) et une faible fréquence des modalités Sick=Sick_n, Abdominals=Abdo_n, Fever=Fever_n, Diarrhae=Diarrhea_n, Vomiting=Vomit_n, Cheese=Cheese_n, Mayo=Mayo_n et Courgette=Courg_n (du plus rare au plus commun).

La classe 4 contient l'individu 6 caractérisé par une forte fréquence de la modalité Fish=Fish_n.

Ce groupe est caractérisé par 23, 28 et 45, avec une forte fréquence de la modalité Courgette=Courg_n Courgette=Courg_y.

La classe 6 est caractérisée par 4, 14, 20, 26, 29 et 51, avec une forte fréquence des modalités Sick=Sick_n, Abdominals=Abdo_n, Diarrhae=Diarrhea_n, Fever=Fever_n, Vomiting=Vomit_n et Nausea=Nausea_n (du plus commun au plus rare) et une faible fréquence des modalités Sick=Sick_y, Abdominals=Abdo_y, Fever=Fever_y, Diarrhae=Diarrhea_y, Vomiting=Vomit_y et Nausea=Nausea_y (du plus rare au plus commun).

La classe 7 est caractérisé par 2, 15, 32 et 38, avec une forte fréquence des modalités Cheese=Cheese_n, Sick=Sick_n, Abdominals=Abdo_n, Diarrhae=Diarrhea_n et Fever=Fever_n (du plus commun au plus rare) et une faible fréquence des modalités Cheese=Cheese_y, Sick=Sick_y, Abdominals=Abdo_y, Fever=Fever_y et Diarrhae=Diarrhea_y (du plus rare au plus commun).

La classe 8 est caractérisée par 47 et 53. avec une forte fréquence des modalités Icecream=Icecream_n et Mayo=Mayo_n (du plus commun au plus rare) et une faible fréquence des modalités Icecream=Icecream_y et Mayo=Mayo_y (du plus rare au plus commun).

On pourrait aussi compléter en faisant une analyse DISQUAL pour expliquer silk.