2016 - 2017

TECHNOLOGIE DE L’E-COMMERCE ET MOBILES

Florent Cardoen & Jérémy Bastin

Trames SSL vues sur WireShark

Exercices DataMining

**setwd('/Users/florentcardoen/Dropbox/HEPL/Troisième/e-commerce/partie 2/')**

# 3.1

**data <- read.table("ex1.csv", h=TRUE, sep=';')**

**data**

enfants appareils

1 1 7

2 2 3

3 3 1

4 4 2

5 5 6

6 6 8

**attach(data)**

**enfants**

[1] 1 2 3 4 5 6

**reg <- lm(enfants~appareils, data=data)**

**reg**

Call:

lm(formula = enfants ~ appareils, data = data)

Coefficients:

(Intercept) appareils

2.6867 0.1807

**summary(reg)**

Call:

lm(formula = enfants ~ appareils, data = data)

Residuals:

1 2 3 4 5 6

-2.9518 -1.2289 0.1325 0.9518 1.2289 1.8675

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) **2.6867** 1.6255 1.653 0.174

appareils **0.1807** 0.3119 0.580 0.593

Residual standard error: 2.009 on 4 degrees of freedom

Multiple R-squared: 0.07745, Adjusted R-squared: -0.1532

F-statistic: 0.3358 on 1 and 4 DF, p-value: **0.5933**

**cor(data)**

enfants appareils

enfants 1.0000000 **0.2783032**

appareils **0.2783032** 1.0000000

**plot(enfants~appareils)**

**abline(lm(enfants~appareils, data=data))**

p-value est très grand, on ne rejette pas l’hypothèse nulle

Rejete l’existence d’une corrélation

0.27 basse corrélation.

Droite de régression : y= 2,6867x + 0.1807

(8-0.1807)/2.6867=x

x= 2.9103

# 3.2

## Exercice 1

**tab = matrix(c(25, 215, 200, 60, 15, 207, 194, 60), ncol=2)**

**tab**

[,1] [,2]

[1,] 25 15

[2,] 215 207

[3,] 200 194

[4,] 60 60

**chi2<-chisq.test(tab)**

**chi2**

Pearson's Chi-squared test

data: tab

X-squared = 2.1542, df = 3, p-value = 0.541

On ne rejete pas l'hypothèse qui dit que le type d'achat n'est pas lié au sexe

## Exercice 2

**data <- read.csv("ex3\_2\_2.csv", header=TRUE, sep = ";", row.names=1)**

**data**

csp type

1 Employe Camping

2 Ouvrier Camping

3 Ouvrier Camping

4 Cadremoyen Camping

5 Employe Locationgite

6 Divers Camping

7 Agriculteur Camping

…

98 Agriculteur Familleamis

99 Ouvrier Camping

100 Ouvrier Camping

//on fait une table pour pouvoir faire le test du chi-carré de R

**tablekhi<-table(data$csp,data$type)**

**tablekhi**

Camping Familleamis Hotel Locationgite

Agriculteur 2 8 0 2

Cadremoyen 4 1 2 5

Chefentreprise 1 1 5 3

Divers 3 2 1 9

Employe 8 3 1 3

Ouvrier 9 3 0 2

Retraite 5 13 2 2

//Graphique

**par(mfrow=c(2,4))**

**barplot(tablekhi[1,], main="Agriculteur")**

**barplot(tablekhi[2,], main="Cadre moyen")**

**barplot(tablekhi[3,], main="Chef")**

**barplot(tablekhi[4,], main="Divers")**

**barplot(tablekhi[5,], main="Employe")**

**barplot(tablekhi[6,], main="Ouvrier")**

**barplot(tablekhi[7,], main="Retraite")**

//Test chi-carré => On rejette l’hypothèse nulle

**khitest <- chisq.test(tablekhi)**

Warning message:

