Statistical Analysis using Python and R

Lab Objective

To perform descriptive and inferential statistical analyses using both Python and R, and to interpret the outcomes for real-world data sets. This lab will reinforce the role of statistics in AI tasks like data understanding, feature engineering, and modeling.

Theory Overview

Statistical techniques help in summarizing, analyzing, and making inferences from data. Key statistical techniques covered include:
- Descriptive statistics: Mean, median, mode, variance, standard deviation, skewness, kurtosis.
- Inferential statistics: Correlation, hypothesis testing (t-test), p-values, confidence intervals.

Lab Tasks

Part A – Descriptive Statistics in Python

1. Load a dataset using pandas (e.g., Titanic, Iris, or custom CSV).
2. Calculate:
   - Mean, median, mode
   - Variance and standard deviation
   - Skewness and kurtosis
3. Display summary using df.describe()
4. Plot distributions using seaborn.histplot() or boxplot()

Part B – Inferential Statistics in Python

1. Compute correlation matrix using df.corr().
2. Conduct a t-test using scipy.stats.ttest_ind() to compare two sample means.
3. Interpret p-values and confidence intervals.

Part C – Descriptive and Inferential Stats in R

1. Load the dataset using read.csv() or datasets::iris
2. Compute:
   - mean(), median(), sd(), var()
   - Use summary() and cor()
3. Run a t-test using t.test()
4. Create boxplots and histograms with ggplot2