
CORRECTION TP 1.2 - Analyse de données quantitatives avec le logiciel R

Exemple 1.1

```
1. > arbres<-c(1,2,4,1,6,3,2,1,2,0,1,2,2,1,3,0,3,2,1,2,2,3,2,3)
2. > sort(arbres)
3. > n<-length(arbres)
   > n
```

2.1 Un exemple

```
1. > souris<-c(74,85,95,84,68,93,84,87,78,72,81,91,80,65,76,81,97,69,70,98)
   > souris
2. > table(souris)
3. > length(souris)
4. > summary(souris)
```

Exercice 1 Le fichier Forbes2000

```
1. > Forbes2000
2. > str(Forbes2000)
3. > class(Forbes2000)
4. > nrow(Forbes2000)
5. > ncol(Forbes2000)
6. > names(Forbes2000)
7. > length(Forbes2000$rank)
8. > Forbes2000$name[1]
9. > length(levels(Forbes2000$category))
10. > levels(Forbes2000$category)
11. > summary(Forbes2000$category)
12. > class(Forbes2000$sales)
13. > summary(Forbes2000$sales)
14. > summary(Forbes2000)
15. La première ligne divise la fenêtre graphique en deux sous fenêtres, situées l'une au dessus de l'autre.
    La deuxième ligne crée l'histogramme associé à la variable « marketvalue ».
    La troisième ligne crée l'histogramme associé à la variable « log(marketvalue) ».
16. Si la première ligne est absente, le second histogramme efface le premier.
17. > plot(log(Forbes2000$marketvalue),log(Forbes2000$sales))
18. > boxplot(log(Forbes2000$marketvalue)~ Forbes2000$country=="Germany") par exemple
```

Exercice 2

On remplace l'exercice sur l'indice de Quételet par un exercice plus complet dont voici l'énoncé et le corrigé.

Partie I.

1. Importer sous R le fichier de données mis à disposition en utilisant la fonction adéquate.

```
> Quetelet<-read.csv(file.choose())
```

2. Afficher, une par une, les variables *taille*, *poids* et *sexe*. Créer 3 vecteurs *taille*, *poids* et *sexe* correspondant aux données de la table.

```
> Quetelet$taille
> Quetelet$poids
> Quetelet$sexe

> taille<-Quetelet$taille
> poids<-Quetelet$poids
> sexe<-Quetelet$sexe
```

3. Quel est le nombre d'individus dans l'échantillon ?

```
> length(Quetelet$taille)
```

4. Afficher le tableau des indicateurs de tendance centrale pour les deux variables quantitatives. Essayer également cette commande en donnant le nom de la table et non le nom d'une variable. Qu'obtient-on ?

```
> summary(taille)
> summary(poids)
> summary(Quetelet$taille)
> summary(Quetelet$poids)
```

5. Décomposer l'échantillon en un échantillon d'hommes et un échantillon de femmes.

```
> genre<-split(Quetelet,sexe)
```

6. Quel est le nombre d'hommes et le nombre de femmes dans l'échantillon ?

```
> nrow(genre$f)
> nrow(genre$h)
```

Partie II.

La masse corporelle d'un individu est mesurée à partir du poids et de la taille. L'indice le plus utilisé est celui proposé par Adolphe Quételet (1796-1874), astronome et mathématicien belge.

7. Construire la variable « indice de Quételet » à partir des données

$$\text{imc} = \frac{\text{poids}}{\text{taille}^2}$$

La taille est exprimée en mètre et le poids en kilogrammes. Cet indice s'appelle aujourd'hui « indice de masse corporelle » (*imc*) ou « body mass index » (*bmi*). Il permet de mesurer la corpulence de l'homme adulte. L'Organisation Mondiale de la Santé a défini les critères suivants : maigreur (inférieur à 18,5), normal (de 18,5 à 25), risque de surpoids (de 25 à 30), obésité (supérieur à 30). Mais l'indice de Quételet n'a qu'une valeur indicative. Pour déterminer l'existence d'une obésité réelle, il faut faire d'autres mesures destinées à établir exactement la proportion de masse grasse, car c'est l'excès de masse grasse qui représente un facteur de risque.

```
> imc<-poids/(taille/100)2
```

