

TP noté: ACP

Dans ce TP, nous aurons besoin de charger, en plus des options de base, les librairies factoextra et dplyr.

```
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
library(dplyr)
```

```
##
## Attachement du package : 'dplyr'
## Les objets suivants sont masqués depuis 'package:stats':
##
##     filter, lag
## Les objets suivants sont masqués depuis 'package:base':
##
##     intersect, setdiff, setequal, union
```

R Markdown

Nous travaillons sur le dataset Abalone qui regroupe 9 variables mesurées sur plus de 4000 individus. Nous ne garderons que les 8 dernières variables (la première étant catégorielle) et 100 individus pour plus de clarté.

```
Data=read.csv(file="abalone.data")
Data = Data[1:100,2:9]
head(Data)
```

```
##      X0.455 X0.365 X0.095 X0.514 X0.2245 X0.101 X0.15 X15
## 1  0.350  0.265  0.090 0.2255  0.0995 0.0485 0.070   7
## 2  0.530  0.420  0.135 0.6770  0.2565 0.1415 0.210   9
## 3  0.440  0.365  0.125 0.5160  0.2155 0.1140 0.155  10
## 4  0.330  0.255  0.080 0.2050  0.0895 0.0395 0.055   7
## 5  0.425  0.300  0.095 0.3515  0.1410 0.0775 0.120   8
## 6  0.530  0.415  0.150 0.7775  0.2370 0.1415 0.330  20
```

Abalone est le nom Anglais de l'Ormeau, un mollusque marin, les différentes variables correspondent à des mesures de différents attributs de ces coquillages.

```
Data = rename(Data, "Longueur"=`X0.455`, "Diamètre" = `X0.365`, "Hauteur" = `X0.095`, "Poids_Total"=`X0.514`, "Poids_Décortiqué" = `X0.2245`, "Poids_Viscères" = `X0.101`, "Poids_Coquille" = `X0.15`, "Sexe" = `X15`)
head(Data)
```

```
##      Longueur Diamètre Hauteur Poids_Total Poids_Décortiqué Poids_Viscères
## 1    0.350    0.265    0.090    0.2255          0.0995          0.0485
## 2    0.530    0.420    0.135    0.6770          0.2565          0.1415
## 3    0.440    0.365    0.125    0.5160          0.2155          0.1140
## 4    0.330    0.255    0.080    0.2050          0.0895          0.0395
## 5    0.425    0.300    0.095    0.3515          0.1410          0.0775
## 6    0.530    0.415    0.150    0.7775          0.2370          0.1415
```

##	Poids_Coquille	Age
## 1	0.070	7
## 2	0.210	9
## 3	0.155	10
## 4	0.055	7
## 5	0.120	8
## 6	0.330	20

1. Centrez les variables, c'est à dire retirer à chacune sa valeur moyenne. Pour cela, on calculera le vecteur des moyennes, on le répétera le nombre de fois nécessaire pour créer une matrice des moyennes de même taille que Data, contenant en position (i, j) la valeur de la moyenne de la variable j . On retirera cette moyenne à Data.
2. Réduisez de même les variables, c'est à dire, diviser les par leur écart-type. Observez les ordres de grandeurs pris maintenant par les variables centrées.
3. Stocker la matrice de variance covariance normée dans une matrice `hatSigma`.
4. Diagonalisez la matrice `hatSigma`
5. Comparer l'inertie du nuage des variables du dataset et la somme des valeurs propres de `hat(Sigma)`. Commentez le résultat.
6. Tracer, en fonction de i , le pourcentage de l'inertie expliqué par les i premières valeurs propres ainsi que celui expliqué par la i -ème valeur propre.
7. Représentez les individus sur le plan factoriel des 2 premières valeurs propres. Donnez un individus très représentatif du premier axe et un autre du second axe.
8. Représentez les variables sur le cercle des corrélations. Quelles variables expliquent selon vous le mieux les premiers axes de l'ACP? Récupérez les individus extrêmes pour ces variables et discuter leur position sur le graphe précédent. Discuter le résultat
9. En pratique, déterminer l'âge des Ormeaux demande un comptage long et fastidieux du nombre d'anneaux sur la coquille, tandis que les autres données sont plus simples à obtenir. Au vue de l'analyse précédente, discutez quelle(s) variable(s) semblent les plus pertinentes pour estimer l'âge et les limites probables d'une telle estimation.