

IE6200: Engineering Probability and Statistics

Assignment: Lab 04

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 Bluebikes dataset provided along with the assignment in R and save it as a data frame.

Task 2

Using the data, compute the following statistics for the number of bikes that were picked up or dropped off (either one or both if needed) from any station of your choosing (except for the ones analysed in class) between 10:00 AM and 11:00 AM

1. Probability Mass Function (PMF)
2. Continuous Distribution Function (CDF)
3. Expected Value
4. Joint Probability
5. Correlation Coefficient

Create any visualizations that you feel appropriate to convey your findings. (Example: heatmap for joint probability)

Task 3 (Bonus Points)

Search for datasets from reputed online data repositories (such as UCI Machine Learning Repository) and provide details about those for which similar statistics (PMF, CDF, Joint Probability) can be calculated.