Day 09

Topic Covered: Statistical Analysis in R

Summary:
Today's session focused on statistical analysis in R, which is an essential part of data analytics and research. R is an open-source statistical programming language widely used across academia and industry. It offers a large collection of statistical packages, with more than 18,000 available on CRAN for specialized analysis.
The session covered two key areas: basic statistical analysis and statistical hypothesis testing. These methods help summarize numerical data, identify patterns, and make informed decisions using sample data.

New Concepts Learned:

1. **Basic Statistical Analysis**
   We used a sample vector y_vector <- c(3, 4, 23, 67, 23, 78, 68, 34) to understand basic functions.
   Key functions learned:

   - mean(): Calculates the average
   - median(): Finds the middle value
   - sd(): Computes standard deviation
   - var(): Calculates variance
   - Mode calculation using table() and sorting
   - Quantiles using quantile() to analyze distribution at different points

2. **Statistical Hypothesis Testing**
   Hypothesis testing compares two claims using sample data.
   Example vectors used:
   vector1 <- c(34, 23, 67, 34, 12, 67)
   vector2 <- c(78, 23, 45, 36, 28, 90, 16, 37)

Key techniques learned:

- t.test(): Compares the means of two groups
- chisq.test(): Performs chi-square test for categorical or frequency data
- var.test(): Compares variances using an F-test
- cor.test(): Tests correlation between two numeric variables

Activity:

- Calculated mean, median, standard deviation, variance, and quantiles for sample vectors
- Computed mode manually using table()
- Performed t-test, chi-square test, F-test, and correlation test on example data
- Observed how R simplifies statistical calculations and interpretations

Challenges Faced:

Understanding which statistical test to apply for a given scenario was slightly confusing. Calculating the mode manually without a built-in function also required careful steps.

Key Takeaway:

Statistical analysis in R helps in summarizing data and comparing groups effectively. Hypothesis testing provides a structured method for making data-driven conclusions. Learning these functions strengthens analytical skills and improves the ability to interpret and present meaningful insights.