In chisq.test(tablekhi) : l'approximation du Chi-2 est peut-être incorrecte

> khitest

Pearson's Chi-squared test

data: tablekhi

X-squared = 56.06, df = 18, p-value = 8.72e-06

// Tester la variable qui a le plus d’influence sur le résultat

**contributions<-round(khitest$residuals^2/khitest$stat\*100,1)**

**contributions**

Camping Familleamis Hotel Locationgite

Agriculteur 1.6 8.8 2.4 0.7

Cadremoyen 0.0 3.5 0.6 2.0

Chefentreprise 2.7 2.5 24.7 0.1

Divers 1.2 2.7 0.5 11.9

Employe 3.8 1.0 0.5 0.4

Ouvrier 8.1 0.7 2.7 1.3

Retraite 1.1 10.0 0.1 4.3

# 3.3

## Exercice 1

**summary(data)**

PAIN BISC VIN BIER

Min. : 87.0 Min. : 1.000 Min. : 63.00 Min. : 12.00

1st Qu.:109.0 1st Qu.: 2.750 1st Qu.: 68.75 1st Qu.: 44.75

Median :124.5 Median : 4.500 Median : 74.00 Median : 61.00

Mean :127.1 Mean : 4.875 Mean : 96.12 Mean : 58.88

3rd Qu.:144.0 3rd Qu.: 6.250 3rd Qu.:123.00 3rd Qu.: 74.75

Max. :167.0 Max. :11.000 Max. :163.00 Max. :111.00

PATA LEG FRU PLATS

Min. :27.00 Min. : 3.000 Min. : 4.00 Min. : 6.00

1st Qu.:33.50 1st Qu.: 4.750 1st Qu.: 6.75 1st Qu.:15.75

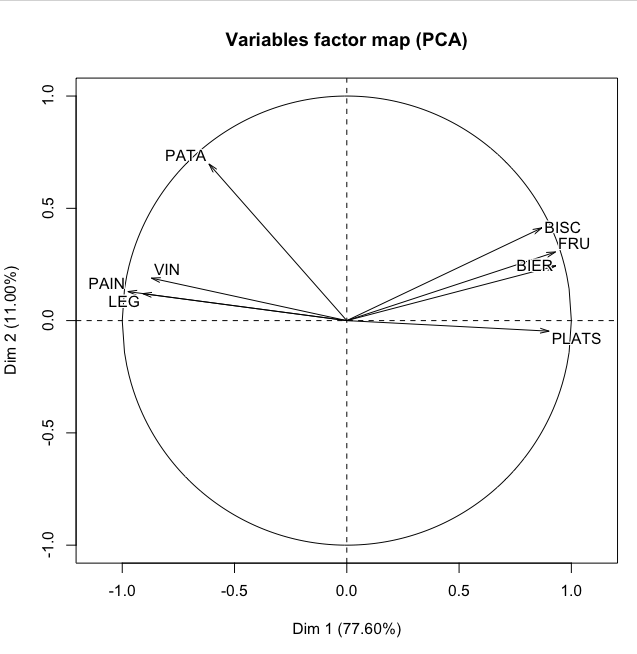
Median :39.50 Median : 6.500 Median :10.50 Median :24.00

Mean :38.62 Mean : 6.625 Mean :10.12 Mean :24.38

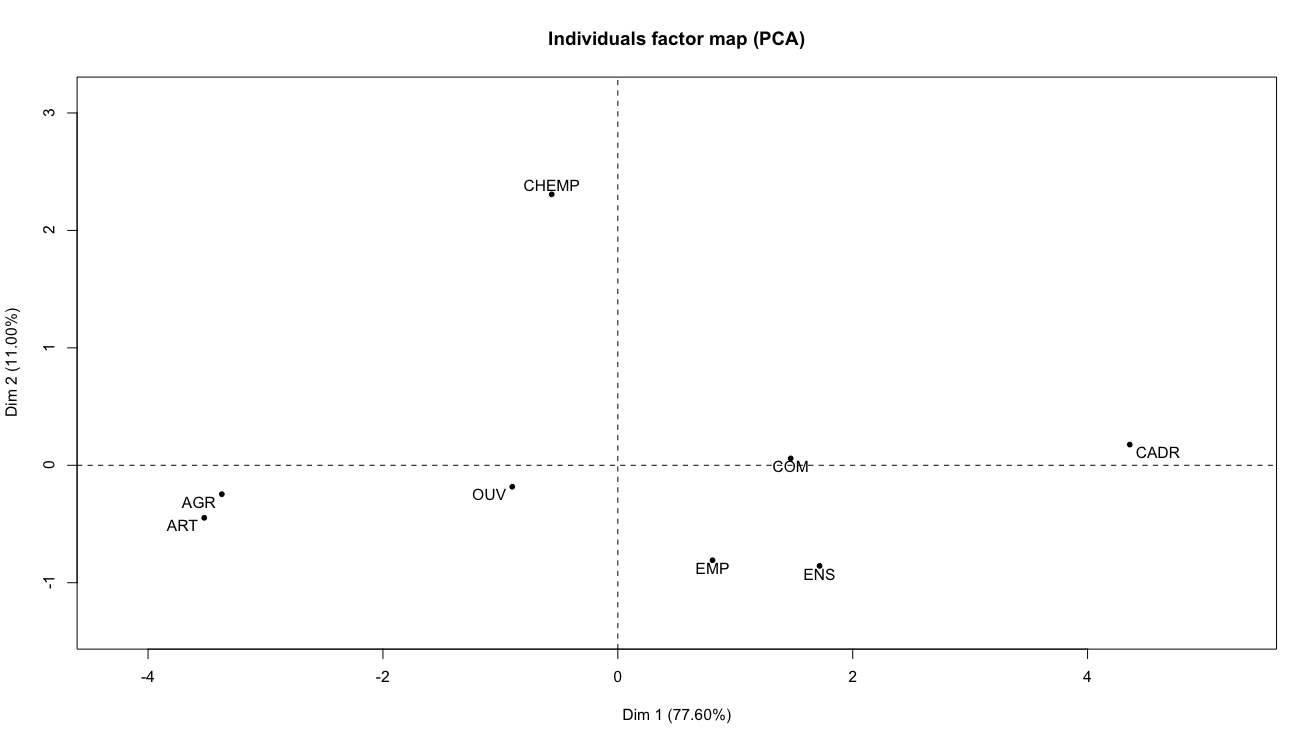
3rd Qu.:41.50 3rd Qu.: 8.000 3rd Qu.:12.25 3rd Qu.:32.25

