UCSC

SCS2211 - LABORATORY II
R Lab Practical Sheet - 14

Instructions

- **Do the Activities given and prepare a report , report should be in PDF format**
- **File name should be <Index number>.pdf (Eg: 2000000.pdf) and upload to the given link.**
- **Any form of plagiarism or collusion is not allowed**

Data Analysis and Hypothesis Testing with the Iris Dataset in RStudio

Assignment Instructions

You are required to analyze and manipulate the **Iris dataset** in **RStudio** while covering various statistical and visualization techniques. Follow the tasks below and compile your findings into a report using the given format.

Activities

1. Dataset Exploration

- Load the **Iris dataset** in RStudio.
- Display the **structure, summary statistics, and first few rows** of the dataset.
- Identify the **number of species** present in the dataset.
- Determine the **mean, median, and standard deviation** of each numerical feature.

2. Data Visualization

- Create a **Pie Chart** to represent the **species distribution** in the dataset.
- Generate a **Bar Chart** to visualize the count of each species.
- Plot a **Histogram** for the Sepal Length and Petal Length to examine their distributions.
- Create a **Scatterplot** between **Sepal Length and Petal Length** and analyze any correlation.

3. Hypothesis Testing

Conduct three different hypothesis tests on the dataset. Use a significance level of **α = 0.05**.

- **Lower Tail Test:**
  - Test whether the **average Sepal Length is significantly lower than 5.8 cm**.
  - State the null and alternative hypotheses.
  - Calculate the test statistic and p-value.
  - Interpret the results.
- **Upper Tail Test:**
  - Test whether the **average Petal Length is significantly greater than 3.5 cm**.
  - Define the hypotheses.
  - Compute the test statistic and p-value.
  - Explain the conclusion.
- **Two-Tailed Test:**
  - Test whether the **average Sepal Width is significantly different from 3.0 cm**.
  - Set up the null and alternative hypotheses.
  - Compute the test statistic and p-value.
  - Provide an interpretation of the findings.

Report Format

Prepare a structured report following this format:

1. **Title Page**
   - Assignment Title, Your Name, Date
2. **Introduction**
   - Overview of the dataset and objectives of the analysis
3. **Methodology**
   - Steps taken in the analysis and justification
4. **Results**
   - Findings with visualizations and tables
5. **Discussion**
   - Interpretation and significance of results
6. **Conclusion**
   - Summary and potential future work
7. **References**
   - Cite sources (if any)