

# ACP

*Matthieu Lucas Etienne Azat*

*11 février 2019*

```
load(".RData")
```

## Analyse des Composantes principales

Réalisation de 3 ACP : 1 sur le jeu de données avec les scores sans les q4 (car non numériques) 1 sur le jeu de données avec seulement deux q8 (car corrélées) 1 sur le jeu de données avec l'essentiel des q4 et des q8

### JDD 1

```
require(FactoMineR)
```

```
## Loading required package: FactoMineR
```

```
require(factoextra)
```

```
## Loading required package: factoextra
```

```
## Loading required package: ggplot2
```

```
## Welcome! Related Books: `Practical Guide To Cluster Analysis in R` at https://goo.gl/13EFCZ
```

```
# summary(dataSum)
```

```
#soit acp sur tout sauf la Q4 car vectorielle
```

```
#soit afm
```

```
# colnames(dataSum)
```

```
noms<-c("Pédagogie", "T1","T2","T3",  
        "T5","T61","T62","T71","T72","T81",  
        "T82","T83","T84","T85","T86","T87",  
        "T88","T89")
```

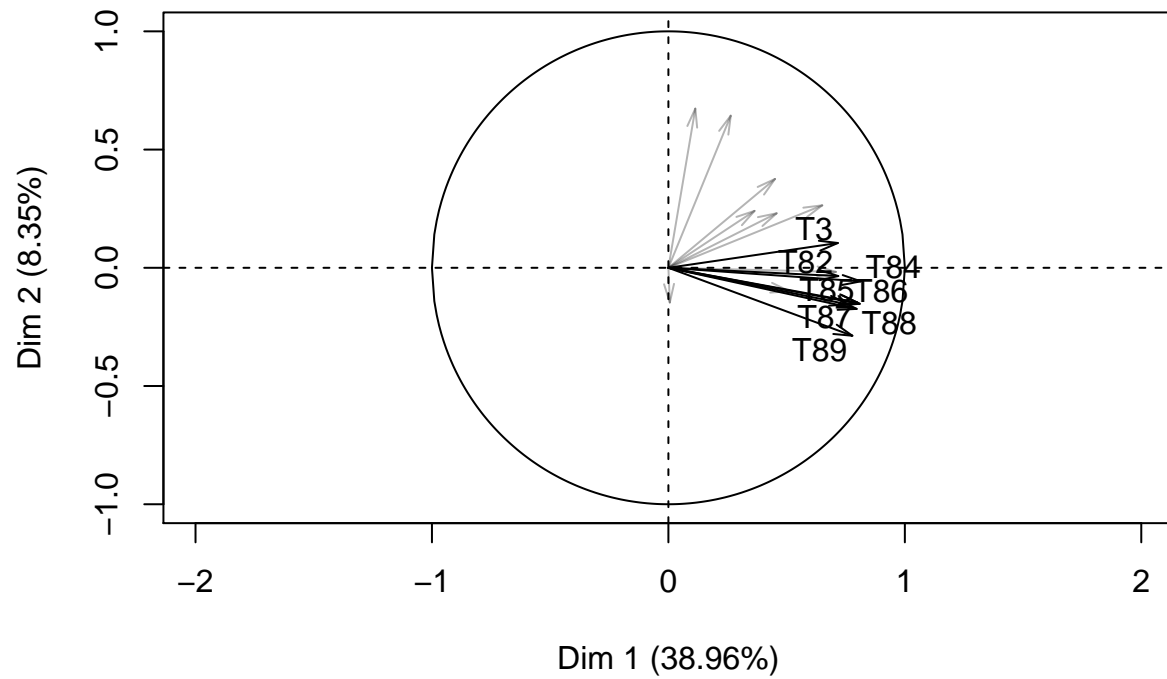
```
valquanti<-dataSum[,noms]
```

```
valquanti[,2:length(valquanti)]<-scale(valquanti[,2:length(valquanti)])
```

```
res.pca<-PCA(valquanti,quali.sup = 1,graph = FALSE)
```

```
plot.PCA(res.pca,choix = "var",select = "contrib 8")
```

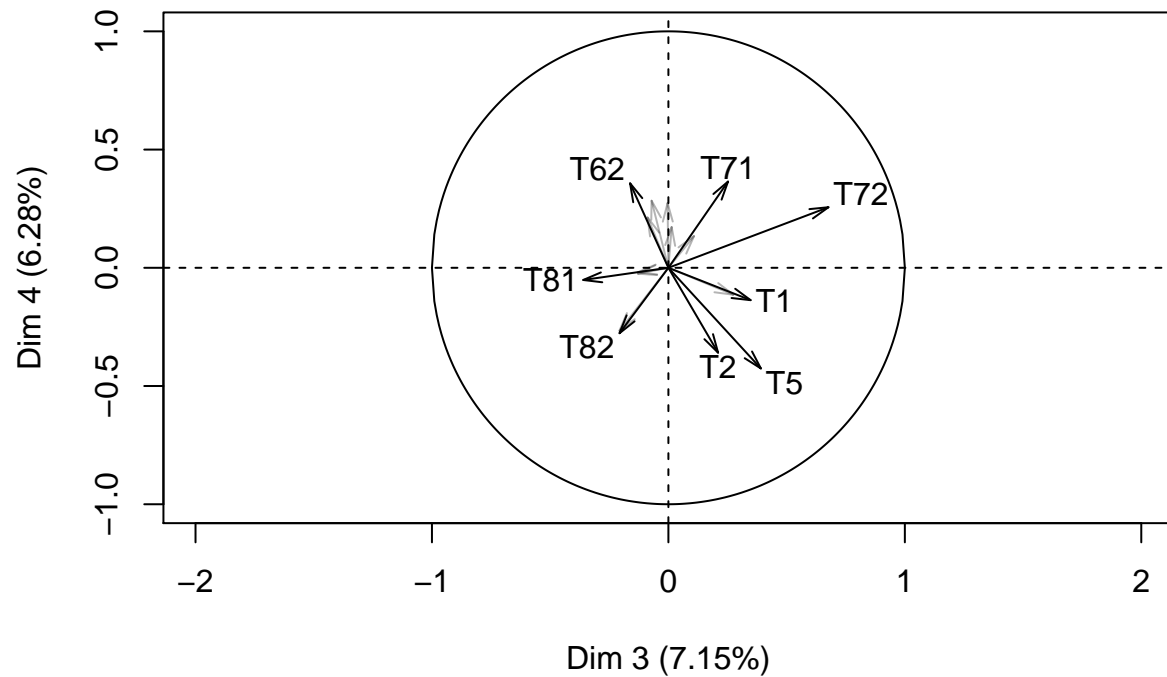
## Variables factor map (PCA)



Les questions 8 ressortent le plus (puis la 3), comme dans les précédentes analyses : cela nous pousse à voir la matrice des corrélations. Seulement 60% de l'info sur les 4 premières dimensions.

```
plot.PCA(res.pca, choix = "var", select = "contrib 8", axes = c(3,4)) #Dim 3 > T_72
```

## Variables factor map (PCA)



```
# summary(res.pca)
```

Dimension 3 expliquée par la T72.

## Matrice des corrélations

```
# install.packages("corrplot")
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.5.2
```

```
## corrplot 0.84 loaded
```

```
summary(dataPropre)
```

```
## Experimentateur Pedagogie Classe TypeClasse
## JB :43 P1:96 MS-8 :26 moyens grands :12
## PC :47 P2:60 Classe 3:13 petits moyens :38
## NA's:66 Classe 6:13 petits moyens grands:40
## 9 :12 NA's :66
## MS-1 :10
## MS-5 :10
```

