DS201/DSL253: Statistical Programming

Assignment 07

27.03.2025

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) A quality control team at a manufacturing plant wants to ensure that the average weight and deviation of their product batches meet industry standards. Assume that the weights follow normal distribution, $W \sim \mathcal{N}(\mu, \sigma^2)$. To do this, they take random samples of size n and calculate confidence intervals at $(1-\alpha)\%$ confidence levels for both the average weight and deviation. However, how reliable are these confidence intervals in truly capturing the actual batch parameters? If this sampling process is repeated m times for each choice of n and α, what proportion of intervals successfully captures the true average weight and variation? Write a program to conduct the above experiments and based on the results, determine what insights can be drawn about the impact of sample size and confidence level on the accuracy of these estimates?

b) In a real-world setting, minor fluctuations in weight measurements can occur due to machine calibration errors. Suppose a random additive uniform distribution noise, η in the range (-1,1), is present in each measurement. How does this affect the confidence intervals? What impact does this have on the proportion of intervals that successfully capture the true average weight and variability?

2.

A pharmaceutical company is testing two different drug formulations to compare their effects on blood pressure reduction. The first formulation is tested on a group of patients, with results modeled as $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ with n_1 number of samples , while the second formulation is tested on another group, modeled as $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ with n_2 number of samples. Write a program to generate the two sets of samples (explore different values of the parameters) and determine the confidence interval for the difference in average effectiveness between the two formulations? If this experiment is repeated m times, what proportion of intervals successfully capture the true difference in effectiveness, and what patterns emerge?

3.

In a closely contested two-way election, a pollster surveys a group of voters to estimate the proportion supporting Candidate A. Assuming the responses follow a Bernoulli distribution, $X \sim$ Bernoulli(p), with an unknown p. Write a program to generate the samples and determine a confidence interval for given p and $(1-\alpha)\%$ values. If this process is repeated m times for different values of p and m, what patterns emerge in the intervals formed?