PW01

Data analysis with R

Exercise 1 :

In this exercise we will do some exploratory data analysis on the famous Iris dataset.

The Iris Dataset contains four features (length and width of sepals and petals) of 50 samples of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). These measures were used to create a linear discriminant model to classify the species.

There is no need to import a dataset file. The datasets library in R already contains it.

1) Load the dataset
1) Return the first 5 rows of the dataset
2) Give the dimensions of the dataset
3) Select the rows where Petal Length <1,5
4) Select the subset containing setosa OU versicolor
5) Return the number of iris where Petal Length = 3.5cm
6) Select the number of flowers where Petal Length <1.5 cm or > 5cm
7) Extract the first and the third row of the dataset
8) Plot Sepal Lenght against Petal Lenght. Add the linear regression line
9) For each species, give the number of rows.
10) Use the previous result to creat a pie chart plot, then a barplot
11) Create a boxplot using the numeric variables of iris.
12) Add a title to the plot and remove the outliers
13) Put the the pie chart and the boxplot
14) Represent the pie chart and the boxplot on the same graphic window, one next to the other, using the par() function and the mfrow option
15) Export the graph obtained from the previous question (the pie chart and the boxplot) to pdf on your machine.

Exercise 2 :

1) Import into a variable named A the data contained in the file named auto2004_original.txt.

Source : https://github.com/PF-BB/Formation-Rrrr/tree/master/docs/source/TP1/data

2) Display the names of the considered variables

3) What is the mode of the objects created by the read.table() function?
4) Show number of rows and columns.
5) Display the first 6 rows of this dataset
6) **Display** a **summary statistic** of the datase
7) Determine th variance and standard deviation of the Puissance variable.
8) Importez dans une variable nommée S le jeu de données auto2004_don_manquante.txt (même source). Combien de valeurs manquantes sont contenues dans le fichier ?
9) Import into a variable named S the data set auto2004_don_manquante.txt (same source). How many missing values are contained in the file ?
10) Insert the VeryWeighty var which displays TRUE where Weight >= 1000

Exercise 3 :

1) Execute the following function and test it.

```
monexemple<-function(A,B){
 out<-(A+B)^2
 out<-out + A
 return (out)
}
```

2) Write in R a function CV which allows to calculate the coefficient of variation of a vector of numerical values.
   Recall that the coefficient of variation is a standardized measure of dispersion of a probability distribution or frequency distribution
   It is defined as the ratio of the standard deviation to the mean
3) Write in R a function PO which allows to determine by the least squares method (OLS) the unknown parameters in a linear regression model (the slope a and the intercept b to form the equation of the line)