

## **Theory & Definitions**

**Question 1** :- Define the following with examples from the dataset:

(a) Types of Data: Numerical & Categorical

(b) Types of Statistics: Descriptive vs. Inferential

(c) What is Descriptive Statistics?

**Answer 1** :-

(a) Types of Data: Numerical & Categorical

1. Numerical Data (Quantitative Data)

Numerical data refers to data that is expressed in numbers and can be measured or counted. Mathematical operations like addition and subtraction can be performed on this data.

**Examples from dataset :-**

- Marks scored (75, 82, 90)
- Height (170 cm, 165 cm)

Numerical data is of two types:

- **Discrete Data** – Countable values (number of students)
- **Continuous Data** – Measurable values (weight, height)

## 2. Categorical Data (Qualitative Data)

Categorical data refers to data that represents categories, labels, or groups.

### **Examples from dataset :-**

- Gender (Male, Female)
- Department (IT, Science, Commerce)

### (b) Types of Statistics: Descriptive vs. Inferential

#### 1. Descriptive Statistics :-

Descriptive statistics is used to summarize and describe the main features of a dataset.

It includes:

- Mean
- Median
- Mode
- Charts and graphs
- Standard deviation

#### 2. Inferential Statistics :-

Inferential statistics is used to make conclusions or make predictions about a population based on a sample.

It includes:

- Hypothesis testing
- Confidence intervals
- Regression analysis

### (c) What is Descriptive Statistics?

Descriptive statistics is the method of organizing, summarizing, and presenting data in a meaningful way to understand it easily.

It helps in:

- Finding average
- Measuring spread
- Creating graphs

It does not predict future results but explains existing data.

**Question 2** :- Explain the difference between:

(a) Mean, Median, Mode

(b) Range, Variance, Standard Deviation

**Answer 2 :-**

**(a) Difference between Mean, Median and Mode**

Mean :-

Mean is the average of all values.

Formula:

Mean = (Sum of all values) / (Total number of values)

Median :-

Median is the middle value when data is arranged in ascending or descending order.

Use when data contains outliers.

Mode :-

Mode is the value that appears most frequently.

Useful for categorical data.

### (b) Difference between Range, Variance and Standard Deviation :-

Range :-

Range is the difference between the highest and lowest value from a dataset.

Formula:

Range = Maximum – Minimum

Variance :-

Variance measures how far each value is from the mean.

Formula:

$$\sigma^2 = \sum (x - \mu)^2 / N$$

## Standard Deviation

Standard deviation is the square root of variance. It tells how spread out the data is.

Formula:

$$\sigma = \sqrt{\text{Variance}}$$

**Question 3** :- Explain the following terms with neat diagram and formula:

- Gaussian Distribution
- Log Normal Distribution
- 3-Sigma Rule (Empirical Rule)
- Percentiles
- Quartiles
- Five Number Summary
- Skewness
- Kurtosis

### **Answer 3 :-**

#### **1. Gaussian Distribution (Normal Distribution)**

Gaussian distribution is a bell-shaped symmetrical distribution where:

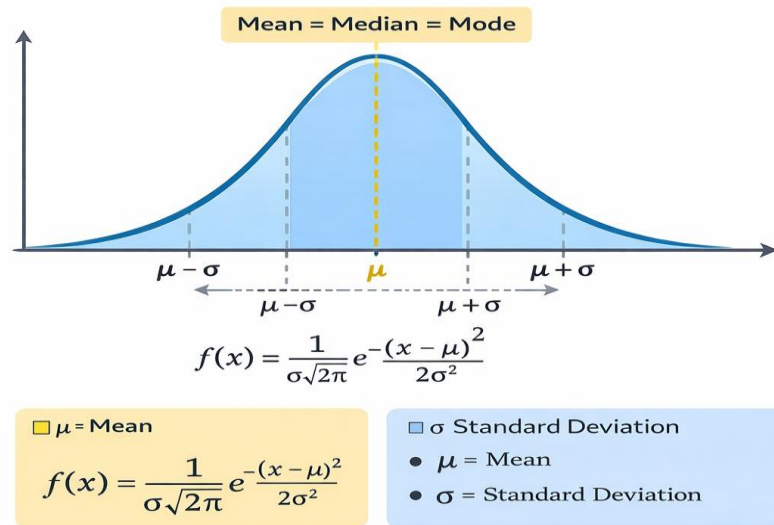
$$\text{Mean} = \text{Median} = \text{Mode}$$

Where:

