DAY 5 – Statistics Basics for Data Analytics

Objective

To understand the basic statistical concepts required for data analytics, business analysis, and interviews.

1. What is Statistics?

Statistics is the science of collecting, organizing, analyzing, and interpreting data to make decisions.

It helps in understanding data patterns, comparing performance, and predicting future trends.

2. Mean (Average)

Mean is the average value of a dataset.

Formula: Mean = Sum of values / Number of values

Example: 20, 30, 40, 40, 50, 60 → Mean = 40

Key Point: Mean is affected by outliers.

3. Median

Median is the middle value of an ordered dataset.

Example: 20, 30, 40, 40, 50, 60 → Median = 40

Key Point: Median is not affected by outliers.

4. Mode

Mode is the most frequently occurring value in a dataset.

Example: 20, 30, 40, 40, 50, 60 → Mode = 40

5. Variance

Variance measures how far data values spread from the mean.

Example Variance = 166.67

Key Point: Higher variance means more spread in data.

6. Standard Deviation

Standard deviation is the square root of variance.

Example Standard Deviation ≈ 12.91

Key Point: Lower SD means more consistent data.

7. Correlation

Correlation measures the relationship between two variables.

Range: -1 to +1

Example: Advertising spend vs Sales shows positive correlation.

Key Point: Correlation does not imply causation.

Excel Formulas Used

Measure	Excel Formula
Mean	=AVERAGE(A1:A6)
Median	=MEDIAN(A1:A6)
Mode	=MODE(A1:A6)
Variance	=VAR.P(A1:A6)
Standard Deviation	=STDEV.P(A1:A6)
Correlation	=CORREL(A1:A6,B1:B6)

Conclusion

Statistics is the foundation of Data Analytics and is essential for Excel, Python, Power BI, and Machine Learning.