8. Calculer les paramètres statistiques élémentaires de cette nouvelle variable sur l'ensemble des individus et en fonction de chaque sexe.

```
> summary(imc)
> genreimc<-split(imc,sexe)
> lapply(genreimc,summary)
```

9. Construire l'histogramme de cette nouvelle variable sur l'ensemble des individus.

```
> hist(imc, col = grey(0.8))
```

10. Comparer graphiquement les indices de masse corporelle chez les hommes et chez les femmes à l'aide d'histogrammes et de fonctions de répartitions empiriques (utiliser la fonction « `ecdf` » de la librairie « `stepfun` »).

```
> par(mfcol = c(2, 2))
> hist(imc[sexe == "h"], main = "Hommes", xlab = "imc", xlim = c(16,27), col = grey(0.9))
> hist(imc[sexe == "f"], main = "Femmes", xlab = "imc", xlim = c(16,27), col = grey(0.9))
> plot(ecdf(imc[sexe == "h"]), main = "")
> plot(ecdf(imc[sexe == "f"]), main = "")
```

Si on souhaite un passage à la modélisation :

```
> par(mfcol = c(2, 2))
> hist(imch <- imc[sexe == "h"], main = "Hommes", xlab = "imc", xlim = c(16,27),
+ col = grey(0.9), prob = T)
> lines(x0 <- seq(16, 28, le = 50), dnorm(x0, mean(imch), sd(imch)),lwd = 2, col = "red")
> hist(imcf <- imc[sexe == "f"], main = "Femmes", xlab = "imc", xlim = c(16,27),
+ col = grey(0.9), prob = T)
> lines(x1 <- seq(16, 26, le = 50), dnorm(x1, mean(imcf), sd(imcf)),lwd = 2, col = "red")
> plot(ecdf(imch), main = "")
> lines(x0, pnorm(x0, mean(imch), sd(imch)), lwd = 2, col = "red")
> plot(ecdf(imc[sexe == "f"]), main = "")
> lines(x1, pnorm(x1, mean(imcf), sd(imcf)), lwd = 2, col = "red")
```

11. En utilisant la fonction « `boxplot` », représenter l'indice de Quételet en fonction du sexe.

```
> boxplot(imc ~ sexe, col = grey(0.8))
```

12. Construire sur un seul graphique les nuages de points de l'indice de Quételet en fonction de la taille, en fonction du poids, pour chaque sexe. Commenter.

```
> par(mfcol = c(2, 2))
> plot(imc[sexe == "h"], tai[sexe == "h"], xlim = c(min(imc), max(imc)),
+ ylim = c(min(tai), max(tai)), pch = 20, xlab = "imc", ylab = "taille",main = "hommes")
> plot(imc[sexe == "h"], poi[sexe == "h"], xlim = c(min(imc), max(imc)),
+ ylim = c(min(poi), max(poi)), pch = 20, xlab = "imc", ylab = "poids")
> plot(imc[sexe == "f"], tai[sexe == "f"], xlim = c(min(imc), max(imc)),
+ ylim = c(min(tai), max(tai)), pch = 20, xlab = "imc", ylab = "taille", main = "femmes")
> plot(imc[sexe == "f"], poi[sexe == "f"], xlim = c(min(imc), max(imc)),
+ ylim = c(min(poi), max(poi)), pch = 20, xlab = "imc", ylab = "poids")
```

13. Représenter, sur un seul graphique, le nuage de points de l'indice de Quételet en spécifiant hommes et femmes. Placer les bornes des critères de l'O.M.S. Commenter.

```
> plot(imc, pch = 19, cex = 2, col = c("deeppink3", "blue")[unclass(sexe)])
> abline(h = 18.5)
> abline(h = 25)
> title("indice de masse corporelle")
> llo = c("femmes", "hommes")
> legend(1, 24.8, llo, pch = 21, pt.cex = 2, cex = 1.25, col = c("deeppink3","blue"),
+ pt.bg = c("deeppink3", "blue"))
```