

## **Statistics:**

Statistics is the science of collecting, organizing, analyzing, and interpreting data. It has two main areas:

- **Descriptive Statistics:** Summarizes data using measures like mean, median, mode, and range, variance, standard deviation.
- **Inferential Statistics:** Draws conclusions from sample data using techniques like hypothesis testing, confidence intervals, and regression analysis.

## **Probability:**

Probability is the study of the likelihood of events occurring. Key concepts include:

- **Random Experiment:** An unpredictable process.
- **Sample Space:** All possible outcomes.
- **Event:** A specific outcome or group of outcomes.
- **Probability of an Event:** A number between 0 and 1 indicating the chance of the event happening.

**Mean:** The average of a set of numbers, calculated by summing all values and dividing by the number of values.

**Mode:** The value that appears most frequently in a dataset.

**Median:** The middle value in a sorted list of numbers, or the average of the two middle values if there is an even number of values.

**Confidence Interval:** A confidence interval is a range of values derived from sample data that is used to estimate the range within which a population parameter such as the mean or proportion is likely to lie. It is accompanied by a confidence level that expresses the level of confidence that the parameter falls within the interval

**Covariance & Correlation:** In data science, **correlation** identifies how closely two sequences vary together: they may rise and fall simultaneously, or move in opposite directions. Correlation quantifies this relationship, ranging from -1 (strong negative correlation) to +1 (strong positive correlation), with 0 indicating no linear relationship. **Covariance**, the basis for correlation, measures how deviations from mean values of two variables relate: positive for joint deviations and negative for opposite deviations. Normalizing covariance by dividing by standard deviations yields correlation, a standardized measure facilitating comparison across different scales.