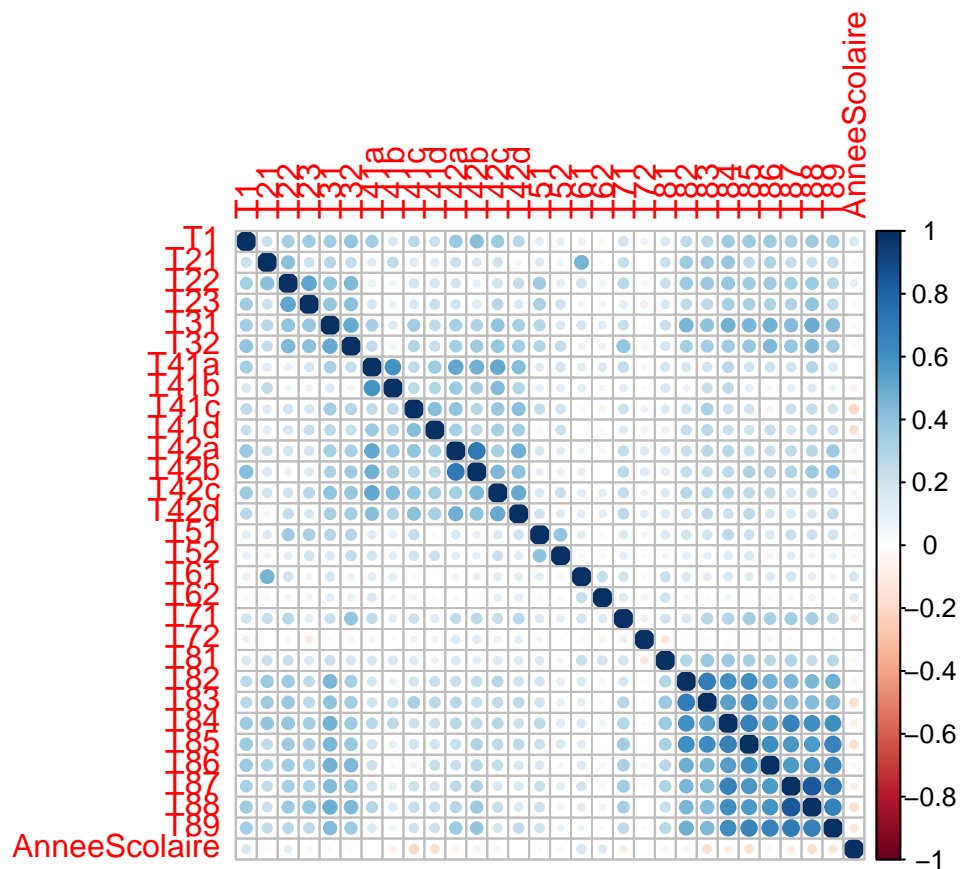
```

##                               (Other) :72
##      Sexe                      Langues
## F   :71   français              : 42
## M   :81   français et arabe     : 2
## NA's: 4    français et tunisien : 1
##                français et turque : 1
##                français et vietnamien: 1
##                NA's              :109
##
##                               Lateralite
## ? (dit droite, utilise les deux pour dessiner): 1
## D                                           :81
## G                                           : 8
## NA's                                       :66
##
##
##      DateNaissance                DateEval
## Min.   :2010-11-27 00:00:00   Min.   :2016-01-18 00:00:00
## 1st Qu.:2011-07-21 00:00:00   1st Qu.:2016-01-25 00:00:00
## Median :2012-03-25 12:00:00   Median :2017-01-24 00:00:00
## Mean   :2012-04-19 05:50:46   Mean   :2016-12-07 08:18:27
## 3rd Qu.:2013-01-18 00:00:00   3rd Qu.:2018-01-22 00:00:00
## Max.   :2013-12-12 00:00:00   Max.   :2018-10-29 00:00:00
##
##      AgeChar      AgeNum      AgeInt      T1      T21
## Length:156      Min.   :4.107   4:139   Min.   : 0.00   0: 21
## Class :character 1st Qu.:4.365   5: 17   1st Qu.: 4.00   1:135
## Mode  :character Median :4.664                Median : 10.00
##                Mean   :4.635                Mean   : 11.83
##                3rd Qu.:4.895                3rd Qu.: 15.00
##                Max.   :5.782                Max.   :100.00
##
##      T22      T23      T31      T32      T41a      T41b      T41c      T41d      T42a
## 0:74      0:120      0:63      0:113      0:99      0:100      0:81      0:84      0:119
## 1:82      1: 36      1:93      1: 43      1:57      1: 56      1:75      1:72      1: 37
##
##
##
##
##
##      T42b      T42c      T42d      T51      T52      T61      T62      T71      T72
## 0:133      0:123      0:123      0:124      0:150      0: 9      0: 39      0:96      0:112
## 1: 23      1: 33      1: 33      1: 32      1: 6      1:147      1:117      1:60      1: 44
##
##
##
##
##      T81      T82      T83      T84      T85      T86      T87      T88      T89
## 0: 13      0: 41      0: 37      0: 54      0:62      0:74      0:78      0:82      0:85
## 1:143      1:115      1:119      1:102      1:94      1:82      1:78      1:74      1:71
##
##

