DS201/DSL253: Statistical Programming

Assignment 08

17/04/2025

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.

Question 1

A manufacturer claims that the average lifetime of its batteries is 500 hours, with a known standard deviation of 100 hours. A quality control team tests 30 randomly selected batteries, obtaining the following lifetimes
495, 520, 510, 505, 480, 500, 515, 495, 510, 505, 490, 515, 495, 505, 500, 510, 485, 495, 500, 520, 510, 495, 505, 500, 515, 505, 495, 510, 500, 495
Perform a hypothesis test at a 5% significance level to determine whether the mean battery lifetime differs significantly from 500 hours. Also, compute the p-value and plot the Operating Characteristic (OC) curve for a range of true means.

Question 2

A public health official claims that the mean home water use is 350 gallons per day. To verify this claim, a study was conducted on 20 randomly selected homes, and their daily water usage (in gallons) was recorded as follows:
340, 344, 362, 375, 356, 386, 354, 364, 332, 402, 340, 355, 362, 322, 372, 324, 318, 360, 338, 370
Consider two cases:

(a) Suppose that the population variance is known to be 144. Perform a hypothesis test at a 5% significance level to determine whether the data provides sufficient evidence to contradict the official's claim. Compute the test statistic and p-value, and state your conclusion.

(b) Now, suppose that the population variance is unknown. Perform a one-sample t-test at a 5% significance level using the given sample to determine whether the data provides sufficient evidence to contradict the official's claim. Compute the test statistic and p-value, and state your conclusion.

Question 3

A nutritionist wants to determine whether a new diet plan significantly affects body weight. She selects 10 individuals and records their weight (in kg) before and after following the diet for one month. The recorded weights are as follows:

Participant	Before (kg)	After (kg)
1	85.2	82.5
2	78.5	75.8
3	92.3	90.1
4	80.0	77.2
5	88.7	85.4
6	76.4	74.5
7	90.5	87.6
8	84.1	81.3
9	79.0	76.8
10	86.2	83.0

Perform a paired t-test at a 5% significance level to determine whether the diet plan has a significant effect on body weight.

Write a program to:

(i) Formulate and perform the hypothesis test.

(ii) Compute and report the test statistic and p-value.

(iii) State the conclusion based on the p-value.

Question 4

The quality control team at a pharmaceutical company needs to verify whether their IV fluid filling machine maintains the advertised consistency. The manufacturer claims the volume variance is no more than 4 mL^2 ($\sigma^2 \leq 4$). A random sample of 15 bottles showed the following volumes (in mL):

502, 498, 505, 497, 503, 499, 504, 496, 501, 500, 506, 495, 502, 498, 504

Write a program to:

(i) Calculate the sample variance from this dataset

(ii) Perform a chi-square test at $\alpha = 0.01$ significance level to determine if the machine violates the variance specification

(iii) Investigate how removing potential outliers (volumes < 495 mL or > 505 mL) affects the conclusion.