PART 1️⃣ Measure of Dispersion

📘 Professional Definition

Measure of Dispersion is a statistical measure used to describe the spread or variability of data values around a central value.

📗 Simple Definition (Classroom Language)

Dispersion batata hai:

Data kitna faila hua hai (spread)

Central tendency → center batata hai
Dispersion → spread batata hai

🎯 Why Dispersion is Needed

Suppose 2 companies salary:

Company A:

20k, 25k, 30k, 35k, 40k

Company B:

10k, 20k, 30k, 40k, 100k

Mean same ho sakta hai, but salary variation different hai.

Dispersion helps understand this difference.

📊 Types of Measure of Dispersion

1. Range
2. Variance
3. Standard Deviation

1️⃣ Range

📘 Professional Definition

Range is the difference between the maximum and minimum value in dataset

---

📗 Formula

$$Range = Maximum - Minimum$$

🧮 Example (Maths Sum)

Data:

10, 15, 20, 25, 30

Solution:

Maximum = 30
Minimum = 10

Range:

$$Range = 30 - 10 = 20$$

🎯 Real Life Use

Used in:

Temperature analysis

Stock price analysis

💻 Python

```
import numpy as np

data=[10,15,20,25,30]

print(max(data)-min(data))
```

2️⃣ Variance

📘 Professional Definition

Variance is the average of squared deviations from mean

📗 Formula

$$Variance = \frac{\sum(X - Mean)^2}{N}$$

🧮 Example

Data:

10, 20, 30

Step 1: Mean

$$Mean = (10 + 20 + 30)/3 = 20$$

Step 2: Square difference

Value	Difference	Square
10	-10	100
20	0	0
30	10	100

Sum = 200

Variance:

$$200/3 = 66.67$$

🎯 Real Life Use

Used in:

Stock market risk

3️⃣ Standard Deviation

📘 Professional Definition

Standard deviation is square root of variance

📗 Formula

$$SD = \sqrt{Variance}$$

🧮 Example

Variance = 66.67

SD:

$$\sqrt{66.67} = 8.16$$

🎯 Real Life Use

Used in:

Salary analysis

Risk analysis

💻 Python

```
import numpy as np

data=[10,20,30]

print(np.std(data))
```

PART 2️⃣ Percentage

📘 Professional Definition

Percentage is a number expressed as fraction of 100

📗 Formula

$$Percentage = \frac{Value}{Total} \times 100$$

🧮 Example

Marks = 45

Total = 50

$$Percentage = (45/50) \times 100$$

= 90

🎯 **Real Life Use**

Student marks

Profit calculation

PART 3️⃣ Percentile

📘 **Professional Definition**

Percentile is the value below which certain percentage of observations fall

---

📗 **Simple Explanation**

Percentile tells:

Position of value in dataset

🧮 **Example**

Data:

10, 20, 30, 40, 50

50 percentile = Median = 30

🎯 **Real Life Use**

Used in:

Competitive exams

GRE, CAT

💻 **Python**

```
import numpy as np

data=[10,20,30,40,50]

print(np.percentile(data,50))
```

PART 5️⃣ Five Number Summary

📘 Professional Definition

Five number summary is set of five values that describe dataset distribution

---

Five values

1 Minimum
2 Q1
3 Median
4 Q3
5 Maximum

📗 Formula

Q1 = 25 percentile

Median = 50 percentile

Q3 = 75 percentile

🧮 Example

Data:

10, 20, 30, 40, 50

Minimum = 10

Q1 = 20

Median = 30

Q3 = 40

Maximum = 50

💻 Python

```
import numpy as np

data=[10,20,30,40,50]

print(np.min(data))
```

```
print(np.percentile(data,25))

print(np.median(data))

print(np.percentile(data,75))

print(np.max(data))
```

PART 5️⃣ OUTLIERS (VERY IMPORTANT)

📘 Professional Definition

Outlier is an extreme value that differs significantly from other values

---

📗 Example

Data:

10, 20, 30, 40, 500

Here:

500 is outlier

📗 Why Outlier Important

Because it affects:

Mean

Standard deviation

📗 Outlier Formula (IQR Method)

$$IQR = Q3 - Q1$$

Outlier:

$$Lower = Q1 - 1.5(IQR)$$
$$Upper = Q3 + 1.5(IQR)$$

🧮 Example

Q1 = 20

Q3 = 40

IQR:

$$40 - 20 = 20$$

Upper:

$$40 + 30 = 70$$

Value above 70 is outlier

💻 **Python Outlier Detection**

```
import numpy as np

data=[10,20,30,40,500]

Q1=np.percentile(data,25)

Q3=np.percentile(data,75)

IQR=Q3-Q1

upper=Q3+1.5*IQR

print(upper)
```

🎯 **Data Science Use**

Dispersion used in:

EDA

Feature engineering

Machine Learning

Outlier removal

Real Life Example

Fraud detection

Outliers detect fraud transactions

⭐ Summary Table

Measure	Formula
Range	Max-Min
Variance	$\Sigma(X\text{-mean})^2/N$
SD	√Variance
Percentage	Value/Total ×100
Percentile	Position
IQR	Q3-Q1