

## TP 3 : Analyse en Composantes Principales (ACP) avec le logiciel R

### 1 Exemple d'une ACP avec le logiciel R

On considère le tableau de données brutes (contenues dans le fichier "eaux.txt") :

Nom de l'eau	intensité d'émission de bulles	saveur salée	appréciation globale
St Yorre	3,9	6,4	2,9
Vichy	1,4	6,0	2,8
Quézac	5,1	4,7	3,5
Salvetat	2,9	4,1	3,4
Perrier	8,2	4,9	2,8

On désire faire une ACP (par défaut ACP centrée réduite) de ce petit jeu de données avec le logiciel R. Pour cela, on va utiliser la fonction `PCAmix` (pour "Principal Component Analysis for mixed data") d'un package dédié à l'analyse des données, le package `PCAmixdata`. Les lignes de commandes suivantes permettent de réaliser cette ACP et d'obtenir les sorties numériques et graphiques nécessaires à l'interprétation :

```
#-----
# Chargement du jeu de donnees
#-----
eaux <- read.table("eaux.txt",header=TRUE,row.names=1,sep=";")
eaux
#-----
# Chargement du package PCAmixdata
#-----
require(PCAmixdata) # permet de charger le package "PCAmixdata"
                    # afin de pouvoir l'utiliser par la suite
help(PCAmix)
#-----
# Mise en oeuvre de l'ACP
#-----
res<-PCAmix(eaux) # tous les calculs de l'ACP sont stockes dans l'objet "res"
                  # NB : par default les graphiques des plans factoriels 1-2
                  # sont affiches a l'ecran
res <- PCAmix(eaux,graph=FALSE) # idem sans les graphiques
res # permet de voir l'ensemble des sorties numeriques disponibles
#-----
# Choix du nombre d'axes à retenir
#-----
round(res$eig,digit=2) # permet d'afficher les valeurs propres et les pourcentages
                      # de variances expliquees par chaque axe
# Graphique de l'ebouli des valeurs propres
barplot(res$eig[,1],main="Eigenvalues",names.arg=1:nrow(res$eig))
abline(h=1,col=2,lwd=2)
#-----
# Graphiques des individus et des variables sur le plan factoriel 1-2
#-----
?plot.PCAmix # permet d'afficher la fenetre d'aide de la commande "plot.PCA"
plot(res,axes=c(1,2),choice="ind") # on retrouve ici le graphique des individus (plan 1-2)
plot(res,axes=c(1,2),choice="cor") # on retrouve ici le cercle des corr\elations
                                   # des variables (plan 1-2)
plot(res,axes=c(1,2),choice="sqload") # on retrouve ici le graphique des "square loadings" (plan 1-2)
#-----
# Sorties numeriques pour les individus et es variables
#-----
res$ind # permet d'afficher l'ensemble des sorties numeriques associees aux individus :
        # coordonnees, contributions, cosinus carres
round(res$ind$cos2,digit=3) # uniquement les cosinus carres

res$quanti # permet d'afficher l'ensemble des sorties numeriques associees aux variables :
          # coordonnees, contributions, cosinus carres
round(res$quanti$cos2,digit=3) # uniquement les cosinus carres
```

**Travail à réaliser.** Faire l'ACP en utilisant les commandes ci-dessus et commenter les différentes sorties graphiques et numériques.

## 2 Autre jeu de données pour faire une ACP

Voici ci-après le jeu de données `decathlon` avec lequel vous pouvez faire également une ACP. Ces données sont disponibles par défaut lorsque le package `PCAmixdata` a été téléchargé. Elles concernent les résultats à des épreuves de décathlon de plusieurs athlètes. Seules les  $p = 10$  premières variables nous intéressent ici (il s'agit de celles correspondant aux résultats des 10 épreuves du décathlon). Les autres variables peuvent éventuellement être mises en variables supplémentaires ou illustratives (deux sont quantitatives et une est qualitative). Pour accéder à ce jeu de données, il suffit de taper dans votre console R :

```
data(decathlon)
```

**Remarque.** Il existe d'autres packages R dédiés à l'analyse des données (donc incorporant la méthode d'ACP). En particulier, on peut mentionner le package `FactoMineR`.