🌟 1. Introduction to Measures of Spread

🎯 Why Do We Measure Spread?

When we look at a set of data, it's not just about the **average** — we also want to know how **spread out** the data is.
Are most of the values close together? ✊
Or are they far apart? 🧭

This is where **measures of spread** come in — they help us understand the **variability** in data. 📊

🔍 What Are Measures of Spread?

Measures of spread give us important information about **how data values differ from each other**.

They help answer questions like:

- 📏 How wide is the range of scores?
- ⚖️ Are there outliers (unusual values)?
- 📉 Is the data tightly clustered or scattered?

📚 Common Measures of Spread

Here are three key tools you'll use to measure spread:

1. 📐 **The Range –** The simplest measure, showing the difference between the highest and lowest values.
2. 📊 **The Interquartile Range (IQR) –** Measures the spread of the middle 50% of data, less affected by outliers.
3. 📈 **The Standard Deviation –** Tells us how much the data varies around the **mean**.

✌️ 2. The Range

📘 What is the Range?

The **range** is the **simplest measure of spread**.
It tells us how far apart the smallest and largest values in a dataset are.

👉 It gives a quick sense of how **wide** the data is spread.

🧮 Formula for the Range

$$\boxed{R = \text{largest value} - \text{smallest value}}$$

Just subtract the smallest number from the largest, and you've got it!
🟢 **Large range** = More spread
🔵 **Small range** = Less spread

🔍 Worked Example (Example 14): Finding the Range

Let's look at marks from **two different tasks** given to a group of students:

📝 **Task A Marks:**

2, 6, 9, 10, ..., **94**

✅ **Smallest value** = 2
✅ **Largest value** = 94

🧠 **Range for Task A**:

$$R = 94 - 2 = \boxed{92}$$

📝 **Task B Marks:**

11, 16, 19, ..., **91**

✅ **Smallest value** = 11
✅ **Largest value** = 91

🧠 **Range for Task B**:

$$R = 91 - 11 = \boxed{80}$$

📊 Interpreting the Results

- Task A has a **larger range** (92) compared to Task B (80).
- At first glance, this suggests Task A's scores are **more spread out**.
- BUT... is that the whole story? 🤔
  (Stay tuned for Section 3 to find out why **range** might not always **tell** the full truth! 📏🔍)

⚠️ 3. Limitations of the Range

🚨 The Problem with the Range

While the **range** is super easy to calculate, it's not always **reliable**.
Why? Because it can be heavily affected by **outliers** — values that are **much higher or lower** than the rest.

🔍 Just one unusual number can make the range look **way bigger** than it should be!

🌿 Visual Check: Stem-and-Leaf Plots

Let's compare Task A and Task B using **stem-and-leaf plots**:

📘 Task A

```
0 | 2 6 9
1 | 0 1 2 3
2 | 3 4 6 6 7
3 | 4 5 8 8 9
4 | 2 6 7
5 | 2 2 6 6 9
6 | 0
7 |
8 |
9 | 1 4
```

📗 Task B

```
0 |
1 | 1 6 9
2 | 1 3 8
3 | 1 1 3 8
4 | 1 9
5 | 2 3 4 6 9
6 | 3 5 8
7 | 1 2 3 5 8 8
```

```
8 | 6 8
9 | 1
```

👁 What Do We See?

- Task A has **two very high scores** (91 and 94) that stretch the range. 🎯
- Most of Task A's scores are **clustered in the middle**, with just a couple of outliers.
- Task B's scores are **more evenly spread** — the range reflects the distribution **a bit more accurately**.

🧠 Conclusion: We Need a Better Tool!

The **range** doesn't always give us the full picture 📉.
We need a measure that:

- Focuses on the **middle values**
- Ignores extreme outliers 🚫

👑 That's where the **Interquartile Range (IQR)** comes in — get ready for it in the next section! 📦 ✨

📦 4. The Interquartile Range (IQR)

🧾 Definition

The **interquartile range (IQR)** tells us how spread out the **middle 50%** of the data is.
Unlike the range, it ignores the **lowest 25%** and **highest 25%** of values.

🎯 It's a **more reliable** measure of spread, especially when outliers are present!

🧮 Formula for IQR

$$\boxed{\text{IQR} = Q_3 - Q_1}$$

Where:

- Q_1 = First Quartile (25th percentile)
- Q_3 = Third Quartile (75th percentile)

🔢 Steps to Find the IQR

1. 🔢 **Order** the data from smallest to largest
2. ✂️ **Split** the data into two equal halves
   - Lower half (first 50%)
   - Upper half (last 50%)
3. 🎯 **Find** Q_1 = median of the lower half
4. 🎯 **Find** Q_3 = median of the upper half
5. ➖ **Subtract**:

$$\text{IQR} = Q_3 - Q_1$$

📘 Worked Example (Example 15): Finding the IQR

Let's find the IQR for **Task A** and **Task B**:

📝 **Task A Data**

- 30 total values
- Split into:
  - **Lower half:** 2, 6, 9, 10, 11, 12, 13, **22**, 23, 24, 26, 26, 27, 33, 34
    - $Q_1 = 22$
  - **Upper half:** 35, 38, 38, 39, 42, 46, **47**, 47, 52, 52, 56, 56, 59, 91, 94
    - $Q_3 = 47$

🧮 IQR for Task A:

$$IQR = 47 - 22 = \boxed{25}$$

📝 **Task B Data**

- 30 total values
- Split into:
  - **Lower half:** 11, 16, 19, 21, 23, 28, **31**, 31, 33, 38, 38, 41, 49, 52, 53
    - $Q_1 = 31$
  - **Upper half:** 54, 56, 59, 63, 65, 68, **71**, 72, 73, 75, 78, 78, 86, 88, 91
    - $Q_3 = 73$

🧮 IQR for Task B:

$$IQR = 73 - 31 = \boxed{42}$$

📊 Comparison & Conclusion

- ✅ Task A IQR = 25 → Scores are more **concentrated**
- ✅ Task B IQR = 42 → Scores are more **spread out**

🧠 What it means:

> The IQR shows that Task A's data is less variable than Task B's data.
> It's a clearer picture of the real spread—without being tricked by outliers! 🎯

📉 5. The Standard Deviation

📘 What Is the Standard Deviation?

