DSL201: Statistical Programming

Dr. Anil Kumar Sao

Assignment 4

Deadline: 8/04/2024 11:59pm

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_Asg2).

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 and 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.

---

1. Consider a random sample $X_1, X_2, \ldots, X_{10000}$ drawn from a Poisson distribution with mean θ. Show that the critical region C defined by $\sum_{i=1}^{10000} X_i \geq 3$ is a best critical region for testing $H_0 : \theta = 0.2$ against $H_1 : \theta = 0.6$. Determine, for this test, the significance level α and the power at $\theta = 0.6$.

    1. Write a Python function to generate a random sample of size 10000 from a Poisson distribution with mean θ.

    2. Define the critical region C as the sum of the sample being greater than or equal to 3.

    3. Perform hypothesis testing by comparing the critical region with the significance level α.

    4. Calculate the power of the test at $\theta = 0.6$.

    Ensure to use Python from scratch to solve the given problem statement.

2. Consider a random sample $X_1, X_2, \ldots, X_{100}$ drawn from a normal distribution with mean μ and standard deviation σ. Show that the critical region C defined by $\bar{X} \geq 20$ is a best critical region for testing $H_0 : \mu = 15$ against $H_1 : \mu = 18$. Determine, for this test, the significance level α and the power at $\mu = 18$.

1. Write a Python function to generate a random sample of size 100 from a normal distribution with mean μ and standard deviation σ.

2. Define the critical region C as the sample mean $(\bar{X})$ being greater than or equal to 20.

3. Perform hypothesis testing by comparing the critical region with the significance level α.

4. Calculate the power of the test at $\mu = 18$.

Ensure to use Python from scratch to solve the given problem statement.

3. In a study on music's effect on memory, 40 participants are split: one group listens to classical music while studying, and the other studies silently. After a memory test, scores are recorded.

Scores of the classical music group:

78, 82, 85, 79, 81, 83, 80, 77, 79, 84, 76, 82, 80, 81, 83, 78, 82, 79, 81, 83,77, 80, 81, 84, 79, 82, 80, 82, 83, 78, 75, 78, 82, 83, 76, 80, 82, 84, 77, 79

The mean score before the study is 75. Researchers hypothesize that listening to classical music improves memory, leading to a sample mean significantly higher than 75.

Test if the sample mean is significantly different from 75 at the 0.05 level. Report p value.

After review, one score of 85 was mistakenly recorded as 75. Are the corrected scores significantly different from 75 at the 0.05 level?