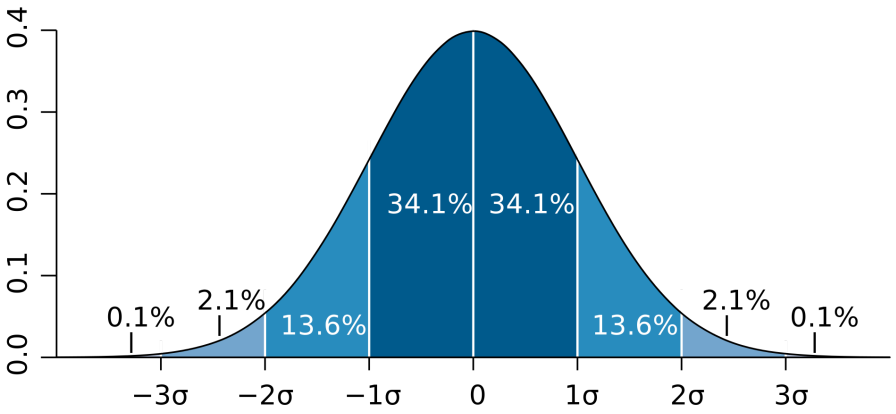
📘 Lecture 2: Measures of Shape

🔶 What Are Measures of Shape?

In descriptive statistics, we don't just look at **averages (central tendency)** or **spread (dispersion)**—we also care about the **shape** of the data distribution.



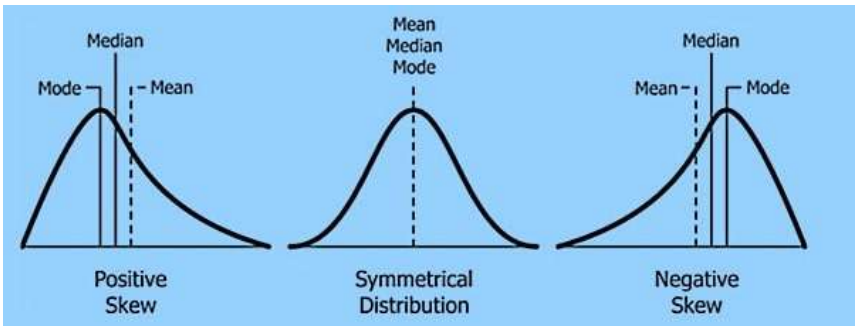
🔷 Why Are These Measures Important?

1. **Understanding Real-World Behavior**
2. **Choosing the Right Statistical Methods**
3. **Detecting Outliers and Anomalies**

🔶 Skewness?

Skewness is a statistical measure that describes the **asymmetry** of a distribution around its mean.

- When a distribution is **symmetrical**, the left and right sides are mirror images.
- When it is **skewed**, one tail is **longer or fatter** than the other.
- Skewness tells us the **direction** and **degree** of the skew.



🔷 Types of Skewness

1️⃣ Symmetrical Distribution (Zero Skewness)

- **Mean ≈ Median ≈ Mode**
- Data is evenly distributed around the center.

🧪 **Example:**

```
Data: 60, 65, 70, 70 , 71, 74, 80

Mean = 70, Median = 70, Mode = 70

Skewness ≈ 0 (symmetrical)
```

2️⃣ Positively Skewed (Right Skewness)

- **Mean > Median > Mode**
- Tail is stretched to the **right** (toward higher values)
- Bulk of data lies on the **left**

🧪 **Example:**

```
Data: 60, 65, 70, 70 , 71, 74, 108
Mean = 74, Median = 70, Mode = 70
Skewness > 0 (positive)
```

- Indicates a **few high outliers**
- Common in **income distributions**, **CO_2 emissions**
- Mean is "pulled" right by extreme values

3️⃣ Negatively Skewed (Left Skewness)

- **Mean < Median < Mode**
- Tail is stretched to the **left** (toward lower values)
- Bulk of data lies on the **right**

🧪 **Example:**

```
Data: 32, 65, 70, 70 , 71, 74, 80
Mean = 66, Median = 70, Mode = 70
Skewness < 0 (negative)
```

- Indicates **few low outliers**
- Can occur in datasets like **exam scores** where most students score high, but a few score very low
- Mean is "pulled" left by lower extreme values

🧮 Mathematical Definition

The skewness coefficient is often calculated as:

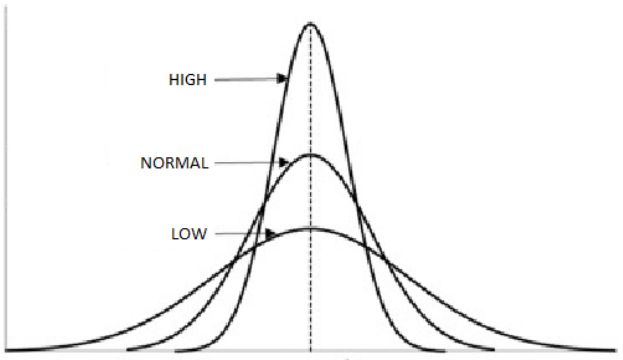
$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X_i - \bar{X}}{s} \right)^3$$

Where:

- X_i = each observation
- $\bar{X}$ = mean of the data
- s = standard deviation
- n = number of observations

🔶 What Is Kurtosis?

Kurtosis measures the **"peakedness"** or **flatness** of a data distribution compared to a normal (bell-shaped) distribution.



🔷 Types of Kurtosis

There are **three main types**:

1️⃣ Normal Kurtosis

- **Reference shape**: Normal distribution
- Moderate **peak**, moderate **tails**

2️⃣ High Kurtosis

- **Tall, sharp peak**

- **Heavy tails**
- Most values are **tightly clustered** near the mean
- A few values are **extremely far** from the mean
- Indicates a **highly concentrated center** with **more extreme values**.

3️⃣ Low Kurtosis

- **Flat or broad peak**
- **Light tails**
- Data is **evenly spread**; there is **no strong central tendency**.

🧮 Mathematical Definition

$$\text{Kurtosis} = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{X_i - \bar{X}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

- **Excess Kurtosis** is often used:
  - $\text{Excess Kurtosis} = \text{Kurtosis} - 3$

Exercise

📊 For the Daily PM2.5 Concentration (μg/m³) in a City Over 15 Days

Day	PM2.5 Concentration (μg/m³)	Day	PM2.5 Concentration (μg/m³)
1	22	9	40
2	25	10	45
3	28	11	55
4	30	12	60
5	32	13	65
6	34	14	70
7	35	15	100
8	36		

Find:

1️⃣ Central Tendency

- Calculate the **Mean** PM2.5 value
- Find the **Median**
- Determine the **Mode** (if any)

2️⃣ Dispersion

- Find the **Range**
- Calculate the **Variance**
- Calculate the **Standard Deviation**

3️⃣ Shape Analysis

- **Skewness**
- **Kurtosis**: