

TP N°5----ANALYSE FACTORIELLE DES CORRESPONDANCES (AFC)

I. Enoncé 1:

Soit le jeu de données suivant (Catégorie-socioprofessionnelle, lieu de séjour en vacances):

Hotel Location Res.Second Parents Amis Camping Sej.org Autres

Agriculteurs 195 62 1 499 44 141 49 65

Patrons 700 354 229 959 185 292 119 140

Cadres.sup 961 471 633 1580 305 360 162 148

Cadre.moy 572 537 279 1689 206 748 155 112

Employes 441 404 166 1079 178 434 178 92

Ouvriers 783 1114 387 4052 497 1464 525 387

Autres.actifs 142 103 210 1133 132 181 46 59

Inactifs 741 332 327 1789 311 236 102 102

Identifier les correspondances
possibles.

```
library("FactoMineR")
library("factoextra")
library(graphics)
CSP_Vacation <-
read.csv("D:/COURSES_FSTBM/MASTER_ID_1/AD_R_2021/CSP_Vacation.txt",
row.names=1, sep=" ", stringsAsFactors=TRUE)
View(CSP_Vacation)
library("gplots")
# 1. convertir les données en tant que table de contingence
dt <- as.table(as.matrix (CSP_Vacation))

# 2. Graphique
balloonplot(t(dt), main = "CSP_Vacation", xlab = "", ylab = "", label = FALSE,
show.margins = FALSE)

summary(dt)

boxplot(CSP_Vacation)
boxplot(t(CSP_Vacation))

# ou
khd=chisq.test(CSP_Vacation)
```

khd

```
I <- dim(CSP_Vacation)[1] # n rows
J <- dim(CSP_Vacation)[2] # n cols
I
J
dimIJ <- (I-1)*(J-1)
Beta <- (khd$Statistic - dimIJ)/(sqrt(dimIJ))
Beta

#### AFC par factoextra ####

res.afc <- CA(CSP_Vacation) #
#round(res.afc$eig,3) #???[1:3]

# Affichage du screen plot
library(factoextra)
plot(res.afc$eig[,1], type="b", xlab="eigen values", ylab="values", main="screen plot")

fviz_eig(res.afc, addlabels = TRUE, ylim = c(0, 50))
fviz_screplot (res.afc) +
  geom_hline (yintercept = 33.33, linetype = 2, color = "red")

#### Profil ligne

Lig <- get_ca_row(res.afc)
Lig
Lig$inertia
Lig$coord

# Graphique
fviz_ca_row (res.afc, col.row = "red", shape.row = 15)
fviz_ca_row (res.afc, col.row = "cos2",
  gradient.cols = c ("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE)
fviz_ca_row (res.afc, col.row = "contrib",
  gradient.cols = c ("#00AFBB", "#E7B800", "#FC4E07"),
  repel = TRUE)

library("corrplot")
corrplot(res.afc$row$contrib, is.corr=FALSE)

# Contributions des lignes ? la dimension 1
fviz_contrib(res.afc, choice = "row", axes = 1)
# Contributions des lignes ? la dimension 2
fviz_contrib(res.afc, choice = "row", axes = 2)

# Aussi, on peut identifier les contributions du profil-ligne i ? l'inertie de l'axe dim 1 :
# cont >= (1/|L|= 4= 25%)
res.afc$row$contrib
# on identifie .....
res.afc$row$coord # le signe donne le sens de contribution
# l'axe 1 oppose ....
# l'axe 2 ....
```

Profil Colonne

```
col <- get_ca_col(res.afc)
col
```

```
fviz_ca_col (res.afc, col.col = "cos2",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
fviz_ca_col (res.afc, col.col = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel = TRUE)
```

```
library("corrplot")
corrplot(col$contrib, is.corr=FALSE)
```

```
fviz_cos2 (res.afc, choice = "col", axes = 1)
fviz_cos2 (res.afc, choice = "col", axes = 2)
fviz_cos2 (res.afc, choice = "col", axes = 1:2)
```

```
# Contributions des lignes ? la dimension 1
fviz_contrib(res.afc, choice = "col", axes = 1)
#fviz_contrib(res.afc, choice = "row", axes = 1)
# Contributions des lignes ? la dimension 2
fviz_contrib(res.afc, choice = "col", axes = 2)
fviz_contrib(res.afc, choice = "col", axes = 1:2)
```

```
### Contribution du profil-colonne j ? l'inertie de l'axe dim 1 :
# cont >= (1/|C|= 4= 25%)
res.afc$col$contrib
# on identifie
res.afc$col$coord # le signe donne le sens de contribution
# l'axe 1 caract?rise ...
# l'axe 2 .....
```

```
### Graphe simultan? /biplot
# Sym?trique
fviz_ca_biplot(res.afc, repel = TRUE)
```

```
# Asym?trique
fviz_ca_biplot (res.afc,
               map = "rowprincipal", arrow = c(TRUE, TRUE),
               repel = TRUE)
```

Si l'angle entre deux flèches est aigu, alors il y a une forte association entre les lignes et les colonnes correspondantes.

Pour interpréter la distance entre les lignes et les colonnes, vous devriez projeter perpendiculairement des points lignes sur la flèche de la colonne.

la contribution des points (M. Greenacre 2013)

les distances entre les points lignes et l'origine du graphique sont liées à leurs contributions aux axes principaux en considération.
Plus une flèche est proche (en termes de distance angulaire) d'un axe, plus la contribution de la ligne sur cet axe par rapport à l'autre axe est importante. Si la flèche est à mi-chemin entre les deux axes, la ligne contribue aux deux axes de manière identique.

```
fviz_ca_biplot(res.afc, map = "colgreen", arrow = c(TRUE, FALSE),  
               repel = TRUE)
```

```
fviz_ca_biplot(res.afc, map = "rowgreen", arrow = c(FALSE, TRUE),  
               repel = TRUE)
```

Dimension

Description de la dimension

Pour identifier facilement les lignes et les colonnes les plus associées aux axes principaux

```
res.desc <- dimdesc(res.afc, axes = c(1, 2))
```

```
res.desc[[1]]$row # par les lignes
```

```
res.desc[[1]]$col # par les colonnes
```

```
res.desc[[2]]$row
```

```
res.desc[[2]]$col
```

II. **Enoncé 2:**

Soit le jeu de données suivant (profession et type de médias suivis):

Profession	Radio	Tel.	Quot.Nat.	Quot.Reg.	Press.Mag.	Press.TV
Agriculteur	96	118	2	71	50	17
Petit.Patr.	122	136	11	76	49	41
Prof.Cad.Sup	193	184	74	63	103	79
Prof.Int.	360	365	63	145	141	184
Employe	511	593	57	217	172	306
Ouvr.Qualif.	385	457	42	174	104	220
Ouvr.Non-Qual.	156	185	8	69	42	85
Inactif	1474	1931	181	852	642	782

Identifier les correspondances possibles.
