

Exercice 3: Durée de thèse selon les domaines scientifiques

```
dat = read.table("fr-esr-effectifs-doctorants-docteurs-ecoles-doctorales-
durees-these-domaines.csv", header=TRUE, sep=";", stringsAsFactors=FALSE,
quote="\")
dim(dat)
```

```
## [1] 11249    23
```

Le jeu de données contient 11249 individus et 23 variables

Question 3 (Jeu de données modifié et le nombre de modalités de la variable
DOMAINE_SCIENTIFIQUE)

```
dat_mod <- na.omit(dat)
levels(dat_mod$DOMAINE_SCIENTIFIQUE)
```

```
## NULL
```

```
table(dat_mod$DOMAINE_SCIENTIFIQUE)
```

```
##
##                               Biologie, médecine et santé
##                               408
##                               Chimie
##                               296
##                               Mathématiques et leurs interactions
##                               274
##                               Physique
##                               276
##                               Sciences agronomiques et écologiques
##                               214
##                               Sciences de la société
##                               463
##                               Sciences de la terre et de l'univers, espace
##                               252
## Sciences et technologies de l'information et de la communication
##                               332
##                               Sciences humaines et humanités
##                               518
##                               Sciences pour l'ingénieur
##                               365
```

Question 4 (Tableau de zeros 10 fois 4)

```

tableau <- array(0, dim = c(10, 4))
print(tableau)

```

```

##      [,1] [,2] [,3] [,4]
## [1,]    0    0    0    0
## [2,]    0    0    0    0
## [3,]    0    0    0    0
## [4,]    0    0    0    0
## [5,]    0    0    0    0
## [6,]    0    0    0    0
## [7,]    0    0    0    0
## [8,]    0    0    0    0
## [9,]    0    0    0    0
## [10,]   0    0    0    0

```

Question 5

```

domaines <- unique(dat$DOMAINE_SCIENTIFIQUE)
domaines <- domaines[!is.na(domaines) & domaines != ""]
tableau <- matrix(0, nrow = length(domaines), ncol = 4)
for (i in 1:length(domaines)) {
  domaine_data <- subset(dat, DOMAINE_SCIENTIFIQUE == domaines[i])
  resultat <- c(
    sum(domaine_data$MOINS_DE_40_MOIS, na.rm = TRUE),
    sum(domaine_data$DE_40_A_52_MOIS, na.rm = TRUE),
    sum(domaine_data$DE_52_A_72_MOIS, na.rm = TRUE),
    sum(domaine_data$PLUS_DE_6_ANS, na.rm = TRUE)
  )
  tableau[i, ] <- resultat
}
print(tableau)

```

```

##      [,1] [,2] [,3] [,4]
## [1,] 2235 1327 179 16
## [2,] 3885 5093 1049 151
## [3,] 680 1458 2152 2329
## [4,] 2665 1617 142 11
## [5,] 1375 854 187 38
## [6,] 3115 3188 583 60
## [7,] 1073 2155 3275 3375
## [8,] 1055 868 125 25
## [9,] 2888 2565 539 73
## [10,] 910 993 112 13

```

Question 6

```

tab_contingence <- as.table(tableau)
row.names(tab_contingence) <- domaines
colnames(tab_contingence) <- c('Moins de 40 mois', 'Entre 40 et 52 mois',
'Entre 52 et 72 mois', 'Plus de 6 ans')
print(tab_contingence)

```

Moins de 40 mois:

Physique
(2235)
Biologie, médecine et santé
(3885)
Sciences de la société
(680)
Chimie
(2665)
Mathématiques et leurs interactions
(1375)
Sciences pour l'ingénieur
(3115)
Sciences humaines et humanités
(1073)
Sciences de la terre et de l'univers, espace
(1055)
Sciences et technologies de l'information et de la communication
(2888)
Sciences agronomiques et écologiques
(910)

Entre 40 et 52 mois

Physique
(1327)
Biologie, médecine et santé
(5093)
Sciences de la société
(1458)
Chimie
(1617)
Mathématiques et leurs interactions
(854)
Sciences pour l'ingénieur
(3188)
Sciences humaines et humanités
(2155)
Sciences de la terre et de l'univers, espace
(868)
Sciences et technologies de l'information et de la communication
(2565)
Sciences agronomiques et écologiques
(993)

