Lecture 10 Summary - Statistics Using Python and Pandas

1. Definition of Statistics:

- Statistics: A branch of math focused on data understanding.

- Applied Statistics includes:

  - Descriptive Statistics: Summarizes data.

  - Inferential Statistics: Draws conclusions from samples.

2. Descriptive Statistics:

Divided into two parts:

  A) Measures of Central Tendency:

    - Mean: Sum of values / count. Good for normal data.

    - Median: Middle value in sorted list. Best with outliers.

    - Mode: Most frequent value.

  B) Measures of Variability:

    - Range: Max - Min.

    - Variance: Average of squared differences from the mean.

    - Standard Deviation: Square root of variance.

    - Quartiles and IQR: Q1, Q2, Q3 and IQR = Q3 - Q1.

3. Skewed Distributions:

- When skewed, mean and median differ more.

- Median is better for central tendency in this case.

4. Statistical Plots:

- Boxplot: Shows median, quartiles, and outliers.

5. Normal Distribution Test using IQR:

- Compare z-values for Q1 and Q3 to check normality.

6. Covariance and Correlation:

- Covariance: Measures joint variability of two variables.

- Correlation: Scaled measure, between -1 and 1.

7. In Python:

- Tools exist for mean, median, mode calculations.

- Boxplot and heatmap are supported.