Data 501 - Assignment 1

David Huijser

2024-07-17

Assignment 1

Total 80 marks

The assignment is to create project / R program and distribute this on Github.

The user will provide the program with a data sets. The program will calculate and return the test statistics W used in a Shapiro–Wilk test used to determine if the sample is normally distributed.

The program will have the following features:

- at least two inputs (1 required and 1 default/optional)
  - required input is the data
  - optional/default input is up to you, for example a switch for QQ-plot output, or text output or display of other calculate quantities
- input validation
- at least two very basic/simple tests
- catches errors
  - NA values
  - inf values
  - wrong format
  - wrong dimensions

These are the suggested steps to follow:

1. Create empty repository on GitHub
2. Create a project in R-studio (and link with git)
3. Link the local git repository to with the Github repository
4. Develop the core R function
5. Develop the tests
6. Perform the test and edit code to catch potential errors

The features that get evaluated are:

1. Implementation of git, GitHub and commit history [10]
2. Functionality of the R-function [10]
3. Required Features
   - input [10]
   - output [10]
   - validation [10]
   - test [10]
   - error handling [10]
4. Presentation of results in Rmarkdown [10]

You will submit Rmarkdown file. This file displays the chunk of code of the main R-file, and explains the two test. The Rmarkdown file will also include the link to the GitHub repository (and if need additional instructions).