

TP: Analyse Factorielle des Correspondances simple

Analyse des données

Master ISEFAR - M1

1 Packages R utiles

```
library("tidyverse")
library("ggplot2") #pour avoir de 'beaux' graphiques
library("FactoMineR") #pour effectuer l'ACP
library("factoextra") #pour extraire et visualiser les résultats issus de FactoMineR
library("questionr")
library("scales")
library("ca")
```

2 Quelques scripts utiles: rappels profils-ligne et profils-colonne, test du χ^2 et réalisation l'AFC (sur jeu de données HEC ici)

```
#####
# Fréquences marginales
P=HEC/n
#ligne
rowSums(P)
#colonne
colSums(P)

##### Profils
# profil-ligne
Profil.Ligne =rprop(HEC,digits = 2)
knitr::kable(Profil.Ligne)
row<-rprop(HEC,total=FALSE) %>% as.data.frame()
row %>% ggplot(aes(x=Hair,fill=Eye))+geom_bar(aes(y=Freq), stat="identity")

# profil-colonne
Profil.Colonne =cprop(HEC,digits = 2)
knitr::kable(Profil.Colonne)
col <- cprop(HEC,total=FALSE) %>% as.data.frame()
col %>% ggplot(aes(x=Eye,fill=Hair))+geom_bar(aes(y=Freq), stat="identity")

# graphe mosaic
mosaicplot(HEC, main = "Relation between hair and eye color")
```

```

mosaicplot(t(HEC), main = "Relation between hair and eye color")

#####
# Test du  $\chi^2$ 
n <- sum(HEC)
h.chi <- chisq.test(HEC)
phi <- h.chi$statistic/n
sum(res.ca$eig[,1])

#####
# AFC
res.ca=PCA(HEC,graph=FALSE)

# Valeurs propres
knitr::kable(res.ca$eig)
fviz_eig(res.ca, addlabels = TRUE)

# Qualité de représentation: cos2
rowSums(res.ca$row$cos2[,1:2]) #ligne
rowSums(res.ca$col$cos2[,1:2]) #colonne

# Contributions
res.ca$row$contrib[,1:2] #ligne
res.ca$col$contrib[,1:2] #colonne

### Graphe de l'AFC
fviz_ca_biplot(res.ca, pointsize = 2, labelsize = 3)
#plot(res.ca, autoLab = "yes")
# bleu: profils-ligne
# rouge: profils-colonne

### Taux de liaison
P = HEC/sum(HEC)
DI <- apply(P, 1, sum)
DJ <- apply(P, 2, sum)
T = (P-t(DI))%*%DJ/(t(DI))%*%DJ
knitr::kable(T,digits=2)

```

3 Couleur des yeux / Couleur des cheveux

On reprend le jeu de données `HairEyeColor` étudié dans le TP d'analyse descriptive bivariée et décrivant la couleur des cheveux et des yeux de 592 individus (sans prendre en compte le sexe). On avait vu que les deux variables n'étaient pas indépendantes. On effectue une AFC afin de voir si on peut établir une typologie des individus, suivant leur couleur de cheveux et des yeux.

1. Charger les données et les packages utiles.
2. Effectuer une AFC

- a. Combien vaut l'inertie totale?
- b. Vérifier le lien entre l'inertie totale et la statistique de test testant l'indépendance entre les deux variables.
- c. Justifier le nombre d'axes total obtenus.
- d. Quels sont les points bien projetés?
- e. Quels sont les modalités qui contribuent aux deux premiers axes?
- f. Faire le graphe des deux nuages de points. Interpréter.

4 Fumeurs

On considère les données 'smoke' issues du package **FactoMineR**. Elles correspondent au tableau de contingence de 4 catégories de fumeur (en colonne) pour 5 catégories de salariés (en ligne) dans une entreprise. Les catégories sont SM=Senior Managers, JM=Junior Managers, SE=Senior Employees, JE=Junior Employees, SC=Secretaries.

Effectuer une AFC. Pour obtenir les données, faire

```
library(ca)
data(smoke)
```

5 Loisirs & Age

On considère le jeu de données 'TV.txt' qui représente le tableau de contingence de 7 catégories de loisirs pour 4 classes d'âge d'une population.

Effectuer une AFC. Pour obtenir les données, faire

```
LoisirsAge <- read.table("TV.txt",header=TRUE,sep=" ",row.names =1)
```