

TP3 : Analyse en Composantes Principales

On a relevé les notes de quelques élèves dans 7 matières : le Français, le Latin, le Dessin, la Musique, le Sport, les Maths et la Physique. Le tableau suivant donne les notes après centrage et réduction. Afin de voir s'il y a des ressemblances entre élèves et des corrélations entre matières, on veut réaliser une ACP sur ce tableau. Le tableau de données "notes1314.txt" est disponible à l'adresse

<https://synapse.math.univ-toulouse.fr/index.php/s/11DANH3PBqzTpWo>

1. Télécharger le tableau de données, enregistrez le dans le dossier contenant votre script R et étudier sa structure. Charger le tableau dans R.

Calculer le tableau T_s des variables numériques centrées réduites.

Calculer la matrice des corrélations. Quelles variables semblent corrélées?

```
notes <- read.table(file="notes 13-14.txt",
  colClasses = c("character",rep("numeric",7)),header=TRUE,row.names=1)
# ou directement,
notes <- read.table("http://www.math.univ-toulouse.fr/l33/fichiers/notes%2013-14.txt",
  colClasses = c("character",rep("numeric",7)),header=T,row.names=1)
```

```
Ts <-as.matrix(notes)
cov(Ts) # Ou bien cor(Ts)
```

2. Diagonaliser la matrice des corrélations avec la fonction `eigen`.
Afin d'avoir une idée du nombre d'axes factoriels à considérer dans la suite, réaliser un diagramme présentant les valeurs propres ainsi qu'un diagramme présentant les pourcentages cumulés de ces valeurs propres. Combien d'axes sont nécessaires ici?

```
E<-eigen(cov(Ts));E
par(mfrow=c(1,2))
barplot(E$values,xlab="Valeurs Propres",cex.axis=2,cex.names=2)
barplot(cumsum(E$values / sum(E$values) * 100),xlab="Sommes cumulées (%)",cex.axis=2,cex.names=2)
par(mfrow=c(1,1))
```

3. Calculer le tableau T_p des coordonnées des individus dans le nouveau repère.
Afficher les projections des individus sur le plan correspondant aux deux premiers axes factoriels (si vous avez le temps vous pouvez regarder ce que l'on obtient sur les autres axes).
Ajouter les anciens axes, représentés avec des flèches depuis l'origine, avec le nom des variables associées.

```
Tp <-Ts %*% E$vectors
plot.ACP=function(axe1, axe2) {
  plot(Tp[,c(axe1,axe2)], xlab=paste("Axe fact",axe1)
    , ylab=paste("Axe fact",axe2), main="ACP non normée")
  text(Tp[,axe1]+.1, Tp[,axe2]+.1 , row.names(notes), cex=.7)
  abline(h=0) ; abline(v=0)
}
plot.ACP(1,2)

ech=2
arrows(0,0,ech*E$vectors[,1],ech*E$vectors[,2])
text(ech*E$vectors[,1],ech*E$vectors[,2],colnames(Ts),cex=.8)
```

4. Calculer les corrélations entre les variables initiales et les variables correspondants aux axes factoriels (les composantes principales). Créer le graphique correspondant sans oublier le cercle unité.

```

Corr <- cor(Ts, Tp)
Corr
x11()
theta <- 2*pi*(1:100)/100
plot(cos(theta), sin(theta), type="l"
, xlim=c(-1.2,1.2),ylim=c(-1.2,1.2)
, xlab="Composante principale 1"
, ylab="Composante principale 2"
, main="Notes\ncercle des corrélations"
)
abline(h=0) ; abline(v=0)
arrows(0,0,Corr[,1],Corr[,2])
text(Corr[,1]+.1,Corr[,2]+.1,row.names(Corr),cex=.8)

```

5. Interpréter les deux premier axes factoriels : quels individus opposent-ils ? Quelles variables y contribuent positivement, négativement ?