

# Utilisation du logiciel R

## Exercices : corrigés

### 1. Utilisation du logiciel R comme une simple calculatrice

#### Exercice n°1 :

a). Selon votre poids (en kilogramme) et votre taille (en mètre) :

```
> 70 / 1.84^2  
[1] 20.67580
```

b).

```
> monPoids <- 70  
> maTaille <- 1.84  
> monPoids / maTaille^2  
[1] 20.67580
```

#### Exercice n°2 :

```
> monBudget <- 300
```

a).

```
> monTotal <- 260 + 30 + 60  
> monTotal  
[1] 350  
> monTotal < monBudget  
[1] FALSE
```

Réalisation de l'achat : non, car le total est supérieur à mon budget.

b).

```
> 260 + 30  
[1] 290  
> (260 + 30) < monBudget  
[1] TRUE
```

Réalisation de l'achat : oui, car le total de ces 2 articles est inférieur à mon budget.

c).

```
> maReduction <- (260*30/100) + (30*30/100) + (60*30/100)  
> maReduction  
[1] 105  
> maDepense <- monTotal - maReduction  
> maDepense  
[1] 245
```

Le montant de la réduction est de 105 euros.

Le coût total en période de soldes sera alors de 245 euros, ce qui est plus intéressant que l'offre faite ce-jour.

d).

```
> monBudget - maDepense  
[1] 55  
> monBudget <- monBudget - maDepense  
> monBudget  
[1] 55
```

Après l'achat il me reste 55 euros.

### Exercice n°3 :

a).

```
> ccont <- c(11, 13, 15.5, 12, 8, 9, 13, 16)
> exam <- c(8.5, 14, 15, 10, 12, 13, 14, 17)
> ccont
[1] 11.0 13.0 15.5 12.0 8.0 9.0 13.0 16.0
> exam
[1] 8.5 14.0 15.0 10.0 12.0 13.0 14.0 17.0
```

b). Nous allons continuer à travailler sur les vecteurs de données. Il faut pondérer les notes au contrôle continu par 0.4 (compte pour 40 % de la note finale) et celles de l'examen par 0.6 (compte pour 60 % de la note finale).

```
> ue <- 0.4*ccont + 0.6*exam
> ue
[1] 9.5 13.6 15.2 10.8 10.4 11.4 13.6 16.6
```

c). On accède à un élément d'un vecteur en indiquant entre crochets le numéro de l'élément voulu :

```
> ccont[6]
[1] 9
> exam[6]
[1] 13
> ue[6]
[1] 11.4
```

d). Moyenne au contrôle continu (soit calculée comme la somme des notes divisées par le nombre de notes, soit directement par la fonction mean) :

```
> sum(ccont) / 8
[1] 12.1875
> mean(ccont)
[1] 12.1875
```

Moyenne à l'examen :

```
> sum(exam) / 8
[1] 12.9375
> mean(exam)
[1] 12.9375
```

Moyenne à l'UE :

```
> sum(ue) / 8
[1] 12.6375
> mean(ue)
[1] 12.6375
```

Note la plus élevée :

```
> max(ue)
[1] 16.6
```

Note la plus basse :

```
> min(ue)
[1] 9.5
```

## 2. Utilisation du logiciel R pour une analyse descriptive

### Exercice n°4 :

a). Variables quantitatives : id, age, nbmedic.

Variables qualitatives : agecl, sexe, sitfamiliale, mode2vie, aidedomicile, chute, CV, Psy, Antidiabetique, autresmed, marcheautonome.

Remarque : aidedomicile, CV, Psy, Antidiabetique, autresmed sont des variables numériques qui prennent les valeurs 0 ou 1 qui correspondent aux valeurs qualitatives « Non » et « Oui ».

Après avoir transformer les fichier « .xls » au format « .csv » il faut l'importer dans un répertoire de R à l'aide de la commande suivante :

```
> chuteurs <- read.csv2(" ... chuteurs.csv", header=TRUE)
```

```
> str(chuteurs)
'data.frame': 153 obs. of 14 variables:
 $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ age     : int  67 82 87 66 66 81 78 75 78 75 ...
 $ agecl   : chr  "65-74" "75 et +" "75 et +" "65-74" ...
 $ sexe    : chr  "Homme" "Femme" "Femme" "Femme" ...
 $ sitfamiliale : chr  "Mari\`e" "Veuf" "Veuf" "Mari\`e" ...
 $ mode2vie : chr  "Pas Seul" "Pas Seul" "Seul" "Pas Seul"
 ...
 $ aidedomicile : int  0 0 0 0 0 0 0 0 0 0 ...
 $ chute       : chr  "Non" "Oui" "Oui" "Non" ...
 $ nbmedic     : int  8 4 6 6 6 6 8 6 7 2 ...
 $ CV          : int  1 0 1 1 1 1 1 0 1 0 ...
 $ Psy         : int  0 1 0 1 1 1 0 1 1 1 ...
 $ Antidiabetique: int  1 1 1 0 0 0 1 0 0 1 ...
 $ autresmed   : int  1 0 1 0 1 0 1 0 0 0 ...
 $ marcheautonome: chr  "Oui" "Oui" "Aide" "Oui" ...
```

