Statistical Methods 2022

# Submission:

The deadline is 18th November 2022, 23:55 o’clock.

## Format:

A single zip file named as firstname\_lastname\_city.zip

Content of the zip file as follows:

1. A document of your report in word or pdf format (5 to 10 pages).
2. A directory (folder) containing R scripts that covers the necessary parts of your report and calculations. Consider proper commenting, variable naming and consistent conventions in the file(s) to help understand the logic in the code.

# Assignment:

You will try to provide a report about a dataset that you choose, I highly recommend that you use one of the default data sets in R. The content of your report should have following structure:

1. Introduces the data and it’s context.
2. Provides a briefed information about its variables, its variable types (from statistician’s perspective), dataset’s size, etc.
3. Proposes a question that can be answered though the quantitative analyze of data.
4. Provides the sufficient amount of documentation regarding how you analyzed and answered the question.

Your assignment heavily circles around problem solving, decision making, creative thinking as well as being capable of understanding and calculating mathematical/statistical/probabilistic topics you’ll face while trying to answer the question.

Having these factors in mind, the student should follow the below criteria to fulfill the competency required to pass the course :

1. Your submission should be original and authentic, meaning all four aspects of the assignment shouldn’t be done prior to this course and/or by anyone else but yourself. (using of previous material regarding the question, calculation and code is not allowed. If a package in R or a dataset provides the insight, question and/or calculation, it’s not allowed to be used in this assignment).
2. The section 4, answering the question should be benefiting the Least-squares line equation and a proof of it’s usability in percentage, coefficient of determination.

For passing the course with distinction, the student’s submission also should have the following:

1. Answer the question with not only how but a deeper explanation of why they took majority of the steps for answering the question.
2. The student also should provide the inner calculation for each formula in a reasonable depth, for example no need to add all of the data for mean, just sum of x equals y and then dividing the y to n is sufficient.
3. The student also should provide some insightful information with help of probability, using the z-score and z-table. For example the gas consumption of the cars with engine size of x or greater, is more than 50 percent of the all cars in the dataset.

Good luck!