Statistical and Predictive Modeling I (DATA 1204)
Assignment #2 – Hypothesis Testing (15% of Total Grade)
Professor: Fatma Tetikoglu

Recently a cholesterol level experiment was conducted on some individuals. The results can be found in the **CholestrolLevel.xlsx** database.

Goal of Analysis

Previous analysis has shown that the population mean of the cholesterol level is 5.95 and the standard deviation is 0.897. The research team is trying to determine whether an individual's cholesterol level ***has changed or not*** after applying a certain diet program properly for the past 6 months.

Questions:

1. State the Hypothesis Statement (by using an 'if' statement) that helps solve the research question. **(1%)**
2. Assess the normality of the dataset by QQ plot **or** a histogram. **(2%)**
3. Provide a written *step-by-step outline (in a total of 5 steps)* on how you would prove (or dis-prove) your hypothesis statements that you developed above. **(8%)**
4. Conduct the analysis you outlined in #3 in R. **(2%)**
5. Create a "Summary of Findings and Conclusion" section for your analysis. **(2%)**

Hint: Leverage the Week #5 – Exercise

Important Note: To use the dataset, we need to import it to R. To do that:

1. Please copy **CholestrolLevel.xlsx** dataset into your RData1204 directory. Please ensure that the file is called **CholestrolLevel.xlsx**.
2. Using the top right box click the "Environment" tab.
3. Now click "Import Dataset".
4. Select "From Excel" since it's an .xlsx file.
5. Click "Browse" on the right top of the pop-up page.
6. Find the **CholestrolLevel.xlsx** file that you downloaded.
7. Click "Open"
8. Click "Import"

Please post your Word Document (.doc or .docx) including your R script via assignments under Assignment #2 by 11:59 pm on Friday, February 9, 2024.

Remember that DC Connect doesn't accept R script files. All you need to do is to copy the code from your top left panel in R and paste it at the bottom of your word document.

Grading Rubric

Question	Exemplary (14-15)	Proficient (10-13)	Incomplete (7-9)	Needs Improvement (0-6)
1	Hypothesis Statement is properly given.	Hypothesis Statement is incomplete.	Hypothesis Statement is incorrect.	Hypothesis Statement is absent.
2	The normality of the dataset by QQ plot or a histogram is assessed properly.	The normality of the dataset by QQ plot or a histogram is incomplete.	The normality of the dataset by QQ plot or a histogram is incorrect.	The normality of the dataset by QQ plot or a histogram is absent.
3	Step-by-step outline on how to prove (or dis-prove) the hypothesis is properly explained in 5 steps.	Steps of the hypothesis testing are incomplete or missing some information and explanation.	Steps of the hypothesis testing are incorrect.	Steps of the hypothesis testing are absent.
4	Analysis described in Question#2 is conducted properly in R and explained clearly. The script and output are presented in the submission.	Analysis described in Question#2 is incomplete or missing some explanations.	Analysis described in Question#2 is given in a generic form or incorrectly.	Analysis described in Question#2 is absent.
5	Summary of findings and conclusion is given properly and summarizes the analysis correctly.	Summary of findings and conclusion is incomplete or missing some results.	Summary of findings and conclusion is incorrect.	Summary of findings and conclusion is absent.