Entre 52 et 72 mois

Physique
(179)
Biologie, médecine et santé
(1049)
Sciences de la société

```

(2152)
Chimie
(142)
Mathématiques et leurs interactions
(187)
Sciences pour l'ingénieur
(583)
Sciences humaines et humanités
(3275)
Sciences de la terre et de l'univers, espace
(125)
Sciences et technologies de l'information et de la communication
(539)
Sciences agronomiques et écologiques
(112)
Plus de 6 ans
Physique
(16)
Biologie, médecine et santé
(151)
Sciences de la société
(2329)
Chimie
(11)
Mathématiques et leurs interactions
(38)
Sciences pour l'ingénieur
(60)
Sciences humaines et humanités
(3375)
Sciences de la terre et de l'univers, espace
(25)
Sciences et technologies de l'information et de la communication
(73)
Sciences agronomiques et écologiques
(13)

```

```
str(tab_contingence)
```

```

## 'table' num [1:10, 1:4] 2235 3885 680 2665 1375 ...
## - attr(*, "dimnames")=List of 2
## ..$ : chr [1:10] "Physique" "Biologie, médecine et santé" "Sciences de
la société" "Chimie" ...
## ..$ : chr [1:4] "Moins de 40 mois" "Entre 40 et 52 mois" "Entre 52 et 72
mois" "Plus de 6 ans"

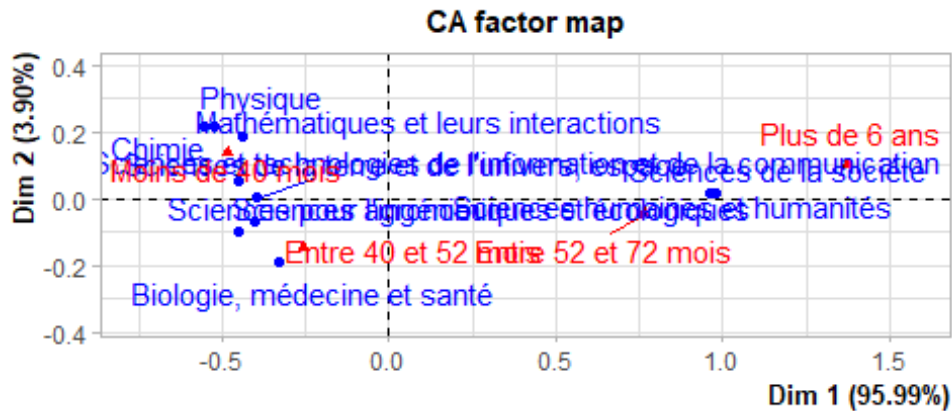
```

Question 7 (AFC)

```

library(FactoMineR)
afc_result <- CA(tab_contingence, graph = TRUE)

```



Interpretation de l'AFC

1) le pourcentage d'inertie est de $95,99 + 3,90 = 99,89$ et donc la qualité de cette AFC est très bonne.

2) Dans les domaines scientifiques: Mathématiques et leurs interactions, Chimie, Physique et sciences de la terre et de l'univers les doctorants ont tendance à finir sur une durée de Moins de 40 mois.

3) Les doctorants dans les domaines comme: Biologie, Médecine et santé, sciences agronomiques et écologiques et sciences pour l'ingénieur finissent sur une durée entre 40 et 52 mois.

4) Ceux des sciences et technologies de l'information sont bien répartis entre les variables "Moins de 40 mois" et "Entre 40 et 52 mois", on peut dire qu'ils finissent leur cycle en moins de 52 mois. Et ceux des sciences de la société et sciences humaines et humanités mettent plus de temps comparés aux autres parce que c'est entre 52 et 72 mois ou plus de 6 ans.

Globalement on peut dire que la durée de thèse pour les doctorants dans les domaines sociales et humaines est plus importante que celle dans les domaines des sciences exactes.

...