### Recodage des variables :

```
# Recodage des variables qui sont au format character
# en variables catégorielles
chuteurs$agecl <- as.factor(chuteurs$agecl)
chuteurs$sexe <- as.factor(chuteurs$sexe)
chuteurs$sitfamiliale <- as.factor(chuteurs$sitfamiliale)
chuteurs$mode2vie <- as.factor(chuteurs$mode2vie)
chuteurs$chute <- as.factor(chuteurs$chute)
chuteurs$marcheautonome <- as.factor(chuteurs$marcheautonome)

# Recodage des variables qui sont au format numérique
# en variables catégorielles
# sauf nbmedic qui est continue
chuteurs$aidedomicile <- factor(chuteurs$aidedomicile,
  labels=c("Non", "Oui"), levels=c(0, 1))
chuteurs$marcheautonome <- factor(chuteurs$marcheautonome,
  levels=c("Oui", "Aide", "Non"),
  labels=c("Oui", "Aide", "Non"))
chuteurs$CV <- factor(chuteurs$CV,
  levels=c(0, 1),
  labels=c("Non", "Oui"))
```

```

chuteurs$Psy <- factor(chuteurs$Psy,
  levels=c(0, 1),
  labels=c("Non", "Oui"))
chuteurs$Antidiabetique <- factor(chuteurs$Antidiabetique,
  levels=c(0, 1),
  labels=c("Non", "Oui"))
chuteurs$autresmed <- factor(chuteurs$autresmed,
  levels=c(0, 1),
  labels=c("Non", "Oui"))

```

b). Le data.frame peut être attaché dans l'environnement de travail à l'aide de la commande suivante :

```
> attach(chuteurs)
```

Fréquences absolues : 2 commandes sont possibles

```
> summary(sexe)
```

```
Femme Homme
   91    62
```

```
> table(sexe)
```

```
sexe
Femme Homme
   91    62
```

Fréquences relatives : fréquences absolues divisées par l'effectif total, qui est obtenu en déterminant le nombre d'éléments (commande length) du vecteur sexe

```
> summary(sexe)/length(sexe)
```

```
   Femme   Homme
0.5947712 0.4052288
```

```
> table(sexe)/length(sexe)
```

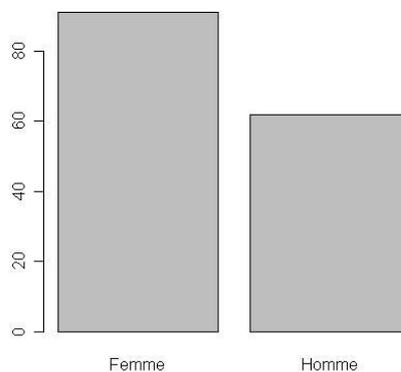
```
sexe
   Femme   Homme
0.5947712 0.4052288
```

```
> prop.table(table(sexe))
```

```
sex
   Man   Woman
0.4052288 0.5947712
```

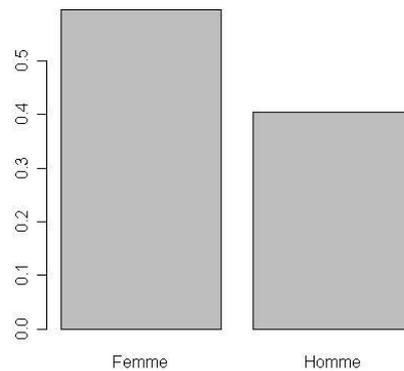
Histogramme des fréquences absolues :

```
> barplot(summary(sexe))
```



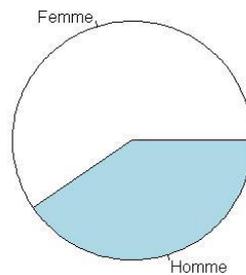
Histogramme des fréquences relatives :

```
> barplot(summary(sexe)/length(sexe))
```



