CS 504 - Fall 2017 - Project 1

Programming languages for data analysis

Due Date: Friday, November 24th, 2017

Create a Matlab interface to calculate statistical tests.

1. If data consist of one sample, the descriptive analysis should be applied.
2. If data consist of two samples, two- sample test (parametric or non-parametric test) should be applied. You must be able to choose the appropriate test based on the descriptive analysis of the two samples separately.
3. If data consist of three sample, three-sample test (parametric or non-parametric test) should be applied. You must be able to choose the appropriate test based on the descriptive analysis of the three samples separately.
4. If data consist of more than 3D, you should reduce dimension and then perform a two-sample test or three-sample test depending on the retained dimension.

To create the interface implementation, follow these steps.

1. Use the main script provided with the project to link your functions (Figure 1).
2. If data dimension is >1, you have to select columns one by one through the dialogue box and display their descriptive statistics then choose the appropriate test. For instance, if you have 2 samples, you select the first column, display the normal plot, and then the second column, and display the normal plot. Finally, you decide which test to use (student t-test or Wilcoxon Rank Sum Test) (Figure 2)
3. You should be able to enter the type of data (paired or unpaired), the significant level and the tail (one-tail test or two tail test) through a dialogue box.
4. If data dimension is > 3, you should apply PCA to reduce the dimension and save the scores in an excel file.

Figure 1

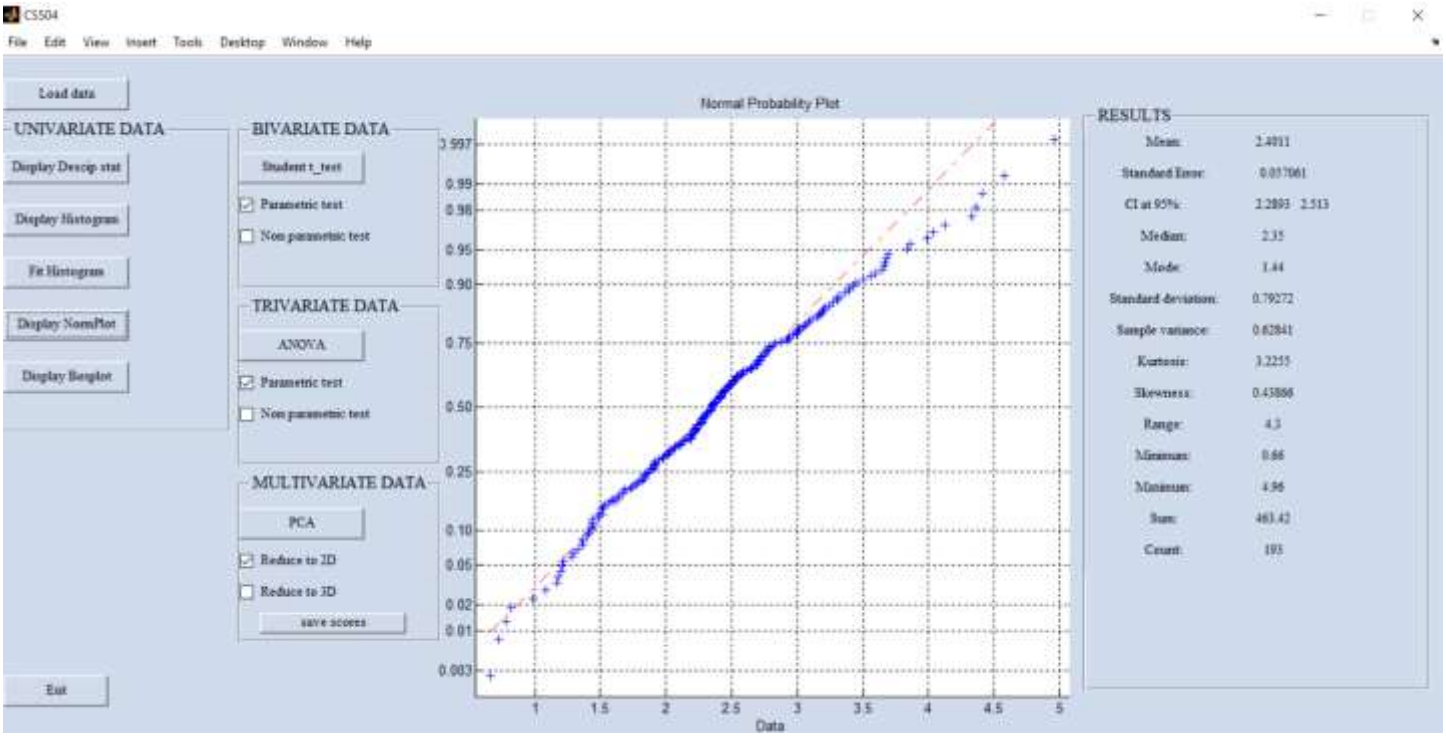

Figure 2