

## Measures of Central Tendency (Statistics in Data Science)

---

### Introduction

#### Professional Definition

**Measures of Central Tendency are statistical measures used to determine the central or representative value of a dataset.**

They summarize the entire dataset into a single value which represents the center point.

---

#### Simple Explanation (Classroom Language)

Central tendency means:

 **Data ka average ya center value**

It tells:

**Data mostly kis value ke around hai**

---

### Real Life Example

Marks of students:

40, 50, 60, 70, 80

Central value = 60

This represents overall performance.

---

### Types of Central Tendency

There are mainly 3 types:

#### **Measure Meaning**

Mean     Average

Median   Middle value

Mode     Most frequent value

---

### MEAN (Average)

---

#### Professional Definition

**Mean is the sum of all observations divided by total number of observations.**

---

### **Formula**

$$Mean = \frac{\Sigma X}{N}$$

Where:

$\Sigma X$  = sum of values

N = number of values

---

### **Example**

Data:

10, 20, 30, 40, 50

Mean:

Sum = 150

Mean = 150 / 5

Mean = 30

---

### **Real Life Example**

Salary of employees:

20000, 25000, 30000

Mean salary = average salary

Used in:

- Business
  - Data Science
  - Machine Learning
- 

### **Limitation**

Mean affected by extreme values (Outliers)

Example:

10, 20, 30, 40, 1000

Mean becomes misleading

---

## MEDIAN

---

### Professional Definition

**Median is the middle value when data is arranged in ascending or descending order.**

---

### Example

10, 20, 30, 40, 50

Median = 30

---

### **Even Data Case**

10, 20, 30, 40

Median:

$20 + 30 / 2$

= 25

---

### Real Life Example

Used in:

Salary analysis

Because median not affected by outliers.

---

Example:

20000, 25000, 30000, 500000

Median gives real picture.

---

## MODE

---

### Professional Definition

**Mode is the value that appears most frequently in the dataset.**

---

### Example

10, 20, 20, 30, 40

Mode = 20

---

### Real Life Example

Used in:

Customer preference analysis

Example:

Most sold product

---

### Comparison Table

Measure Definition		Use Case
Mean	Average	General analysis
Median	Middle value	Skewed data
Mode	Most frequent	Category data

---

### Outlier Effect (Important for Masters)

Data:

10, 20, 30, 40, 500

#### Measure Value

Mean    affected

Median   stable

Mode    no effect

---

### Conclusion

Mean → Sensitive

Median → Robust

Mode → Frequency based