

## TP 4 : ANALYSE FACTORIELLE DES CORRESPONDANCES (AFC)

## 1 Introduction

Plusieurs packages fournissent des outils permettant de réaliser une analyse factorielle des correspondances. On peut citer :

- dans le package **MASS** : `mva`
- dans le package **ade4** : `dudi.acm`
- dans le package **FactoMineR** : `CA`, `MCA`.

Les résultats les plus complets semblent être fournis par le package **FactoMineR**.

Installer puis charger le package **FactoMineR**

```
library(FactoMineR)
```

Une première section présente les généralisations de l'AFC et leurs options R sur un exemple trivial. Cette méthode est ensuite appliquée à des données regroupant le nombre de médailles attribuées lors de plusieurs olympiades et des données de dégustation de vins blancs.

## 2 Données fictives

On illustre ici les différentes façons d'obtenir une généralisation de l'AFC.

### 2.1 Création des données

```
csp <- as.factor(c("csp3","csp1","csp2","csp2","csp3","csp4"))
sport <- as.factor(c("S","R","R","S","T","T"))
fic.tab <- table(sport,csp)
fic.conting <- data.frame(unclass(fic.tab))

# Paramètres graphiques #
x11()
par(mfrow=c(1,3))
```

### 2.2 AFC de la table de contingence

```
fic.afc <- CA(fic.conting,graph=FALSE)
fic.afc
plot(fic.afc,title="AFC \n classique")
```

## 2.3 AFC du tableau disjonctif complet

```
# Fonction construisant le tableau disjonctif complet #
disj.comp <- function(tab) {
  n <- dim(tab)[1]
  m <- dim(tab)[2]
  ci <- sapply(tab,max)
  c <- sum(ci)
  disj.tab <- matrix(0,nrow=n,ncol=c)
  id <- 0
  for (j in 1:m) {
    for (i in 1:n) {
      disj.tab[i,id + tab[i,j]] <- 1
    }
    id <- id + ci[j]
  }
  disj.tab
}

# AFC du tableau disjonctif #
fic.disj.comp <- disj.comp(as.data.frame(cbind(csp,sport)))
colnames(fic.disj.comp) <- c(levels(csp),levels(sport))
fic.disj.afc <- CA(fic.disj.comp,graph=FALSE)
fic.disj.afc
plot(fic.disj.afc,title="AFC \n Tableau Disjonctif")
```

## 2.4 AFC du tableau de Burt

```
fic.burt <- t(fic.disj.comp) %*% fic.disj.comp
fic.burt.afc <- CA(fic.burt,graph=FALSE)
fic.burt.afc
plot(fic.burt.afc,title="AFC \n Tableau de Burt")
```

Comparer les résultats (valeurs propres, graphiques) obtenus par AFC de différents tableaux construits à partir des deux mêmes variables fictives.

## 3 Données J.O.

Le tableau des données croise en ligne les épreuves d'athlétisme et en colonnes les différents pays. Chaque cellule contient le nombre total de médailles obtenues lors des olympiades de 1992 à 2008.

Il s'agit bien d'un tableau de contingence, les individus sont les 360 médailles. A chaque médaille sont associées deux variables qualitatives : l'épreuve à laquelle elle se réfère et le pays auquel appartient celui qui l'a obtenue.

L'objectif est de répondre aux problématiques suivantes :

- Peut-on considérer que tous les pays ont le même profil athlétique ou, au contraire, certains pays sont-ils plutôt performants dans certaines épreuves ?

- Peut-on considérer que toutes les épreuves ont le même profil géographique ou, au contraire, certaines épreuves sont-elles la spécialité de certains pays ? Peut-on synthétiser ces spécialisations en faisant apparaître des oppositions entre d'une part des épreuves ayant le même profil (i.e. remportées par les mêmes pays) et d'autre part les épreuves ayant un profil opposé (i.e. remportées par d'autres pays) ?
- Charger les données et réaliser l'AFC du tableau de contingence.

```
data(J0)
J0
J0.ca <- CA(J0)
J0.ca
```

- Choisir le nombre d'axes à retenir.

```
round(J0.ca$eig,2)
barplot(J0.ca$eig[,1],main="Valeurs propres",
        names.arg=paste("dim",1:nrow(J0.ca$eig)))
```

- Réaliser les graphiques des représentations simultanées sur le premier et le second plan factoriel. Commenter les représentations obtenues en définissant les profils athlétiques des pays et les profils géographiques des épreuves.

```
plot(J0.ca)
plot(J0.ca,axes=3:4)
```

- Analyser les contributions de chaque pays sur les deux premiers axes factoriels.

```
# Contributions pays à l'axe 1 #
round(J0.ca$col$contrib[rev(order(J0.ca$col$contrib[,1])),1],2)
# Contributions pays à l'axe 2 #
round(J0.ca$col$contrib[rev(order(J0.ca$col$contrib[,2])),2],2)
```

- Conclure en répondant aux problématiques citées.

## 4 Données Vins Blancs

Dans le cadre d'une recherche sur la typicité des vins provenant du cépage Chenin en Val de Loire, un ensemble de dix vins blancs secs de Touraine, cinq Touraine AOC, issus du cépage Sauvignon, et 5 Vouvray AOC, issus du cépage Chenin, ont été livrés à une analyse sensorielle, combinant différents types de jury et différents protocoles.

Ces données sont rassemblées dans un tableau dont les lignes sont les vins, dont les colonnes sont les mots et dont le terme général  $x_{ij}$  est le nombre de fois que le mot  $j$  a été associé au vin  $i$ . Ce tableau peut être vu comme un tableau de contingence en considérant que l'on dispose de  $n$  descriptions sensorielles (une description est l'association entre un mot et un vin) et que ces descriptions sont ventilées selon deux variables qualitatives : le vin auquel elles se réfèrent d'une part et le mot utilisé d'autre part.

L'objectif de l'AFC est de fournir une image synthétique de la diversité de ces vins.

- Charger les données `vins` et réaliser l'AFC du tableau de contingence.

```
vins <- read.table(file.choose(),header=TRUE,row.names=1,sep=";")
vins
vins.ca <- CA(vins,col.sup=11,row.sup=nrow(vins))
vins.ca
```

- Choisir le nombre d'axes à retenir.

```
round(vins.ca$eig,2)
barplot(vins.ca$eig[,1],main="Valeurs propres",
        names.arg=1:nrow(vins.ca$eig))
```

- Réaliser les graphiques des représentations simultanées sur le premier et le second plan factoriel. Commenter les représentations obtenues en définissant les typicités de chaque vin ou de groupes de vins.

```
plot(vins.ca)
plot(vins.ca,axes=3:4) # Souci d'exhaustivité
```