📘 Day 3 — Central Tendency

🌟 Why Central Tendency?

When we have a dataset, we often want to answer:

👉 "What is a typical value here?"

For example, if I ask **"How much do students score in SSC?"**, you won't read me 100 marks one by one — you'll give me one **representative number**.

That's what central tendency gives us.

There are **3 tools**:

1. **Mean (average)**
2. **Median (middle)**
3. **Mode (most frequent)**

1️⃣ Mean (Average)

📌 **Formula**

$$x^- = Sum\ of\ observations Number\ of\ observations \bar{x} = \frac{\text{Sum of observations}}{\text{Number of observations}}$$

📖 **Example**:

Marks = 91, 81, 92, 89, 90, 94

$$x^- = 91 + 81 + 92 + 89 + 90 + 946 = 5376 = 89.5 \bar{x} = \frac{91 + 81 + 92 + 89 + 90 + 94}{6} = \frac{537}{6} = 89.5$$

👉 **Interpretation**: The average student scored ~90.

💡 **Strength**: Uses all values.

⚠️ **Weakness**: Very sensitive to outliers.

2️⃣ Median (Middle Value)

📌 **Steps**

1. Sort the data.
2. If odd count → middle element.
3. If even count → average of 2 middle elements.

📖 **Example 1 (Odd count)**

Data: 1, 5, 20, 21, 16, 17, 3

Sorted → 1, 3, 5, 16, 17, 20, 21

Median = **16**

📖 **Example 2 (Even count)**

Data: 1, 5, 20, 21, 16, 17, 3, 7

Sorted → 1, 3, 5, 7, 16, 17, 20, 21

Median = (7+16)/2 = **11.5**

👉 **Interpretation**: Half of values are below, half above.

💡 **Strength**: Not affected by outliers.

3️⃣ Mode (Most Frequent)

📌 **Definition**

The value that occurs most often.

📖 **Example**

Data = 10, 15, 20, 20, 25, 30, 20

Mode = **20** (appears 3 times).

👉 **Best for** categorical data:

- "Most common blood group?"
- "Most sold pizza flavor?"

⚖️ Mean vs Median (Impact of Outliers)

📖 **Case 1 – Balanced Salaries**

50K, 75K, 1L, 2L

- Mean = 1.06L
- Median = 87.5K

  👉 Both give fair idea.

📖 **Case 2 – With Outlier**

0.5 Paise, 50K, 1L, 1000 Cr

- Mean = ~200 Cr 🤯
- Median = 1L ✅

  👉 Median is more reliable when extreme values exist.

🔑 Summary Table

Measure	Formula/Logic	Good For	Weakness
Mean	Sum ÷ Count	Balanced, normal data	Affected by outliers
Median	Middle value	Skewed data, income, property prices	Ignores exact magnitudes
Mode	Most frequent	Categories, popularity	Can be multiple / not exist

💡 Extra Insights (for deeper understanding)

1. **Multiple Modes**:
   - 1 peak → **Unimodal**
   - 2 peaks → **Bimodal**
   - Many peaks → **Multimodal**
2. **Link with Shape**:
   - In **normal distribution** → Mean = Median = Mode.
   - In **skewed data** → they spread apart (we'll explore in Day 4).
3. **Why we need all three**:
   - **Mean** tells "mathematical average."
   - **Median** tells "middle typical person."
   - **Mode** tells "most common occurrence."

     👉 Together, they give a **full picture** of the data.

🏋️ Practice Problems

1. Data: {5, 7, 8, 10, 10, 15, 20}
   - Find Mean, Median, Mode.
2. Salaries (in ₹000): {20, 22, 25, 28, 30, 90}
   - Compute Mean & Median. Which represents the data better?
3. Survey results: {Poor, Good, Excellent, Good, Fair, Good, Excellent}
   - Which central tendency measure is appropriate?