📘 Day 5 – Deviation & Measures of Spread

1️⃣ Deviation – Why does total deviation = 0?

👉 **Deviation** = how far a data point is from the mean.

Formula for each value:

$d_i = x_i - \bar{x}$

- If x_i > mean → deviation is **positive**.
- If x_i < mean → deviation is **negative**.

Example: Data = {1, 2, 3, 4, 5}, Mean = 3

Deviations = (-2, -1, 0, +1, +2).

Sum = -2 -1 + 0 + 1 + 2 = **0** ✅

💡 **Reason:** The mean is the balancing point of data, so deviations cancel.

👉 That's why we **don't use plain deviation** as a measure of spread.

2️⃣ Absolute Mean Deviation (AMD or MAD)

👉 To avoid negative + positive canceling, we take **absolute value**:

$MAD = \frac{1}{n} \sum |x_i - \bar{x}|$

Example: Data = {1, 2, 3, 4, 5}, Mean = 3

- |1-3| = 2
- |2-3| = 1
- |3-3| = 0
- |4-3| = 1
- |5-3| = 2

Sum = 6 → MAD = 6/5 = **1.2**

👉 Interpretation: On average, data points are **1.2 units away from mean**.

⚠️ Math note: Absolute value is hard in calculus (nondifferentiable at 0), so we prefer squaring → leads us to variance.

3️⃣ Variance

👉 To avoid cancellation and make math smooth → **square deviations**.

$$\sigma^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

(for population variance).

For sample variance (statistical estimation):

$$s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

Example: Data = {1, 2, 3, 4, 5}, Mean = 3

Deviations = (-2, -1, 0, 1, 2)

Squares = (4, 1, 0, 1, 4)

Sum = 10

- Population variance = 10/5 = **2**
- Sample variance = 10/4 = **2.5**

👉 Variance measures **average squared spread** of data.

⚠️ But it's in **squared units** (marks², km², rupees²). Hard to interpret directly.

4️⃣ Standard Deviation (SD)

👉 To bring back original units → take square root of variance.

$$\sigma = \sqrt{\sigma^2}, \quad s = \sqrt{s^2}$$

From above example:

- Population SD = √2 ≈ **1.41**
- Sample SD = √2.5 ≈ **1.58**

👉 Interpretation: On average, data points are ~1.4 units away from mean.

📌 SD is the **most widely used measure of spread**.

5️⃣ Scaling Effect (Important Concept)

If you multiply data by a constant cc:

- Mean → ×c
- Variance → ×c2c^2
- SD → ×c

Example:

Data = {2, 3, 4, 5, 6} km

Multiply by 10 → {20, 30, 40, 50, 60} rupees

- Variance (km) = 2
- Variance (rs) = 200 = 100 × bigger
- SD (km) = 1.41
- SD (rs) = 14.1 = 10 × bigger

👉 This is why variance has **squared units**.

6️⃣ Coefficient of Variation (CV)

👉 To compare spread across different units/scales, use **CV** (unitless).

CV=SDMean×100CV = \frac{SD}{Mean} \times 100

Example:

- For km → CV ≈ 35.35%
- For rupees → CV ≈ 35.35% (same, scale doesn't matter).

👉 CV helps compare consistency across different measurements.

7️⃣ Practical Summary

- **Deviation**: cancels to 0 (not useful).
- **MAD**: uses absolute deviation, easy to understand but less used in advanced math.
- **Variance**: squared average deviation, math-friendly but hard units.
- **SD**: square root of variance, best real-world measure of spread.
- **CV**: compares spread across different scales, unitless.

📝 Practice Problems

1. Data = {2, 4, 6, 8, 10}
    - Find Mean, MAD, Variance, SD.
2. A company's monthly profits (in lakh ₹): {20, 22, 25, 23, 100}
    - Find mean and SD. Which is more reliable, mean or median, and why?
3. Two batsmen scored:
    - Player A: {40, 50, 60, 70, 80}
    - Player B: {10, 30, 50, 70, 90}
    - Who is more consistent? (Use SD).
