

ikSaan.com

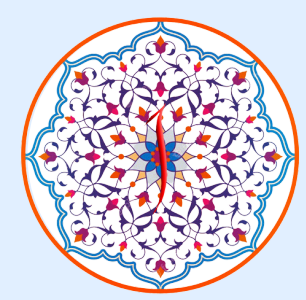
# What is Statistics?

## What is Statistics?

**Statistics is the study of collecting, analyzing, interpreting, and presenting data. It allows us to make sense of data and draw conclusions from it.**

### **Two Main Areas of Statistics:**

- 1. Descriptive Statistics: Summarizing and organizing data to describe its key features.**
- 2. Inferential Statistics: Drawing conclusions and making predictions about a population based on a sample of data.**



# What is Statistics?

## What is Statistics?

### Descriptive Statistics Basics

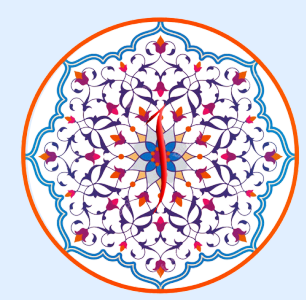
- **Mean (Average):**
- 

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

**Example: The mean of the numbers**

$$\frac{2 + 4 + 6}{3} = 4$$

- **Median: The middle value when the data is ordered. Example: The median of 1,3,7,8,9 is 7 (the middle number).**



# What is Statistics?

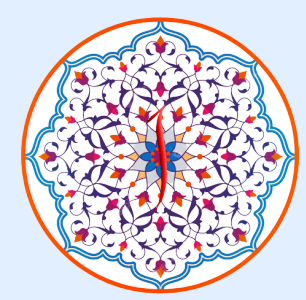
## What is Statistics?

- **Variance: A measure of how much the data values differ from the mean.**

$$\text{Variance} = \frac{\sum (x_i - \text{mean})^2}{n}$$

- **Standard Deviation: The square root of the variance, indicating how spread out the data is.**

$$\text{Standard Deviation} = \sqrt{\text{Variance}}$$



ikSaan.com

# What is Statistics?

## What is Statistics?

### Inferential Statistics Basics

Inferential statistics helps us make predictions or inferences about a population based on a sample of data.

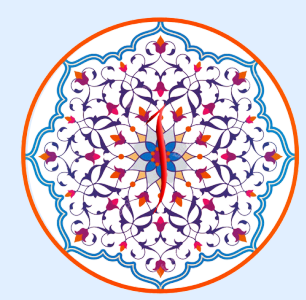
#### Key Concepts:

##### 1. Population vs. Sample:

- **Population:** The entire group of individuals or items we want to study.
- **Sample:** A subset of the population used to make inferences.

##### 1. Hypothesis Testing: A method to test if an assumption about a population is true, using sample data.

- **Null Hypothesis ( $H_0$ ):** The default assumption (e.g., no effect or no difference).
- **Alternative Hypothesis ( $H_A$ ):** The assumption we want to test (e.g., there is an effect or difference).



# What is Statistics?

## What is Statistics?

### Probability Distributions

In statistics and machine learning, understanding probability distributions is critical for analyzing random variables and data behavior.

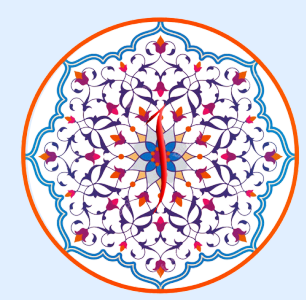
#### Types of Distributions:

**1. Discrete Probability Distributions:** Deals with variables that take specific, distinct values.

- **Example: Binomial Distribution** (number of successes in a fixed number of trials).

**2. Continuous Probability Distributions:** Deals with variables that can take any value within a range.

- **Example: Normal Distribution** (bell-shaped curve, common in nature and machine learning).



**ikSaan.com**

# What is Statistics?

## What is Statistics?

### **Normal Distribution (Gaussian Distribution):**

**The normal distribution is important in AI and statistics because many natural phenomena follow this pattern.**

- **It is symmetric around the mean.**
- **Most of the data falls within 1 standard deviation from the mean.**

### **Example:**

**In a normal distribution with mean 50 and standard deviation 5:**

- **About 68% of the data falls within 45 to 55 (1 standard deviation).**
- **About 95% of the data falls within 40 to 60 (2 standard deviations).**