

# TD de la 4<sup>e</sup> séance

---

1) ouvrir R et vérifier votre répertoire de travail

`getwd()`

si votre répertoire de travail n'est pas le répertoire documents/AtelierR alors changer le répertoire de travail avec la fonction `setwd()`

`setwd(dirname(file.choose()))`

2) Manipuler des fichiers de données

Ouvrir un nouveau script pour conserver vos commandes

Utiliser la fonction `rbind()` pour associer le fichier `DataTD4.csv` et le fichier `DataTD4rbind.csv`.

Attention il faudra vérifier que vous avez le même nombre de variables (fonction `dim()`) et qu'elles sont placées dans le même ordre et qu'elles ont les mêmes appellations. Vérifier aussi quelles sont du même type avec la fonction `str()`

Utiliser la fonction `merge()` pour concaténer deux tables par colonne. On se propose donc de concaténer la table créée au dessus avec les données du fichier `DataTD4merge.csv`.

Ne pas oublier de sauvegarder ce fichier en format csv `DataTD4MAJ date du jour.csv`

`write.csv2(Mydata5,file=paste("DataTD4MAJ",as.character(Sys.Date()),"%Y%m%d"),".csv")`

3) Découverte de l'outil graphique (1)

**Ouvrir un nouveau script pour garder vos commandes**

`set.seed(13)` # permet que le tirage aléatoire suivant soit toujours identique

`Echant1<-rnorm(20,mean=55,sd=10)` # construit un vecteur "a" contenant 20 points, extraits d'une population de moyenne 55 et d'écart type de 10

Analyse univariée des données – épidémiologie descriptive

`mean(Echant1)`

`sd(Echant1)`

`max(Echant1)`

`summary(Echant1)`

découverte des capacités graphiques :

`par(mfrow=c(3, 2))` #partage la fenêtre graphique en 3 lignes et 2 colonnes

`hist(Echant1)` #histogramme

`boxplot(Echant1)` #boite à moustache

`stripchart(Echant1, vertical=T)` #nuage de point

`stripchart(Echant1,pch=16,cex=1,method='jitter', vertical=T)`

`stripchart(Echant1,pch=12,cex=0.5,col=2,method='jitter', vertical=T)`

`stripchart(Echant1,pch=10,cex=1.2,col="violet",method='jitter', vertical=T)`

#### 4) Découverte de l'outil graphique (2)

```
set.seed(13)
x1<-rnorm(10,mean=100,sd=10)
set.seed(9)
x2<-rnorm(10,mean=110,sd=10)
par(mfrow=c(2,1))#partage la fenêtre graphique en 2 lignes et une colonne
boxplot(x1,x2)
t.test(x1,x2)#test T de student
plot(x1,x2)
summary(lm(x2~x1))
```

#### 5) Découverte de l'outil graphique (3)

On va utiliser le fichier créé lors de la séance 1 et complété lors de la séance2

Vérifier que Mydata a bien été créé avec la fonction ls()

Vérifier que l'objet créé contient bien l'information ?

```
names(Mydata)
head(Mydata)
summary(Mydata)#permet de relever quelles variables sont à factoriser et
lesquelles sont à garder en format numérique
```

Créer un vecteur contenant les numéros des colonnes à transformer

```
Transvar<- c(x1,x2,...,xn) # à vous de définir ces x
```

Programmer une boucle qui va transformer chacune des variables choisies en variable qualitative

```
for (i in 1:X) { Mydata[, Transvar[i]] <- factor(Mydata[, Transvar[i]))# à vous de définir X
```

6) Créer une table de données qui ne contienne que les données pour les classe d'âge sont : (20,22] et (25,28.7]

```
Mydata.1<-subset(Mydata, CLASSAGE=="(18,22]"|CLASSAGE=="(25,28.7]"
```

7) diagramme en bâton données catégorielles

```
plot(Mydata$CLASSAGE)
plot(Mydata$CLASSAGE,col=rainbow(3))
plot(Mydata$CLASSAGE,col=rainbow(3), xlab="Classe d'âge", ylab="Effectifs", cex.lab=1.5, cex.axis=0.5)
plot(Mydata$CLASSAGE,col=rainbow(3), xlab="Classe d'âge", ylab="Effectifs",cex.lab=1.5,cex.axis=0.5,
xaxt="n")
axis(side=1,labels=paste(1:4,"°GROUPE"), tck=0, at=plot(Mydata$CLASSAGE,plot=F) )
```

8) camembert

```
pie(summary(Mydata$SEXE)) #ou pie(table(Mydata$SEXE)
nom<-c("Hommes", "Femmes")
```

```
titre<-"répartition par sexe"
pie(table(Mydata$SEXE),label=nom, main = titre,cex.main=2, col=c("red", "blue"))
```

la fonction palette() permet de connaître les numéros des couleurs.

```
[1] "black" "red" "green3" "blue" "cyan" "magenta" "yellow" "gray"
```

