Chapter 7: Statistics

Class 9 – Artificial Intelligence

Introduction

Statistics is a branch of mathematics that deals with **collecting, organizing, analyzing, and interpreting data** to make informed decisions. In the field of Artificial Intelligence (AI), statistics plays a **foundational role** because AI systems rely on **data** to learn and improve. Whether it's training a machine learning model or visualizing patterns, statistical tools are essential.

In this chapter, you will explore how **statistics** is used to understand **data**, perform basic **data analysis**, and extract meaningful **insights**, which form the backbone of intelligent systems.

7.1 What is Data?

◈ Definition:

Data refers to **raw facts or figures** that by themselves may not make sense. Once processed, data becomes **information**.

◈ Types of Data:

1. **Qualitative Data (Categorical):**

    o Represents categories or labels.
    o Examples: Gender (Male/Female), Type of AI (Narrow/General).
2. **Quantitative Data (Numerical):**

    o Represents numbers or quantities.
    o Examples: Age, Number of students using AI tools.

7.2 Collection of Data

◈ Primary Data:

- Collected **directly by the investigator**.
- Example: Conducting a survey among students.

◈ Secondary Data:

- Collected by **someone else** and used for analysis.

- Example: Data from government records or published reports.

7.3 Organization of Data

Once data is collected, it must be organized to **observe patterns and trends**.

◆ Frequency Distribution Table:

- Shows how often each data value occurs.
- Example:

Marks	Frequency
0–10	3
11–20	5
21–30	8

◆ Tally Marks:

- A simple way to count frequency using vertical bars.

7.4 Graphical Representation of Data

Graphs and charts help in **visualizing data**, making interpretation easier.

◆ Bar Graph:

- Used for **categorical data**.
- Bars represent frequency of each category.

◆ Histogram:

- Used for **continuous data**.
- Bars are joined with no gaps.

◆ Pie Chart:

- Represents data as **portions of a circle**.
- Useful to compare parts to the whole.

◆ Line Graph:

- Shows **trends over time**.
- Useful for time-series data.

7.5 Measures of Central Tendency

These measures represent the **center or typical value** of a dataset.

1. **Mean (Average):**

- Formula:

$$\text{Mean} = \frac{\text{Sum of all observations}}{\text{Number of observations}}$$

- Example: For data [5, 10, 15], Mean = (5 + 10 + 15) / 3 = 10

2. **Median:**

- The **middle value** when data is arranged in ascending order.
- If even number of observations: Median = average of two middle numbers.

3. **Mode:**

- The value that occurs **most frequently** in the dataset.

- Example: [4, 6, 6, 7, 9] → Mode = 6

7.6 Importance of Statistics in AI

⬦ AI Relies on Data:

- AI algorithms require **large datasets** for training and testing.

⬦ Pattern Recognition:

- Statistics helps identify **patterns, correlations**, and **outliers** in data.

⬦ Data Preprocessing:

- Cleaning and preparing data involves **statistical methods**.

⬦ Predictive Modeling:

- Many **machine learning models** use statistical theories to make predictions.

7.7 Applications of Statistics in AI

AI Field	Statistical Use
Healthcare	Predicting disease risk from patient data
Education	Analyzing student performance trends
Finance	Forecasting stock prices

AI Field	Statistical Use
Agriculture	Yield prediction and climate pattern analysis
Social Media	User behavior analysis

Summary

- **Statistics** is the science of working with data—collecting, organizing, analyzing, and interpreting it.
- It helps in making **data-driven decisions**, which is the core of **Artificial Intelligence**.
- Key concepts include **mean, median, mode**, data types, and graphical representations.
- AI heavily uses statistics for **training algorithms**, **evaluating models**, and **deriving insights**.
- Real-world applications in fields like healthcare, education, and finance show the relevance of statistics in building **intelligent systems**.