The **standard deviation (s)** is a powerful measure of spread that tells us **how much data values deviate from the mean** (average).

If values are close to the mean ➡️ 📏 **Small standard deviation**
If values vary a lot from the mean ➡️ 🌪️ **Large standard deviation**

🧮 Formula

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

Where:

- x = individual data value
- $\bar{x}$ = mean (average) of the data
- n = number of values
- $\sum$ = sum of all the values
- s = sample standard deviation

🧠 Breaking It Down

🔹 **1.** $\bar{x}$ **– The Mean**

Find the average of all your values:

$$\bar{x} = \frac{\sum x}{n}$$

🔹 **2.** $x - \bar{x}$ **– Deviation from the Mean**

For each value, find how far it is from the mean.

🔹 **3.** $(x - \bar{x})^2$ **– Squared Deviations**

Square each deviation so that positive and negative differences don't cancel out. 📐

🔹 **4.** $\sum(x - \bar{x})^2$ **– Total Variation**

Add all the squared deviations together.

◆ **5. Divide by $n - 1$**

This is called **Bessel's correction —** it adjusts for the fact that we're working with a **sample** (not the whole population).
✅ It helps get a more accurate estimate of variability.

◆ **6. Take the Square Root**

This brings the units back to the original scale of the data.

❗ Why Use $n - 1$ Instead of n?

Because we're often working with a **sample**, not the full population.
Using $n - 1$ prevents us from **underestimating** the spread.
📊 It corrects the bias and gives a **more accurate picture** of how variable the data really is.

✅ So, while the **range** and **IQR** give you a sense of spread...
📉 **Standard deviation** gives you a **precise, mathematical** look at how data clusters around the mean!

🧠 6. Worked Example: Calculating the Standard Deviation

Let's calculate the **sample standard deviation** for this simple data set:

🔢 Data Set:

$$\{2, 3, 4\}$$

🧮 Step 1: Find the Mean $\bar{x}$

$$\bar{x} = \frac{2+3+4}{3} = \frac{9}{3} = 3$$

✅ The mean is 3

📋 Step 2: Create a Table

x	$x - \bar{x}$	$(x - \bar{x})^2$
2	2 - 3 = -1	$(-1)^2 = 1$
3	3 - 3 = 0	$(0)^2 = 0$
4	4 - 3 = 1	$(1)^2 = 1$

🧮 Sum of squared deviations:

$$1 + 0 + 1 = 2$$

✅ Step 3: Apply the Formula

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} = \sqrt{\frac{2}{3-1}} = \sqrt{\frac{2}{2}} = \sqrt{1}$$

$$\boxed{s = 1}$$

🎉 Final Answer:

The **sample standard deviation** of the data set $\{2, 3, 4\}$ is:

$$\boxed{1}$$

📏 What Does Standard Deviation Tell Us?

The **standard deviation** is a number that tells us **how spread out the values in a dataset are** from the mean (average).

- ✅ **Small SD (e.g., SD = 1):**
  Most values are **close to the mean** → data is **tightly clustered**.
  → Example: Test scores where most students scored around 80.
- 🔄 **Larger SD (e.g., SD = 2 or more):**
  Values are **more spread out** → more **variation or inconsistency**.
  → Example: Some students scored 50, some 90, even though the average is still 70.

🧠 How to Interpret SD Values

Imagine a class of students with a mean test score of 70:

- 🎯 **SD = 1:** Most students scored **between 69 and 71** → very consistent performance
- 🌪️ **SD = 10:** Scores vary **widely**, from **60 to 80** and beyond → large performance gap

So, SD helps you understand:

- Are the results **similar** (low SD)?
- Or are they **all over the place** (high SD)?

🌍 Real-Life Examples of Standard Deviation

🧪 1. Education – Test Scores

Let's say two classes have the same **average** test score (75), but:

- **Class A** has **SD = 2** → Most students scored around 75
- **Class B** has **SD = 15** → Some students got 40, some got 100

✅ This helps a teacher identify which class had **more consistent learning outcomes**.

📈 2. Business – Sales Performance

A company tracks weekly sales across branches:

- **Branch X** has **SD = $200** → Stable sales each week
- **Branch Y** has **SD = $1,000** → Some weeks are great, some terrible

✅ Managers can use this to decide where to **invest training or resources** to stabilize performance.

📉 3. Health – Blood Pressure Readings

In a clinical trial, if one group has blood pressure with **low SD**, the treatment is **working consistently**.
But if the SD is high, it suggests the treatment has **varying effects** — it might not be reliable.

🧮 4. Manufacturing – Quality Control

If a machine produces metal rods with length = 10cm:

- **SD = 0.1cm** → Most rods are between 9.9 and 10.1cm → Excellent consistency
- **SD = 0.5cm** → Lots of variation → May cause issues with product quality

✅ Lower SD = more **precision and quality**

🎓 Key Takeaway

> **Standard deviation is a tool to understand consistency.**
> **The smaller the SD, the more predictable and reliable the data.**
> **The larger the SD, the more uncertainty or variation there is.**