```

```
##
##
##
##  AnneeScolaire
##  15/16:66
##  16/17:47
##  17/18:43
##
##
##
##
```

```
m.cor<-cor(sapply(dataPropre[,13:ncol(dataPropre)],as.numeric))#matrice des corrélations pour les quest
corrplot(m.cor,method = "circle")
```

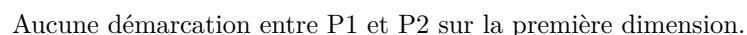


Grosse corrélation entre les 8\* : donc ACP biaisée (et légère sur les q4 mais suffisante pour biaiser l'analyse)  
Etonnement la T1 ne ressort pas comme grosse contrib

## JDD 2

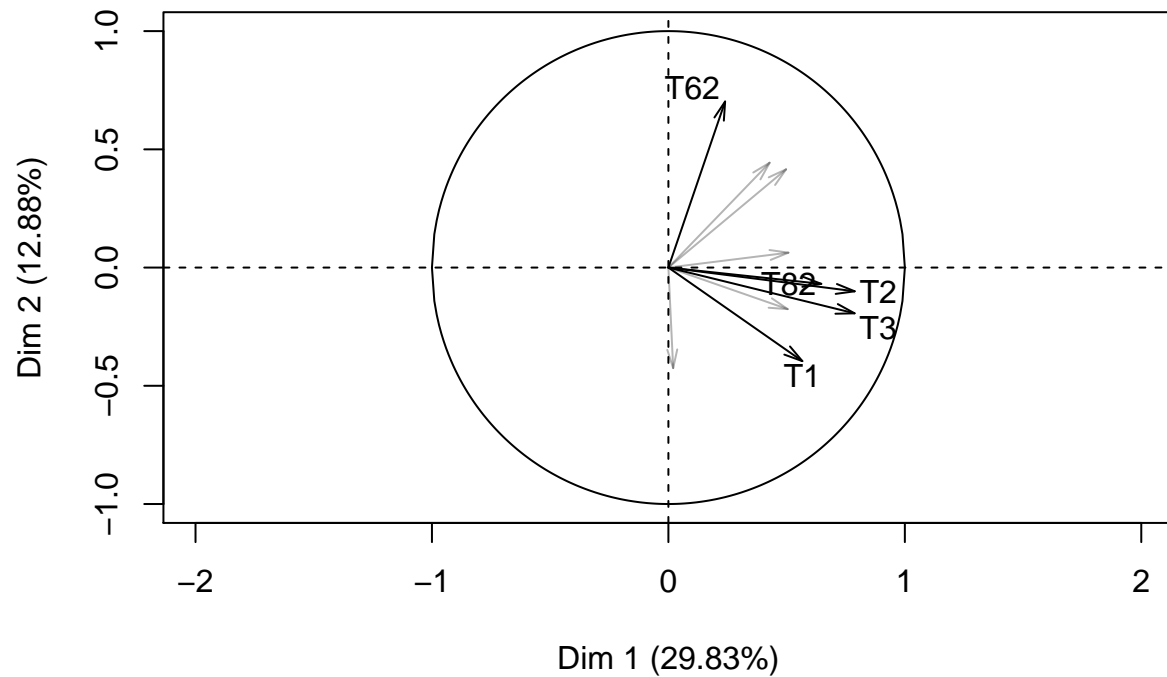
```
##Nouvelle PCA sans les T8 super corrélées entre elles (que T81 et T82 car ce sont les moins corrélées)
valquanti1<-valquanti[, -c(12:18)]
summary(valquanti1)
```

```
res.pca1<-PCA(valquanti1,quali.sup = 1,graph = FALSE)
plot.PCA(res.pca1,axes = c(1,2),choix = "ind")
```



```
plot.PCA(res.pca1,axes = c(1,2),choix = "var",select = "cos2 5")
```

