Data 501 - Assignment 2

David Huijser

Assignment 2

Total 100 marks

Due date: 23:00 16 August 2024

The assignment is to create R package project distributed on Github.

The aim of the package will be to create a graph similar to a Cooks distance plot (but with more distance measures).

The user will provide the program with a data set as well as a model (which is an object of class `lm`)

The functionality of the program entail performing a simple diagnostics on the influence of points.

The code should be able to calculate at least three measures of influence:

- Cooks Distance Measure (Cook, 1977)
- DFFITS (Welsch and Kuh, 1977; Belsley, 1980)
- Hadis Influence Measure (Hadi, 1992)

Note: Regression Analysis by Example (Hadi, Ali S; 2012; 5th Edition,Chapter 4) might be helpful.

The program will have the following features:

- at least three inputs (2 required and 1 default/optional)
    - required input is the data and model (a `lm` object)
    - optional/default input is up to you
- input validation
- perform different distance measure depending on arguments
- catches errors
    - NA values
    - inf values
    - wrong format
    - wrong dimensions
- create graphs visualizing so results
- typical package features
    - installable
    - proper unit tests
    - passes Checks
    - Manual AND Vignettes
- generate some return output: either influence measure or a plot-object

You will submit pdf generated from a Rmarkdown file. This file gives opportunity to explain some background, particular choices and rational of your implementation. This pdf will also include a link to a GitHub repository and/or instructions to install.