PW02

R for data analysis

Exercise 1 :

Let M be the following matrix :

$$M = \begin{pmatrix} 2 & 4 & 8 \\ 2 & 5 & 9 \\ 2 & 3 & 8 \\ 1 & 4 & 7 \\ 1 & 5 & 1 \\ 1 & 2 & 3 \end{pmatrix}.$$

1) Create the matrix M in R
2) Create a new matrix M2 obtained by adding a new column to M of the following elements: 8, 7, 1, 2, 1, 2.
3) Transform the matrix into a data frame.
4) Assign names to the columns: case, month, week, value and to the rows, the first 6 letters of the alphabet. Show matrix
5) Extract the 2nd, 4th and 6th rows of M2.
6) Order the rows of M2 according to the column « value » in descending order

Exercise 2 :

Consider the dataset “airquality” which is available within R

1) Load the dataset
2) Consult the documentation of this dataset to understand its structure.
3) Display the names of the variables considered.
4) Display the dimensions of the dataset.
5) Give the quantiles of the Time/Temps variable.
6) Give the standard deviation for each of the variables in the dataset
7) Determine the coefficient of variation of the Time variable. Interpret
8) Extract from the base:

(a) the 3rd line,
(b) the 3-th column,
(c) lines 1, 2 and 4 with a single command,

(d) lines 3 to 6,
(e) all except columns 1 and 2,
(f) all lines with a temperature above 80.
9) Give a histogram which represents the frequencies for the values of the variable Temp with 20 classes.
10) Insert a new variable VeryWindy in the airquality dataset which displays TRUE if the value of Wind >= 10, and FALSE otherwise.
11) Delete the VeryWindy variable.
12) Create a new aq dataset corresponding to the airquality collection without the Ozone rows containing missing values
13) Install the ggplot2 package.
Test and analyze the output of the following command:
qplot(Temp, Ozone, data = airquality, color = Month)
14) Add the linear regression line and determine its coefficients

Exercise 3 :

Write a function with R that calculates the BMI (Body Mass Index). This index is given by a formula depending on the weight and height of the person.

$$BMI = \frac{weight}{height^2}$$

The function must display a message to the user depending on the BMI value :

- If imc<17, the function displays the message "Lean";
- If 17<=imc<27, the function displays "Normal";
- Otherwise (imc>=27), the function returns the message "obese"

Exercise 4 :

In statistics, the variance measures the variability from the average or mean in a dataset. It is calculated as follows :

$$var(x) = \frac{\sum_i^n (x_i - \overline{x})^2}{N - 1}$$

1) Write a function with R that calculates the variance.
2) Test the function and check the result with that of the predefined function in R.