Représentation en camembert :

```
> pie(summary(sexe))
```



c).

```
> class(age)
[1] "integer"
```

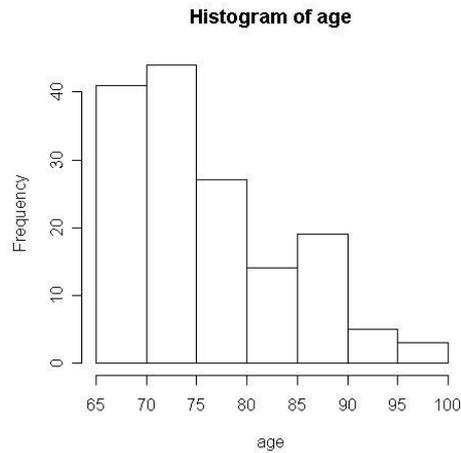
Il s'agit d'une variable quantitative définie comme étant un entier (`integer`) dans R. Elle peut être décrite par : minimum, maximum, médiane, moyenne, 1<sup>er</sup> et 2<sup>ème</sup> quartile

```
> summary(age)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	70.00	75.00	76.29	81.00	100.00

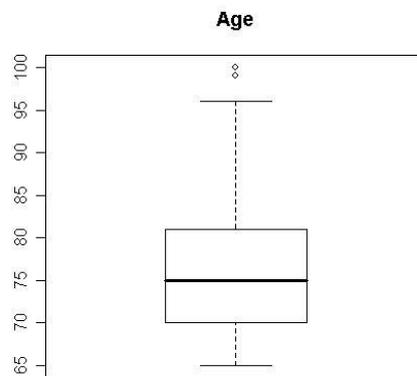
Histogramme des fréquences :

```
> hist(age)
```



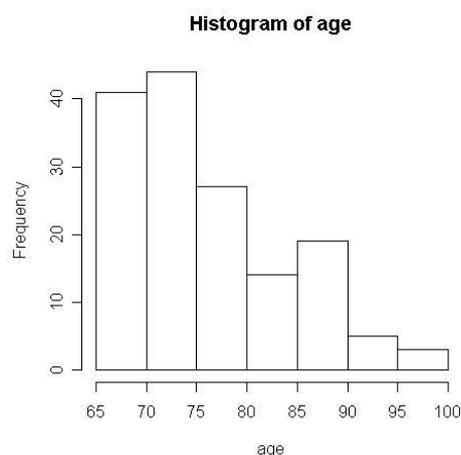
Représentation sous la forme d'une boîte à moustache :

```
> boxplot(age, main = "Age")
```



d). Utilisez `?hist` pour comprendre le fonctionnement de cette commande. Il faut utiliser l'argument `breaks` pour construire ses propres intervalles.

```
> hist(age, breaks=c(65, 70, 75, 80, 85, 90, 95, 100))
```



### Exercice n°5 :

a). Il s'agit d'une variable quantitative que l'on veut décrire en fonction des niveaux d'une variable qualitative.

La commande `summary` passée dans `by` permet d'avoir les résumés statistiques concernant l'âge en fonction du niveau de la variable `chute`.

```
> by(age, chute, summary)
```

```
INDICES: Non
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	69.00	74.00	75.58	80.50	100.00

```
-----  
INDICES: Oui
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
66.00	75.00	77.00	78.76	84.75	96.00

Autre possibilité : construire les vecteurs d'âge en fonction du niveau de la variable `chute`, puis utiliser la commande `summary` :

```
> ageNon <- age[chute=="Non"]
```

```
> summary(ageNon)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
65.00	69.00	74.00	75.58	80.50	100.00

```
> ageOui <- age[chute=="Oui"]
```

```
> summary(ageOui)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
66.00	75.00	77.00	78.76	84.75	96.00

b).