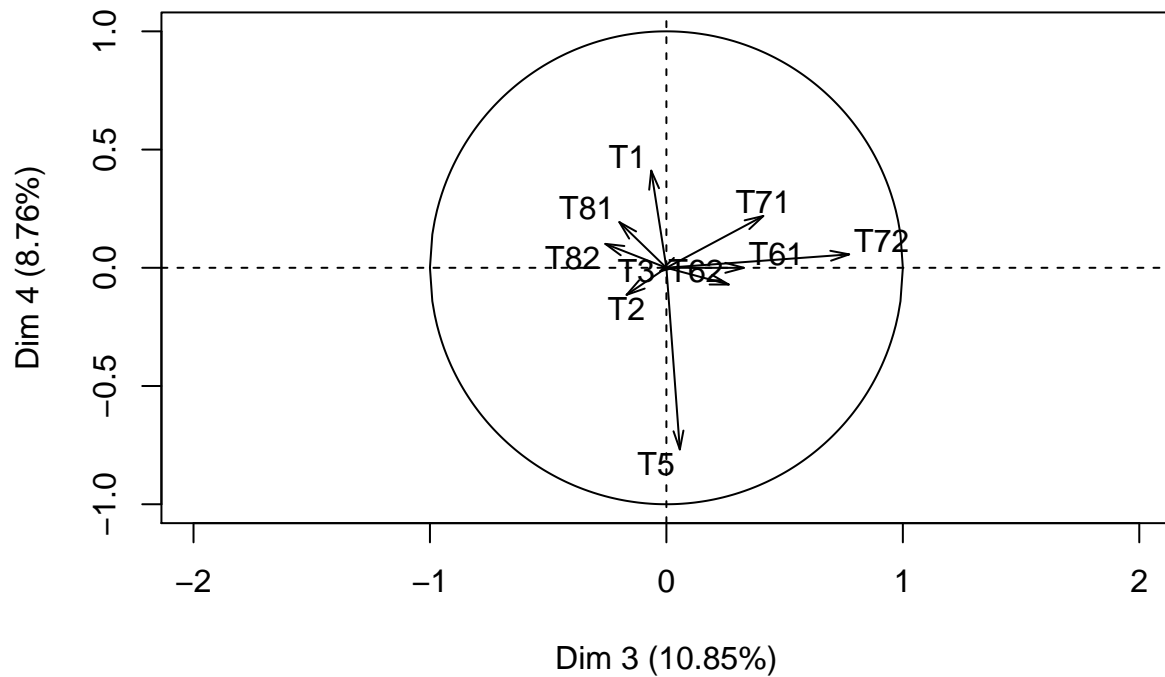
### Variables factor map (PCA)



T2 / T3 sont proches et ont un cos2 élevé : un gros score en T2 implique un gros score en T3 La dimension 1 porte sur les T2 et T3 La deuxième dimension porte sur la T62

```
plot.PCA(res.pca1,axes = c(3,4),choix = "var")
```

