1 Introduction to R

You must make these exercises with R.
We recommend to use R-Studio (an interactive R development environment).
Please visit the class website.

Exercise 1.1
A R code is given hereafter :

```
1  # Data
2  data <- c(23,24,24,25,24,24,24,25,26,22,21)
3  # Center/location
4  mean(data)
5  median(data)
6  # Dispersion
7  sd(data)
8  var(data)
9  IQR(data)
10 # Rank
11 min(data)
12 max(data)
13 quantile(data,0.25)
```

1. Save the code in a file "exo1.R" and run it with the command "source("exo1.R")". You can modify the working directory with the command "setwd("*myWorkingDirectory*")" where *myWorkingDirectory* is the path of your working directory.

2. Explain each line of the code

Exercise 1.2
A R code is given hereafter :

```
1  n = 30
2  k = 1000
3  mu = 5; sigma = 2; sdclt = sigma/sqrt(n)
4  x = matrix(rnorm(n*k,mu,sigma),n,k)
5  x.mean = apply(x,2,mean)
6  x.down = mu - 4*sdclt; x.up = mu + 4*sdclt; y.up = 1.5
7  hist(x.mean,prob=T,xlim=c(x.down,x.up),ylim=c(0,y.up),main='Sampling
8  distribution of the sample mean, Normal case')
9
10 par(new=T)
11 x = seq(x.down,x.up,0.01)
12 y = dnorm(x,mu,sdclt)
13 plot(x,y,type='l',xlim=c(x.down,x.up),ylim=c(0,y.up))
```

1. Save the code in a file "exo2.R" and run it.

2. Explain each line of the code.

3. What does this code make ?

4. What can you conclude if n increases ?

Exercise 1.3

To install the package "RColorBrewer" : install.packages("RColorBrewer",dependencies=TRUE)
A R code is given hereafter :

```
1  data(iris)
2  dim(iris)
3  class(iris)
4  summary(iris)
5  head(iris)
6  sapply(iris, class)
7  plot(iris$Sepal.Length)
8  plot(iris$Sepal.Length, iris$Sepal.Width)
9  library(RColorBrewer)
10 display.brewer.all(n=3)
11 plot(Sepal.Width ~ Sepal.Length, data=iris, col=brewer.pal(3, "Set2")[iris$
       Species])
12 legend(x=6.5, y=4.5, legend=levels(iris$Species), col=brewer.pal(3, "Set2"),
       pch=1)
13 x <- hist(iris$Sepal.Length)
14 segments(x0=x$mids-0.25, x1=x$mids+0.25, y0=x$counts, y1=x$counts, lw=4, col="
       red")
```

1. Save the code in a file "exo3.R" and run it.

2. Explain each line of the code. What is "iris" ?

Exercise 1.4

Download the file "2008.csv" from the Jalon website.
This file comes from http ://stat-computing.org/dataexpo/2009/the-data.html
Load the dataset in R under the name "flights" :

```
1  flights <- read.csv('2008.csv')
```

The following commands can be useful to solve this exercise : str, summary, dim, nrow, ncol, names, is.na, and match.
Please answer the following questions.

1. Print the structure of the data. What do you think about it ?

2. Print the summary statistics of the data. What do you think about the values ? (format, consistency, completeness)

3. Print the dimensionality of the data (number of rows and columns)

4. Print the number of rows. This may seem like a silly command, but it is quite useful for loops and if statements.

5. Print the number of columns.

6. Print the names of the variables.

7. Print whether the first column has missing values (NAs). Try to answer this question with two ways. Hint : %in%

8. Print the number of variables that contain missing values.

9. Find the portion of the variables that contain missing values. What do you think about it ?

10. Print the names of the variables that contain missing values.