

## TD 1

# 1 Logiciel R

## 1.1 Installation et découverte du logiciel

Le logiciel libre R est disponible pour les plateformes Linux, Windows et Mac. Il est téléchargeable gratuitement à l'adresse suivante :

`http://cran.r-project.org/`

La procédure d'installation diffère évidemment que l'on soit sous Linux, Windows ou Mac.

## 1.2 Découverte de R

Pour lancer le logiciel, il suffit de cliquer sur l'icone du logiciel (Windows et Mac) ou de taper R dans une console (Linux). La fenêtre du logiciel apparaît et nous pouvons commencer à travailler. Pour découvrir les possibilités, notamment graphiques, de R, vous pouvez appeler les différentes démonstrations proposées par R. Pour obtenir la liste des démonstrations, il suffit de taper :

```
demo()
```

La démonstration des outils graphiques est quant à elle accessible en tapant :

```
demo(graphics)
```

La *Reference Card* vous indiquera les éléments clés du langage R et en particulier la façon d'affecter une valeur à une variable :

```
x <- 2
```

Pour afficher l'aide d'une fonction, il faut utiliser l'une des commandes suivantes :

```
?nom_de_la_fonction  
help.start()
```

Vous pouvez également consulter l'introduction à R qui est disponible à l'adresse suivante :

`http://cran.r-project.org/doc/contrib/Paradis-rdebuts\_fr.pdf`

## 1.3 Installation de packages supplémentaires

Dans la suite de cette séance, nous aurons peut-être besoin de packages supplémentaires. Ceux-ci pourront être installés directement depuis le logiciel en tapant :

```
install.packages('nom_du_package')
```

## 2 Premiers pas avec R

### 2.1 Un exemple simple

Commençons avec un exemple simple :

```
weight <- c(60, 72, 57, 90, 95, 72)
height <- c(1.75, 1.80, 1.65, 1.90, 1.74, 1.91)
```

et calculez l'indice de masse corporelle (Body Mass Index) :

$$bmi = \frac{weight}{height^2},$$

puis affichez le résultat :

```
bmi
```

Calculez ensuite la moyenne  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  et la variance  $\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$  de ces données.

### 2.2 Les vecteurs

Dans l'exemple précédent, nous avons déjà créé des vecteurs grâce à la commande `c()`. Tapez les commandes suivantes et observez leur action :

```
c(1,2,3,4,5)
seq(4,9)
seq(4,10,2)
4:9
rep(c(0,1,2),3)
rep(c(0,1,2),1:3)
```

### 2.3 Les matrices

Les matrices en R peuvent être créées grâce à la commande `matrix()`. Tapez les commandes suivantes et observez leur action :

```
matrix(1:12,nrow=3,ncol=4)
matrix(1:12,nrow=3,byrow=TRUE)
```

Observez également l'action des commandes ci-dessous :

```
cbind(1:3,4:6,7:9)
rbind(1:3,4:6,7:9)
```

Il est donc aussi possible de créer des matrices en concaténant des vecteurs. C'est en fait la façon la plus simple de créer des matrices.

### 2.4 Les data-frames

Il existe en R un objet particulièrement adaptée aux données statistiques : la data-frame. Les data-frames permettent de combiner des objets "homogènes" dans une même structure, par exemple, des caractéristiques d'une même série d'individus. On peut construire une telle structure grâce à la commande `data.frame()`. Tapez les commandes suivantes et observez leur action :

```
weight <- c(60, 72, 57, 90, 95, 72)
height <- c(1.75, 1.80, 1.65, 1.90, 1.74, 1.91)
X <- data.frame(weight,height)
X
```

L'accès aux colonnes et aux lignes peut se faire de différentes manières :

```
X$weight  
X[,1]  
X[1,]
```

Trouvez à présent les commandes nécessaires pour afficher :

1. la deuxième colonne de X,
2. la 5ème observation de X,
3. la valeur de la 2ème colonne de la 3ème observation de X,
4. la valeur de la 2ème colonne des 3ème et 4ème observations de X.

## 2.5 Sélection conditionnelle

Il est souvent utile d'avoir accès à une partie seulement des données et pour cela la sélection conditionnelle est très utile. Tapez les commandes suivantes, observez leur action et tentez de les interpréter :

```
weight[height<=1.80]  
height[height<=1.80]  
sum(weight[height<=1.80] / length(weight[height<=1.80]))  
sum(height[height<=1.80] / length(height[height<=1.80]))
```

Trouvez à présent les commandes nécessaires pour :

1. calculer le nombre de personnes de l'échantillon dont la taille est inférieure à 1.75,
2. calculer la proportion de personnes de l'échantillon dont le poids est supérieur à 70,
3. calculer le nombre de personnes de l'échantillon dont la taille est inférieure à 1.75 et dont le poids est supérieur ou égal à 60.

## 2.6 Les graphiques

Les capacités graphiques de R sont très grandes. Tapez les commandes suivantes pour vous familiariser avec les graphiques sous R :

```
plot(height,weight)  
plot(height,weight,col='red',pch=2)  
hist(height)  
hist(weight,col='green')  
bmi <- weight / height^2  
prop <- c(sum(bmi<20)/6,sum(bmi>=20&bmi<25)/6,sum(bmi>=25)/6)  
names(prop) <- c("Anorexique","Normal","Obèse")  
pie(prop)
```

Trouvez à présent les commandes nécessaires pour :

1. tracer l'indice de masse corporelle en fonction de la taille pour l'échantillon donné,
2. tracer l'histogramme des indices de masse corporelle pour l'échantillon.

## 2.7 Lecture / écriture de données

Utilisez la commande `write.table()` pour sauver la data.frame X créée précédemment dans un fichier texte. Utilisez ensuite la commande `read.table()` pour recharger ces données dans R.