Max. :53.00 Max. :12.000 Max. :18.00 Max. :41.00

**acp <- PCA(data, graph=T)**



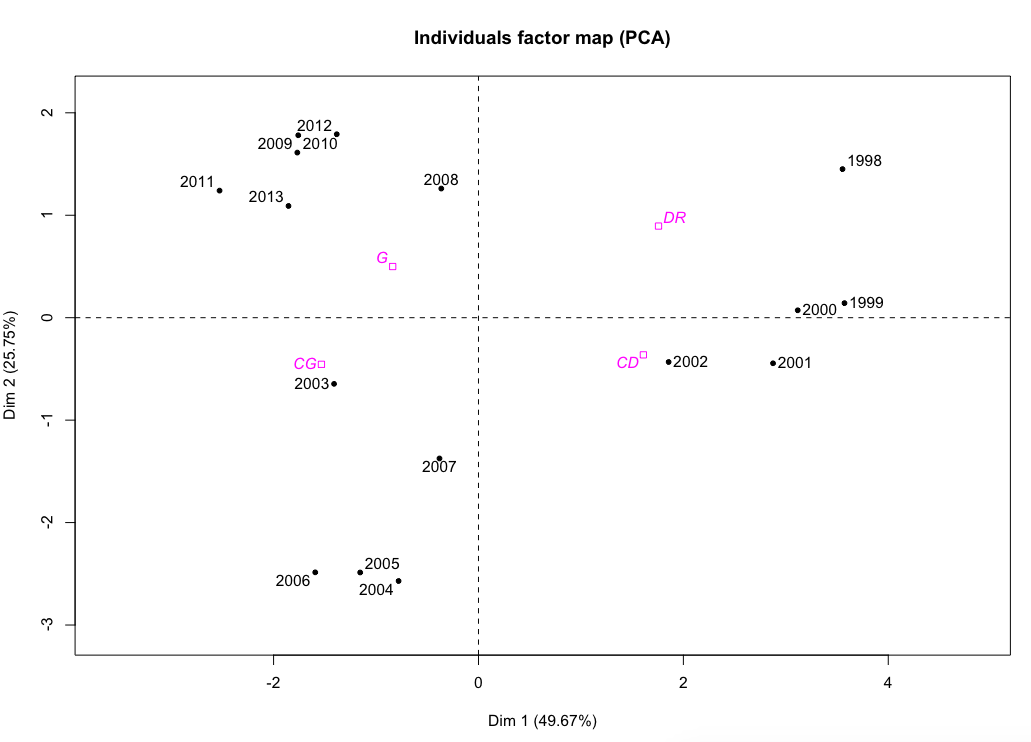
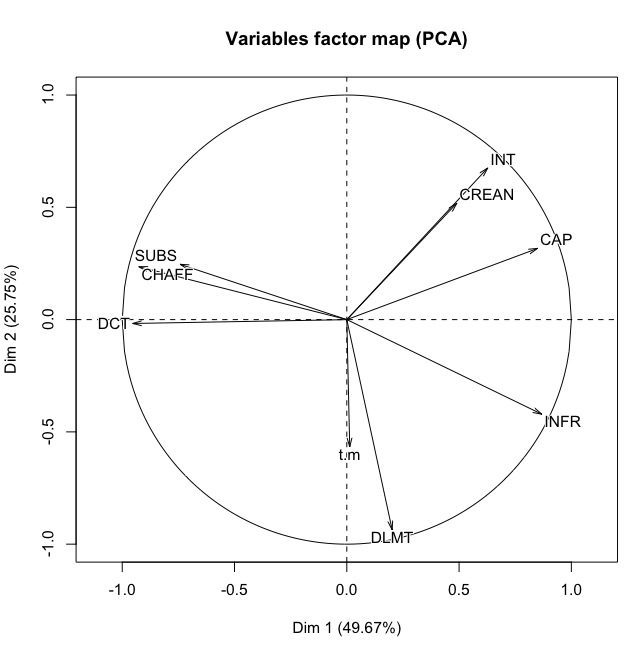
L’axe 1 oppose les individus qui consomment des pains ordinaires, des légumes et du vin « gros rouge » à ceux qui mangent des fruits, de la bière, des plats préparés et des biscuits.

****

Il y a une opposition entre les habitudes alimentaires des agriculteurs et des cadres. Les autres catégories socio-professionnelles sont bien représenter par l’axe dans l’ordre croissant. Les catégories plus modestes consomment des produits plus basiques, tandis que les cadres mangent des choses plus recherchées.

On peut voir aussi que les personnes recherchant de l’emploi comme plus de pommes de terre.

## Exercice 2



Les postes dettes à court terme, chiffre d’affaire et les subventions accordées sont opposé à l’ensemble des capitaux net et Immobilisation. Tout est bien représenté sauf les dettes à long terme et VRD.

Le bilan de l’année 2011 est en opposition avec les bilans des années entre 2001 et 1998