

IE6200: Engineering Probability and Statistics

Assignment: Lab 03

Prof. Mohammad Dehghani



Lab Assignment Guidelines

1. Students need to complete the Lab assignment individually.
2. All the Lab assignments are required to be done in RStudio.
3. Provide necessary comments using '#' for better understanding of your script.
4. The lab report needs to include the following sections:
 - **Problem statement:** A brief about your understanding on the assignment questions (maximum 3 lines)
 - **Output:** What were your finding after creating the code and running it in R. This section may include:
 - Graphs / charts / plots
 - Final data frame for your result
 - Results obtained
 - **Conclusion:** What were the statistical inferences and observations from the results obtained.
✓Students are not required to include codes in reports.
5. If you take help from any external source, please mention that in reference. Violating academic integrity policies may include zero credit on the work.

Deliverables:

1. Please submit a *.zip file including the following items
 - i. R script (just 1 file including all your codes)
 - ii. Lab Report: Report with a maximum length of 10 pages including all appendices, tables, and graphs.
2. All of the above mentioned files have to be labeled as: 'Lab # - IE 6200 – Sec # - <Student Name>'
3. Submit your Lab deliverables via Canvas.

Task 1

Import the Health dataset provided on Blackboard in R and save it as a data frame.

Task 2 (30 points)

- a) Find the frequency of people in each of the BMI categories depicted in the image below.

Category	BMI
Underweight	$\text{BMI} < 18.5$
Normal	$18.5 \leq \text{BMI} < 25$
Overweight	$25 \leq \text{BMI} < 30$
Obese	$\text{BMI} \geq 30$

- b) Find the number of people under the normal and overweight category who suffered from *Hypertension* (Hypertension == **TRUE**) and compare their results. Use barplot to illustrate your answer. What can you interpret from the barplot ?

[Hint: Use geom_bar\(\) to get the barplot](#)

- c) Calculate the following statistics for Normal BMI.

- Mean
- Median
- Range
- Interquartile Range
- Variance
- Standard Deviation

Task 3 (20 points)

- a) Find the percentage of people when each of following conditions are met,

Stroke	Hypertension
TRUE	TRUE
TRUE	FALSE
FALSE	TRUE
FALSE	FALSE

Obtain a table as shown below.

Stroke	Hypertension	Percentage
TRUE	TRUE	%
TRUE	FALSE	%
FALSE	TRUE	%
FALSE	FALSE	%

Task 4(30 points)

- a) Calculate the following for the Avg_Glucose_Level
- Coefficient of Variation
 - Skewness
 - Kurtosis
- b) Plot a histogram of Avg_Glucose_level.
- c) Compare the histogram and the results obtained in part (a) and interpret the results.

Task 5 (10 points)

- a) Plot a scatterplot for Age and Avg_Glucose_Level for people who suffered from Heart_diseases (Heart_diseases = **TRUE**) and those who did not (Heart_diseases = **FALSE**) `geomvline()` and `geomhline()` functions so data can be categorised in the following age groups
- Age < 30
 - 30 < Age < 60
 - Age > 60

What can you infer from the scatterplot?

Task 6 (10 points)

- a) Analyze the data and discover any interesting facts (atmost 2) about the dataset and relationship between variables.
- Note: The analysis should be unique and not exceed more than one page.**