11) données quantitatives

```
plot(Mydata$AGE)#nuage de point
```

```
hist(Mydata$AGE)#histogramme
```

```
hist(Mydata$AGE,nclass=3)#on choisit le nombre de classe
```

```
hist(Mydata$AGE,nclass=3,plot=F)#pour avoir les données du graphiques
```

```
hist(Mydata$AGE,nclass=3,plot=F)$breaks; hist(Mydata$AGE,nclass=3,plot=F)$counts
```

```
hist(Mydata$AGE,freq=T,breaks=c(0:70))
```

9) boites à moustache simple

```
boxplot(Mydata$AGE)
```

```
boxplot(Mydata$POIDS)
```

```
boxplot(Mydata$TAILLE)
```

```
boxplot(Mydata$IMC)
```

10) boites à moustache par catégorie

```
boxplot(split(Mydata$AGE,Mydata$SEXE))
```

```
boxplot(Mydata$AGE~Mydata$SEXE)
```

```
boxplot(Mydata$POIDS~Mydata$SURPOIDS)
```

```
boxplot(Mydata$IMC~Mydata$SEXE)
```

```
boxplot(Mydata$IMC~Mydata$CLASSAGE)
```

11) ajouter un titre

```
boxplot(Mydata$AGE~Mydata$SEXE, main="Moyenne d'âge par sexe")
```

```
#ajouter un nom à l'axe des abscisses
```

```
boxplot(Mydata$AGE~Mydata$SEXE, main="Moyenne d'âge par sexe", xlab="SEXE")
```

12) partager la fenêtre graphique pour présenter plusieurs graphiques

```
par(mfrow= c(3,2))
```

```
boxplot(Mydata$AGE)
```

```
boxplot(Mydata$POIDS)
```

```
boxplot(Mydata$AGE~Mydata$SEXE, main="Age moyen par sexe",xlab="SEXE")
```

```
pie(summary(as.factor(Mydata$MALADIE)),label=c("Male","Female"), main = "Répartition par sexe")
```

```
plot(Mydata$CLASSAGE)
```

```
plot(Mydata$SEXE)
```

```
#supprimer le partage
```

```
par(mfrow=c(1,1))
```

13) autre méthode fonction layout() pour avoir des fenêtres de taille différente

```
zones<-matrix(c(1,1,1,2,3,4),ncol=2)
```

```
layout(zones)
```

```
layout.show(max(zones)) #affiche les régions de figure
```

```
boxplot(Mydata$AGE~Mydata$SEXE, main="Age moyen par sexe",xlab="SEXE")
```

```
pie(summary(as.factor(Mydata$MALADIE)),label=c("non malade","malade"), main = "Repartition en fonction du statut en lien avec la maladie")
```

```
plot(Mydata$CLASSAGE, main ="répartition par age",col="blue")  
plot(Mydata$SEXE, main ="répartition par sexe", col ="red")
```

#### 14) Exporter un graphique

##### #Exemple1

```
graphics.off()  
pdf(file="exempleA.pdf")  
hist(Mydata$AGE,nclass=5, main ="répartition par âge",col="blue")  
plot(Mydata$SEXE, main ="répartition par sexe", col ="red")  
dev.off()
```

##### #Exemple2

```
graphics.off()  
pdf(file="exempleB%03d.pdf",onfile=FALSE)  
hist(Mydata$AGE,nclass=5, main ="répartition par âge",col="blue")  
plot(Mydata$SEXE, main ="répartition par sexe", col ="red")  
dev.off()
```

#### 15) Exercice pour la découverte de l'outil graphique et des tests statistiques

Utiliser ici le fichier DataTD4\_2°Partie.

```
Mydata2<-read.table(file.choose(), header=T, sep=";", dec=",")  
by(Mydata2$AGE,Mydata2$SEXE,mean)  
par(mfrow=c(1,1))  
Titre<-"Age moyen en fonction du sexe"  
boxplot(Mydata2$AGE~Mydata2$SEXE,col=c(2,4), main=Titre, ylab="Age en année", xlab="Sexe")  
jpeg(file.choose())  
boxplot(Mydata2$AGE~Mydata2$SEXE,col=c(2,4), main=Titre, ylab="Age en année", xlab="Sexe")  
dev.off() # ferme la fenêtre graphique active  
wilcox.test(Mydata2$AGE[Mydata2$SEXE==2],Mydata2$AGE[Mydata2$SEXE==1],paired=FALSE)  
wilcox.test(Mydata2$AGE~Mydata2$SEXE,paired=FALSE)
```