

Statistics & Probability

Overview:

Statistics and probability form the mathematical foundation of data science. They help describe data characteristics, relationships, and uncertainties.

Key Statistical Measures:

- **Mean:** Average value of a dataset.
- **Median:** Middle value when data is sorted.
- **Mode:** Most frequent value in the dataset.
- **Variance & Standard Deviation:** Measure how much data points differ from the mean (spread of data).
- **Correlation:** Measures the strength and direction of a relationship between two variables (ranges from -1 to +1).

Probability Concepts:

Probability estimates how likely an event is to occur. In data science, it is essential for understanding uncertainty and supporting model predictions.

Correlation Analysis:

- Used to find how features relate to the target variable.
- Helps identify which variables are most predictive or influential.
- Can be visualized using **correlation matrices** or **heatmaps**.

Importance:

Understanding statistical relationships guides feature selection and model interpretation, improving both accuracy and reliability.