## Variables factor map (PCA)



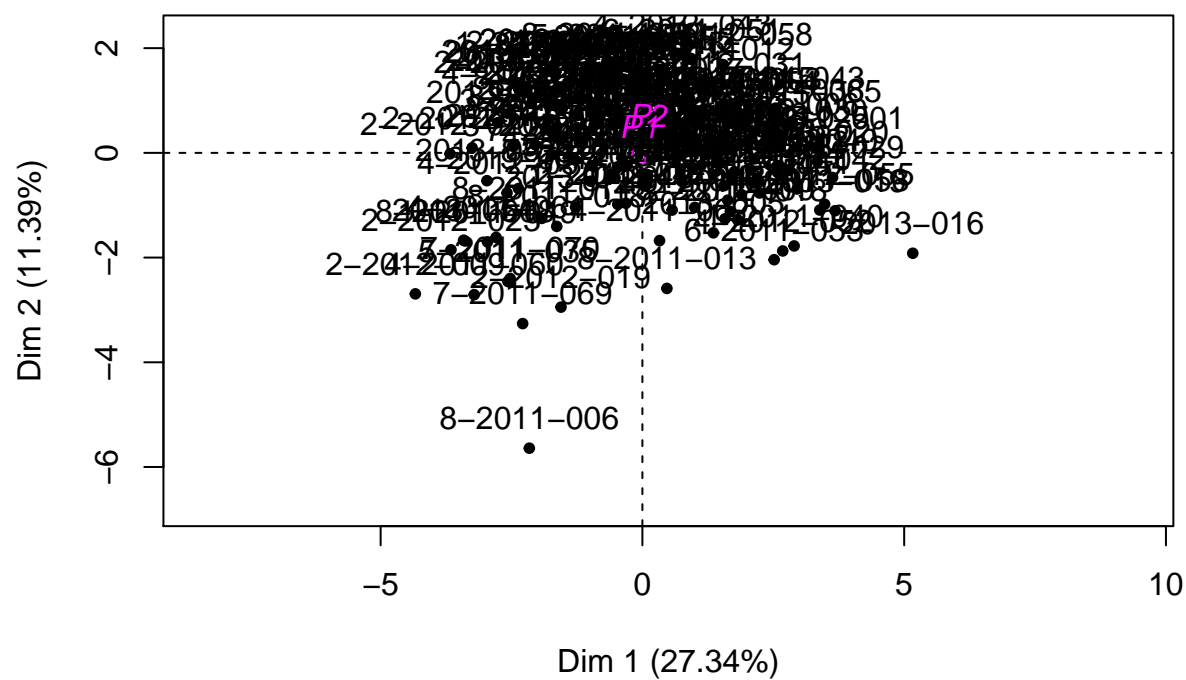
La dimension 3 porte sur la T72 La dimension 4 porte sur la T5 En conséquent nous avons quasi 1 variable / axe l'utilité de l'acp peut être remise en question

## JDD 2

```
#Ajout d'une acp avec les q4 de datapropre (seulement celles les plus corrélées aux autres)
valquanti2<-cbind(valquanti1,
                  apply(dataPropre[,c("T41a","T41c","T41d")],2,as.numeric))
res.pca1<-PCA(valquanti2,quali.sup = 1,graph = FALSE)
plot.PCA(res.pca1,axes = c(1,2),choix = "ind")
```



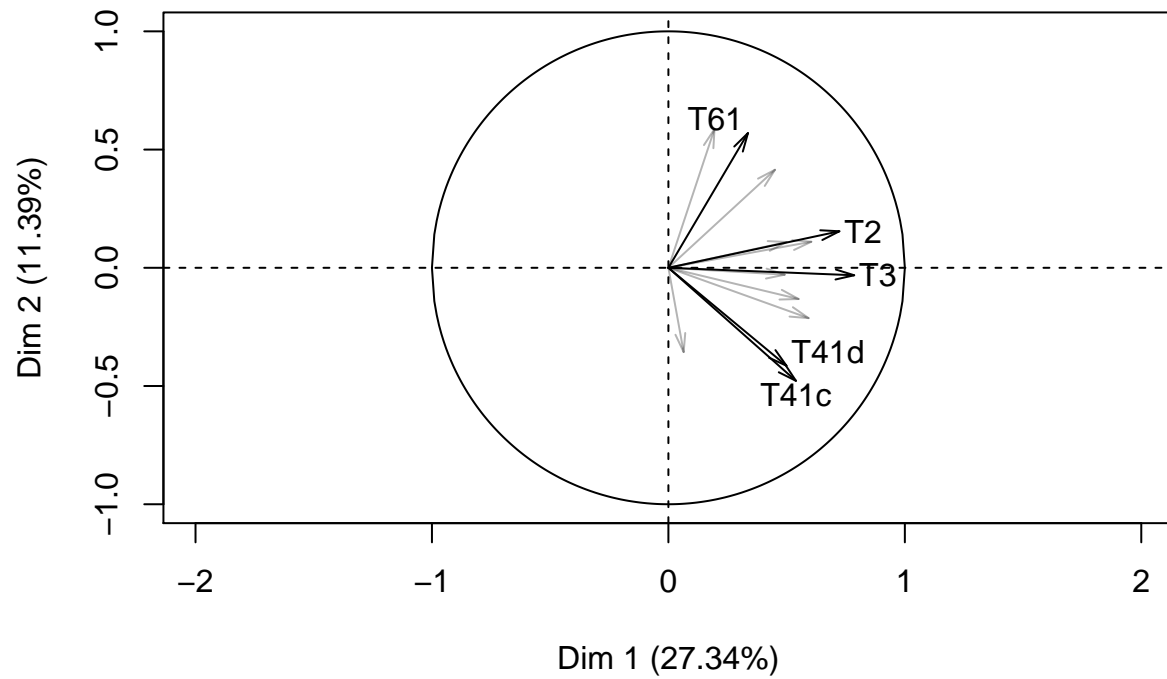
## Individuals factor map (PCA)



