DS201/DSL253: Statistical Programming

Assignment 3

30.01.2025

Instructions for Submission: You can submit your solution as a Jupyter Notebook/Matlab file with comments and discussions on the results obtained in each step.

1. Follow Standard Report Format: Include sections like Introduction, Data, Methodology, Results, Discussion, and Conclusion.
2. File Naming Convention: Adhere to the specified naming convention for each file you submit (e.g., RollNumber FirstName Asg1).
3. Refrain from using zip files. If necessary, submit multiple files.
4. Include comments in the code explaining the logic and any assumptions made.
5. Include References: Cite any external sources or references used in your assignment.
6. Code Quality: Ensure your code follows best practices, is well-organized, and avoid plagiarism as a plagiarism check will be conducted.
7. Be aware that late submissions are not permitted; ensure timely submission.
8. Coding can be done in any language.

1. A company is testing a new marketing strategy to attract potential customers. They run several campaigns and analyze the response rates to understand their effectiveness.

(a) Suppose the marketing team reaches out to 15 potential customers, and the probability of a positive response (conversion) from each customer is 20%. Plot the probability mass function (pmf) for the number of successful responses. This will help the team understand the likelihood of achieving different levels of success. [Hint : You can use the inbuilt packages]

(b) The team wants to explore how different response probabilities might affect campaign performance. Repeat part (a) by considering probabilities of conversion as 10%, 20%, ..., 90%. Analyze and comment on how the distribution changes as the probability of a successful response increases.

(c) To measure long-term campaign performance, the team introduces a success rate metric defined as the number of successful responses divided by the total number of people contacted. Let

$$Y = \frac{X}{n}$$

where X represents the number of successful responses from a campaign targeting n individuals, with a conversion probability of 5%. Generate plots for campaigns targeting 10, 20, 50, and 200 customers. Discuss how the success rate distribution changes as the number of targeted customers increases and what this might imply for large-scale marketing efforts.

2. The lifetime (in hours) of a certain battery used under extremely cold conditions follows a Gamma distribution with parameters :

- Shape parameter ($\alpha = 5$)
- Scale parameter ($\theta = 4$)

Using any programming language, answer the following questions:

(a) Calculate the average expected lifetime of the battery and its variability display the values.

(b) Find the median lifetime of the battery, representing the point at which half of the batteries are expected to last longer and half shorter. The median can be found using the inverse cumulative distribution function:

$$P(X \leq \text{median}) = 0.5$$

(c) Generate a plot of the probability density function (PDF) for battery lifetimes ranging from 0.1 to 50 hours to better understand the distribution of expected lifetimes.

(d) Repeat the same for the following cases of parameters:

  (i) Shape parameter ($\alpha = 10$), Scale parameter ($\theta = 0.9$)

  (ii) Shape parameter ($\alpha = 7$), Scale parameter ($\theta = 2$)

  (iii) Shape parameter ($\alpha = 1.5$), Scale parameter ($\theta = 0.2$)

3. A company has developed a new type of battery whose lifetime (in hours) follows a chi-squared distribution with 10 degrees of freedom. Using any programming language, answer the following questions:

(a) Compute the mean and standard deviation of the battery lifetime. Also, compute the median, first quartile (Q1), and third quartile (Q3) of the lifetime.

(b) Generate a plot of the battery lifetime distribution (PDF) from 0 to 24 hours, clearly indicating the mean, median, and quartiles on the graph. Also, make the mean, median, and quartiles on the plot.

(c) Using MGF function, generate the first, second, third moments of the chi-squared distribution