

## TP 4: Analyse des Correspondances

### **Exercice 1: Travail des femmes - Analyse des correspondances binaires**

On utilise ici un tableau de données issu d'un questionnaire réalisé sur des françaises en 1974, disponible sur le site du cours: women\_work.txt

#### **1) Présentation du tableau de données**

1724 femmes ont répondu à différentes questions à propos du travail des femmes, parmi lesquelles :

- Quelle est selon vous la famille parfaite ?
  - L'homme et la femme travaillent
  - L'homme travaille plus que la femme
  - Seul l'homme travaille
- Quelle activité est la meilleure pour une mère quand les enfants vont à l'école ?
  - Rester à la maison
  - Travailler à mi-temps
  - Travailler à temps complet
- Que pensez-vous de la phrase suivante : les femmes qui ne travaillent pas se sentent coupées du monde ?
  - Complètement d'accord
  - Plutôt d'accord
  - Plutôt en désaccord
  - Complètement en désaccord

Le tableau de données à analyser est formé de deux tableaux de contingence qui croisent les réponses de la première question à celles des deux autres.

Pour charger le package et le tableau de données, écrivez la ligne de code suivante :

```
> library(FactoMineR)
> women_work=read.table("women_work.txt", header=TRUE,
row.names=1, sep="\t")
```

#### **2) Première ACB**

On utilisera les trois premières colonnes (correspondant aux réponses de la deuxième question) comme colonnes actives et les quatre dernières (correspondant à la troisième question) comme colonnes illustratives. À quelle préoccupation correspond ce choix ?

#### ***Lignes et colonnes actives seulement***

Pour voir les nuages des lignes et des colonnes, tapez :

```
> res.ca.act = CA(women_work[,1:3])
```

Interprétez les résultats de cette première ACB: inerties, axes s'ils sont interprétables (s'aider des CO2 et CTR), relations barycentriques.

### *Addition de colonnes illustratives*

On ajoute les colonnes qui correspondent à la troisième question en tant que variables illustratives:

```
> res.ca.actsup = CA(women_work, col.sup=4:ncol(women_work))
```

Enrichir l'interprétation de la première ACB à l'aide de ces projections.

### **3) Seconde ACB**

On utilisera maintenant les quatre dernières colonnes (correspondant à la troisième question) comme colonnes actives et les trois premières colonnes (correspondant aux réponses de la deuxième question) comme colonnes illustratives. À quelle préoccupation correspond ce choix ?

Reprendre pour ce choix toutes les sous-questions de la question (2).

## **Exercice 2: Comment consomme-t-on le thé? - Analyse des correspondances multiples.**

On utilise ici un tableau de données issu d'un questionnaire sur la consommation de thé présent dans la package FactoMineR:

```
> data(tea)
```

### *Présentation des données*

300 consommateurs de thé ont répondu à un questionnaire sur leur consommation de thé.

Les questions portaient sur leur façon de consommer le thé et leur image du thé. Le questionnaire comportait également des questions descriptives telles que le sexe, l'âge, la catégorie socio-professionnelle et la pratique régulière d'un sport.

Excepté l'âge, toutes les variables sont qualitatives. Le tableau de données comporte deux variables différentes pour l'âge : une continue et une qualitative.

### **1) L'ACM**

a) Effectuez les manipulations suivantes, examinez les résultats, puis modifiez les commandes pour modifier le rendu des résultats selon ce que vous voulez.

```
> res.mca = MCA(tea, quanti.sup=19, quali.sup=c(20:36))
> plot.MCA(res.mca, invisible=c("var","quali.sup"), cex=0.7)
> plot.MCA(res.mca, invisible=c("ind","quali.sup"), cex=0.7)
> plot.MCA(res.mca, invisible=c("ind"))
> plot.MCA(res.mca, invisible=c("ind", "var"))
```

#tea: le tableau de données utilisé

#quanti.sup: vecteur des index des variables continues  
illustratives

#quali.sup: vecteur des index des variables qualitatives  
illustratives

#invisible: les éléments à ne pas représenter

```
#cex: taille des caractères
```

b) Quel pourcentage d'information les deux premières dimensions captent-elles?

N.B. Pour obtenir une description des dimensions, tapez :

```
> dimdesc(res.mca)
```

Quelles sont les variables et les individus liés à chacune des deux premières dimensions?

## **2) Etude des individus :**

a) Y a-t-il des groupes d'individus ?

b) Signalez et décrivez les individus "atypiques".

## **3) Etude des variables et des modalités :**

Les questions se posent dans le même ordre que pour l'ACP. Premièrement, on veut étudier les associations entre modalités. Deuxièmement, on recherche une ou plusieurs variable(s) synthétique(s) continue(s) pour résumer les variables qualitatives. Troisièmement, on cherche à caractériser des groupes d'individus par des modalités.

Dans cette étude, les variables sur l'attitude de consommation sont actives et les autres variables sont illustratives.

a) Quelles sont les modalités opposées par la première dimension? Comment interprétez-vous alors cette dimension?

b) Mêmes questions pour la deuxième dimension.

c) La variable "age" est-elle bien représentée? Quelle est sa corrélation avec la deuxième dimension? Qu'est-ce qui distingue partiellement l'achat de thé par les jeunes et les plus âgés?

d) Cachez les modalités actives pour s'intéresser seulement aux modalités illustratives. Les modalités de la variable "age\_Q" sont-elles ordonnées le long de la deuxième dimension? Expliquez comment ceci corrobore la projection en quantitative supplémentaire de la variable "age".

## **4/ Ellipses de confiance**

Il est possible de tracer, dans les plans factoriels, des ellipses de confiance autour des centres de gravité correspondant aux modalités de variables qualitatives supplémentaires. Ces ellipses utilisent la normalité asymptotique de ces centres de gravité.

On utilise la fonction *plotellipses()* :

```
> plotellipses(res.mca, keepvar=c(20:23))
```

```
#res.mca: le résultat d'une ACM
```

```
#keepvar: un vecteur d'index (ou de noms) des variables à représenter
```