A nouveau aucune démarcation entre P1 et P2

```
plot.PCA(res.pca1, axes = c(1,2), choix = "var", select = "cos2 5")
```

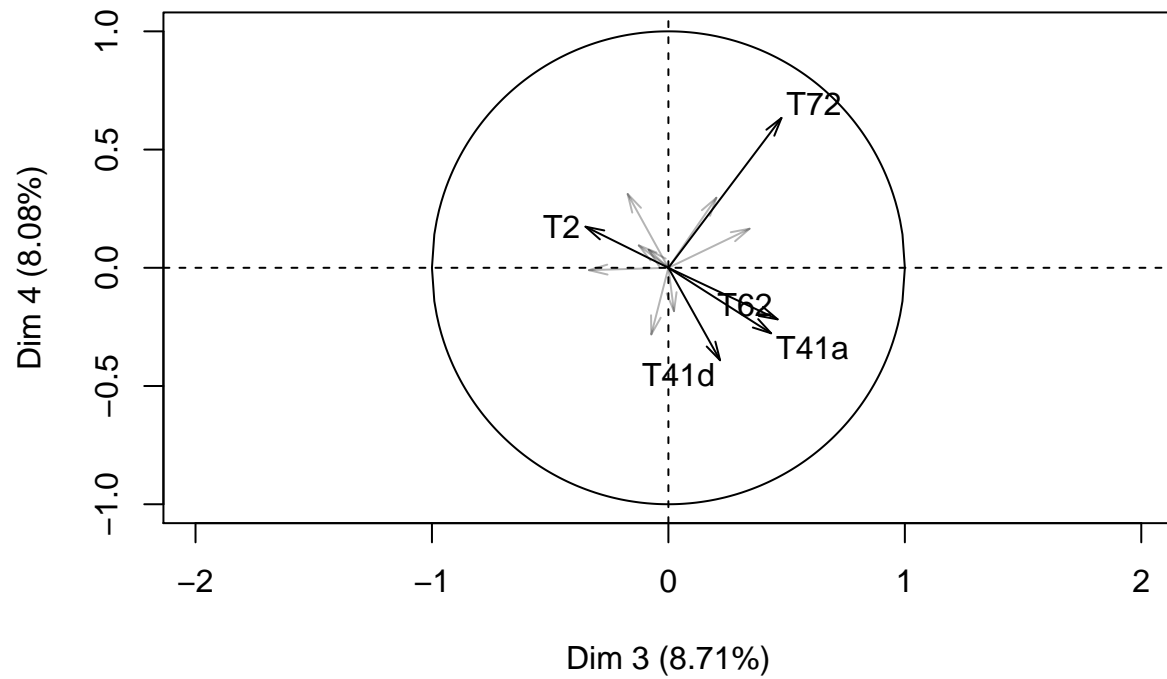
## Variables factor map (PCA)



La dimension 1 porte sur les questions T2 et 3 La dimension 2 porte sur les questions T41c/d (très corrélé alors qu'on ne le voit pas dans le cor) et T61

```
plot.PCA(res.pca1,axes = c(3,4),choix = "var",select = "cos2 5")
```

### Variables factor map (PCA)



Les dimensions 3 et 4 portent sur la question T72

L'ACP permet donc dans les 3 cas d'observer des différences au sein de la population, mais qui n'est pas significative avec la pédagogie suivie par les individus.