Data Distribution Shape

The shape of a data distribution provides insight into how data values are spread and whether they deviate from the normal distribution.

Two important measures used to describe the distribution shape are:

- Skewness (asymmetry of the distribution)
- Kurtosis (peakedness or flatness of the distribution)

Understanding these concepts helps analysts decide on appropriate statistical models and detect outliers.

1. Skewness

Skewness measures the asymmetry of a data distribution.

- **Symmetrical distribution**: Mean ≈ Median ≈ Mode
- **Positively skewed** (right-skewed): Tail on the right is longer; Mean > Median > Mode
- **Negatively skewed** (left-skewed): Tail on the left is longer; Mean < Median < Mode

- Formula

Skewness = 3(Mean-Median) / SD

SD --> Standard Deviation

Where,

$$SD = \sqrt{\frac{\Sigma|x-\bar{x}|^2}{n}}$$

x = random variable
$\bar{x}$ = mean of the data
n = total no. of data

- Example

- Income distribution in most countries is positively skewed
  (A few very high incomes pull the mean to the right.)

- Exam scores with many high achievers but few low scores may be negatively skewed.

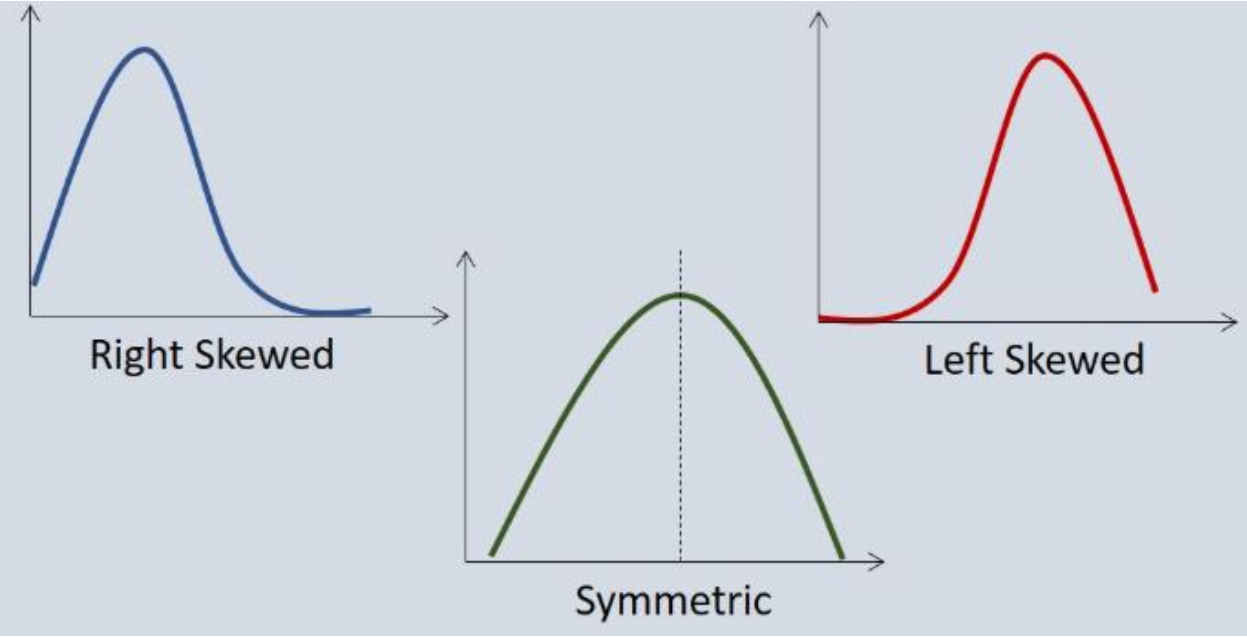
- Interpretation

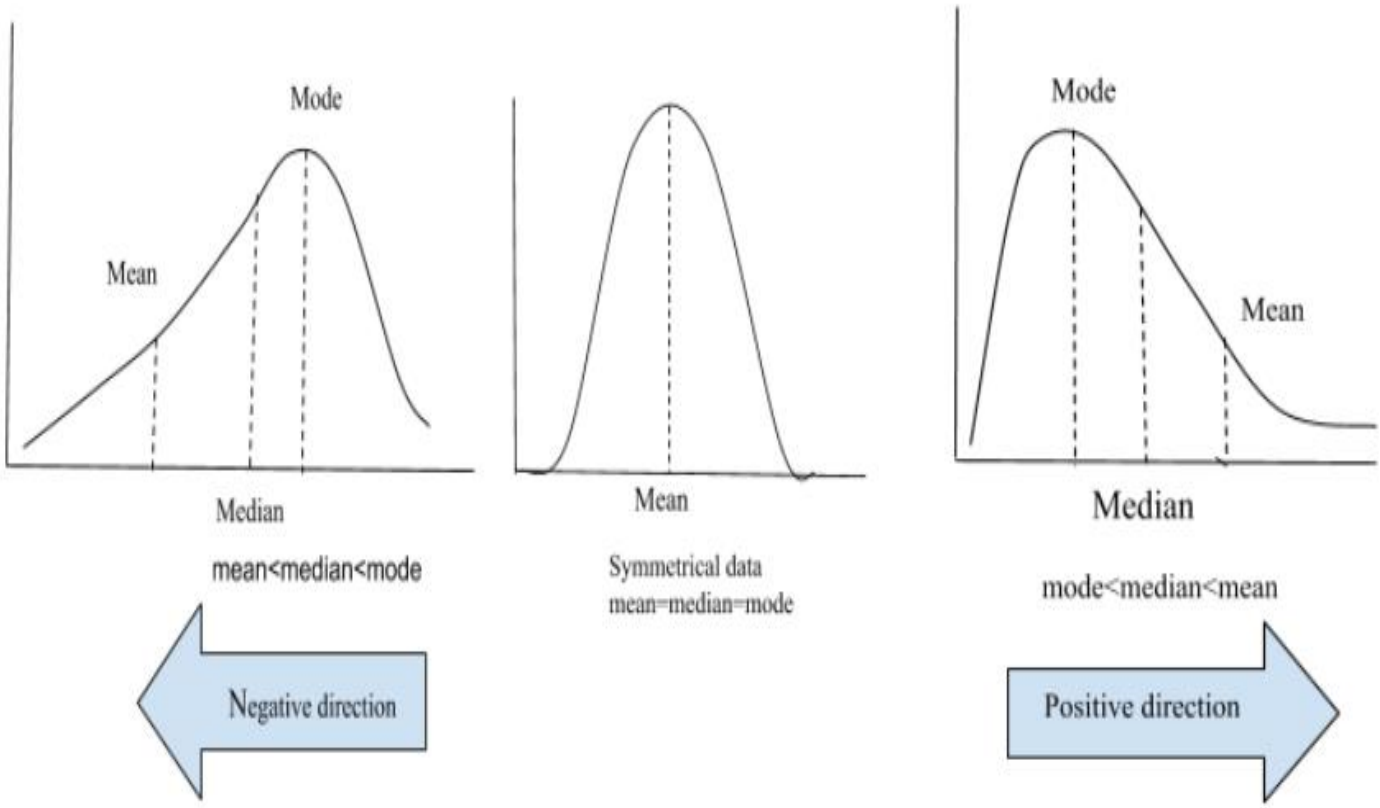
Skewness ≈ 0 → symmetric distribution.

Skewness > 0 → positively skewed.

Skewness < 0 → negatively skewed.

- Visual Examples





2. Kurtosis

Kurtosis measures the "tailedness" or peakedness of a distribution, compared to a normal distribution.

It helps to identify how prone the data is to outliers.

Types:

1. Mesokurtic (Normal distribution)
    - Kurtosis ≈ 3 (Excess Kurtosis = 0).
    - Standard bell curve shape.

2. Leptokurtic (Heavy tails, peaked)
    - Kurtosis > 3 (Excess Kurtosis > 0).
    - More values in the tails; higher chance of outliers.

3. Platykurtic (Flat distribution)
    - Kurtosis < 3 (Excess Kurtosis < 0).
    - Flatter than normal, fewer extreme values.

- Formula

$$\text{skewness} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^3}{(N-1)s^3}$$

where:

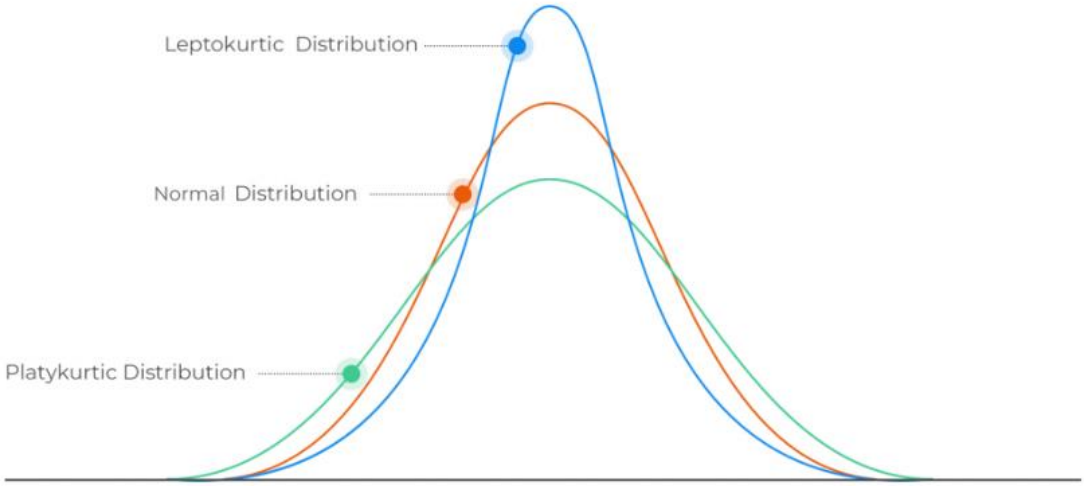
- s is the standard deviation
- $\overline{x}$ is the mean of the distribution
- N is the number of observations of the sample

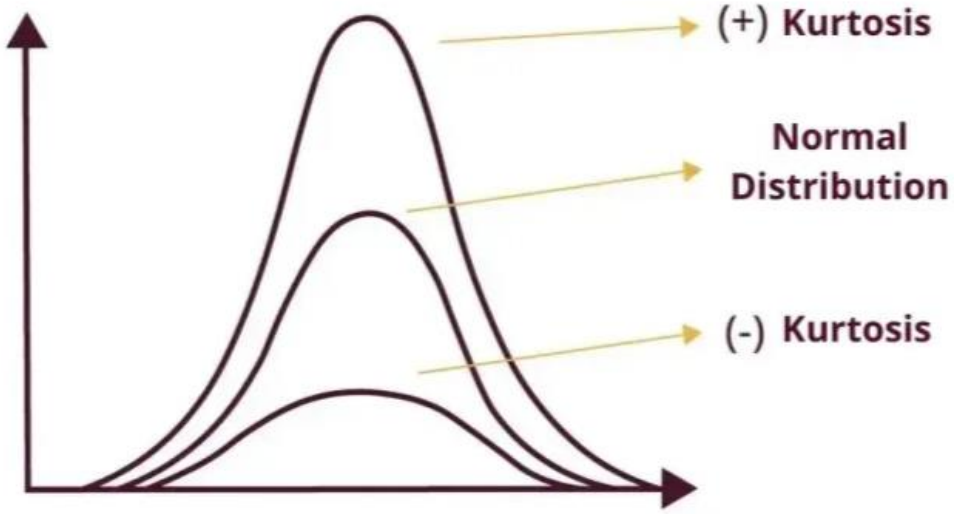
- Example

- Stock market returns often show leptokurtic behavior (heavy tails, extreme values more likely).

- Uniform-like data may be platykurtic.

- Visual Examples





3. Comparison

Property	**Skewness**	**Kurtosis**
Measures	Asymmetry of distribution	Tailedness / peakedness
Ideal value	0	3
Indicates	Direction of tail	Outlier likelihood
Example	Income distribution (positive skew)	Stock returns (leptokurtic)