Name-Surname : **Email :**

No : **Signature :**

HOME WORK 1 (Return by 09.01.2021)

BLM3590 – Statistical Data Analysis

T1(10)	T2(15)	T3(15)	T4(15)	T5(15)	T6(15)	T7(15)				Total(100)

The attached Excel file (**SdA-HW**) consists of two data types (embolic signals (**type 1**), and Doppler speckle (**type 2**)) recorded from stroke patients and some relevant numerical variables (**tpthrt**, **pkthrt**, **dfdrrt**, **time**, **rrt**, **frt**). Using this data file, implement the following tasks in **R**. You must include the **R** scripts in your answers.

==========

T1: Show how to read this Excel datafile into **R** environment.

==========

T2: This data file requires some preprocessing as it inludes a column with no value, some cells with no numerical value (divide by 0 error, etc.), and some cells with zero. Write required script in **R** to remove the empty column and correct the cells with no numerical value and zero by using simple interpolation.

==========

T3: Find **Five-number data summary** of the **variables** for each **data type** in this dataset.

==========

T4: Plot **boxplots** of the **variables** for each **data type** and determine if there is any outlier in these variables.

==========

T5: Plot histograms of the **variables** for each **data type**, compare the histograms, and comment on the distributions.

==========

T6: First, normalize the **variables** for each **data type** so that the values of these variables range between **0** and **1**, and then line-plot (using different colors) each variables for both data types in one figure (total 5 figures). Comment on the similarities of the variables for each plot.

==========

T7: First, determine how similar the variables **tpthrt** and **pkthrt** are for each **data type**, and then determine how similar **tpthrt of data type 1** and **tpthrt of data type 2** by using similarity metric (correlation).
