Day30_ Introduction_to_Statistics

June 27, 2025

Day 30 – Introduction to Statistics (for Machine Learning)

Welcome to Day 30! Today we started with the **basics of statistics**, a powerful tool that helps us understand data and make smart decisions.

What is Statistics?

> **Statistics** is the science of collecting, organizing, analyzing, interpreting, and presenting data.

It helps in:

- Identifying patterns
- Making predictions
- Supporting decision-making

Simple Example: You want to find out how many students like cricket. You ask 100 of them and analyze the results. That's statistics in action!

Use of Statistics in Machine Learning

Statistics is the **foundation** of machine learning.

Here's how it helps:

Area	Statistical Concept	ML Application
Data Understanding	Mean, Median, Mode, Variance	Descriptive analysis before modeling
Pattern Recognition	Probability, Distribution	Naive Bayes, Hidden Markov Models
Prediction	Inference, Regression	Linear Regression, Logistic Regression
Hypothesis Testing	p-value, confidence interval	A/B Testing, model validation
Classification	Categorical Data Analysis	Binary/Multi-class classification

In short, **ML is built on statistics**.

Let's Brush Up – Key Statistical Foundations

We'll go step by step:

- Population vs Sample
- Types of Data
- Levels of Measurement

Let's begin

---

1 Population vs Sample

1.1 Population:

The **entire group** you're interested in studying.

1.2 Sample:

A **subset** of the population that represents the whole.

Example:

- Population: All customers of Amazon in India
- Sample: 1,000 customers selected from Delhi, Pune, and Mumbai

---

2 Inference Techniques (From Sample to Population)

Using a **sample to make conclusions** about a population is called **statistical inference**.

Common methods:

- **Confidence Intervals**
- **Hypothesis Testing**
- **Z-test / T-test**
- **P-values**

Example:
If 70% of your sample likes a product, you infer the population might also like it.

---

3 Sampling Techniques (From Population to Sample)

Technique	Description	Example
Simple Random	Everyone has equal chance	Lottery draw of 100 students
Stratified Sampling	Divide into groups, pick from each	Sample male/female separately
Cluster Sampling	Divide into clusters, pick entire group	2 colleges out of 10
Systematic Sampling	Every k-th item	Every 10th person in a list

Technique	Description	Example
Convenience Sampling	Easy access	Ask people nearby only

Good sampling = Reliable Inference