```
> ageOui <- age[chute=="Oui"]
```

```
> ageNon <- age[chute=="Non"]
```

```
> old.par <- par(no.readonly = TRUE)
```

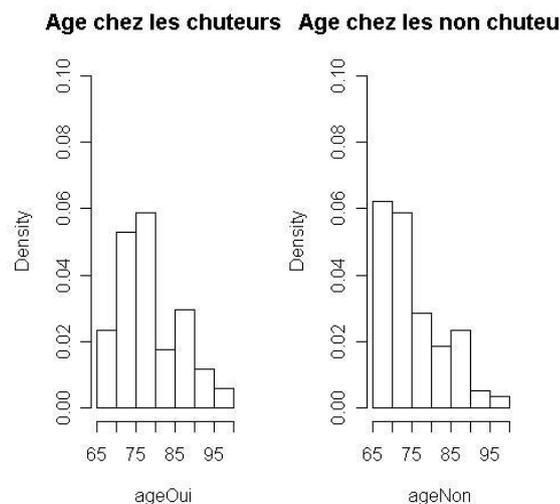
```
> par(mfrow = c(1, 2))
```

```
> hist(ageOui, breaks = seq(65, 100, 5), freq = FALSE,  
+      ylim = c(0, 0.1), main = "Age chez les chuteurs")
```

```
> hist(ageNon, breaks = seq(65, 100, 5), freq = FALSE,  
+      ylim = c(0, 0.1), main = "Age chez les non chuteurs")
```

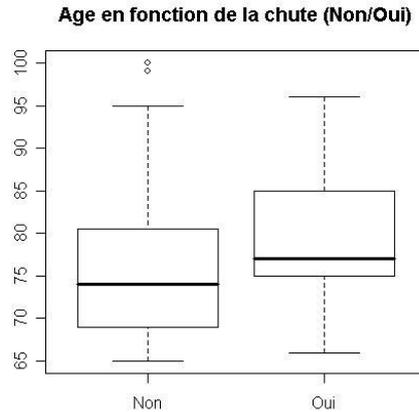
```
> par(mfrow = c(1, 1))
```

```
> par(old.par)
```



c).

```
> boxplot(age ~ chute, main = "Age en fonction de la chute  
(Non/Oui) ")
```



**Exercice n°6 :**

a). Il s'agit de 2 variables qualitatives dont l'association est décrite par un tableau de contingence :

```
> table(mode2vie, chute)
      chute
mode2vie Non Oui
Pas Seul  93  25
Seul      26   9
```

b). Il faut dans un premier temps créer les vecteurs chute chez les personnes ne vivant pas seule et chez celles vivant seule :

```
> chutePasSeul <- chute[mode2vie=="Pas Seul"]
> chuteSeul <- chute[mode2vie=="Seul"]
```

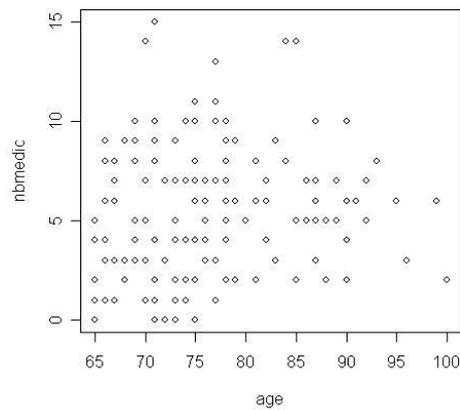
Pour dans un deuxième temps, faire les tableaux des fréquences relatives de chute chez les personnes ne vivant pas seul puis chez celles vivants seule :

```
> table(chutePasSeul)/length(chutePasSeul)
chutePasSeul
      Non      Oui
0.7881356 0.2118644
> table(chuteSeul)/length(chuteSeul)
chuteSeul
      Non      Oui
0.7428571 0.2571429
```

**Exercice n°7 :**

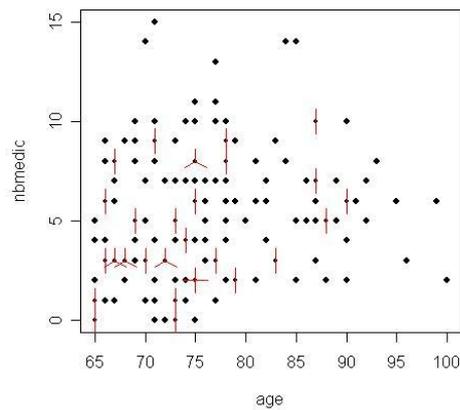
a). Il s'agit de 2 variables quantitatives. La représentation graphique est un nuage de points :

```
> plot(age, nbmedic)
```



Une autre fonction peut être utilisée : `sunflowerplot` qui donne un graphique similaire à celui obtenu avec la commande `plot` mais les points superposés sont dessinés en forme de fleurs dont le nombre de pétale représente le nombre de points.

```
> sunflowerplot(age, nbmedic)
```



b). Il s'agit de 2 variables quantitatives. Une statistique possible est le coefficient de corrélation :

```
> cor(age, nbmedic, method=c("pearson"))
> cor(age, nbmedic, method=c("spearman"))
```