$\mu$  = Mean

$\sigma$  = Standard deviation

## Gaussian Distribution (Normal Distribution)

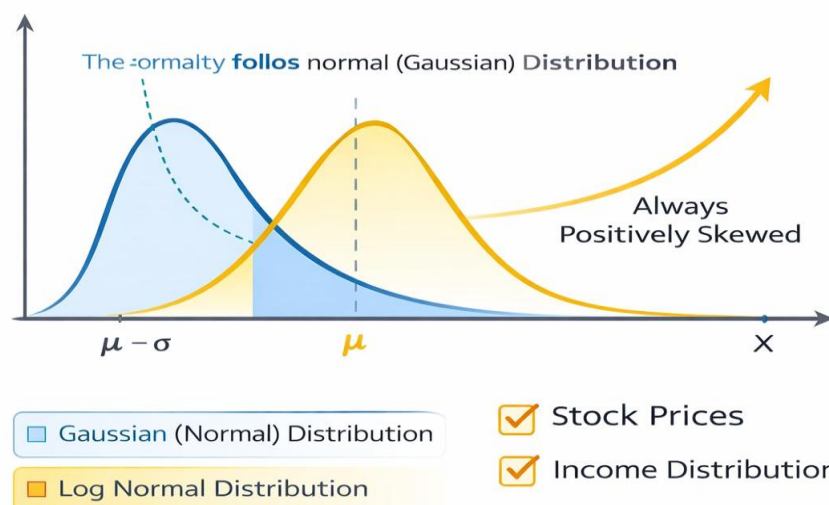


## 2. Log Normal Distribution

A distribution where the logarithm of the variable follows a normal distribution.

- It is positively skewed.
- Used in income distribution and stock prices.

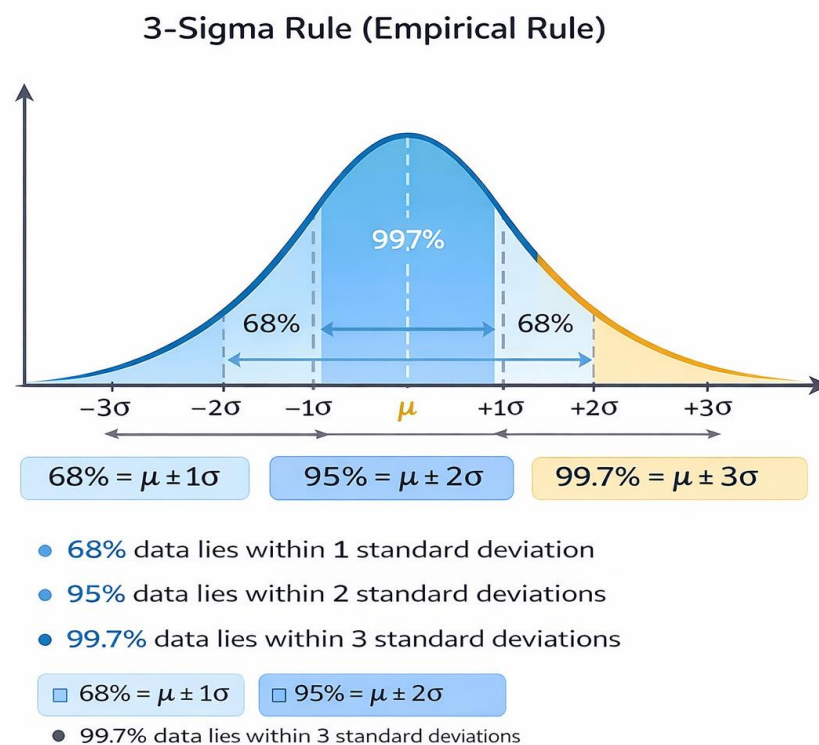
### Log Normal Distribution



### 3. 3-Sigma Rule (Empirical Rule)

In normal distribution:

- 68% of data lies within  $\pm 1$  standard deviation
- 95% lies within  $\pm 2$  standard deviations
- 99.7% lies within  $\pm 3$  standard deviations

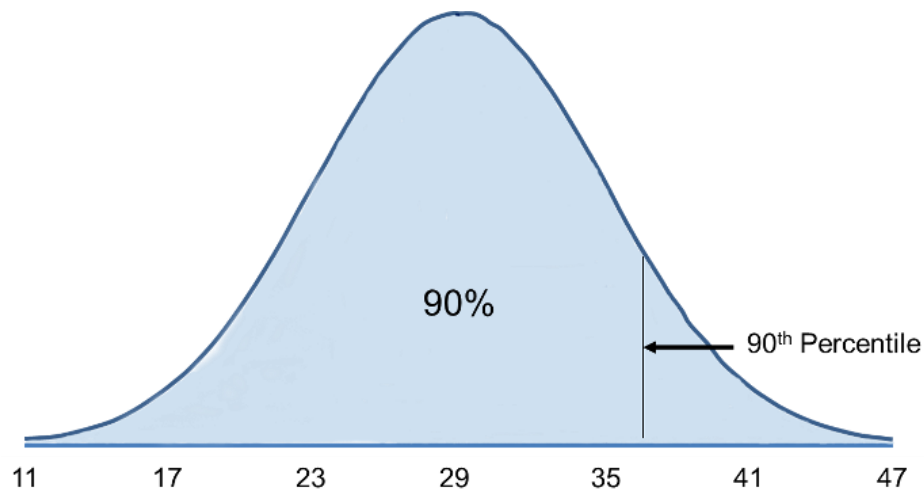


### 4. Percentiles

Percentiles divide data into 100 equal parts.

Example:

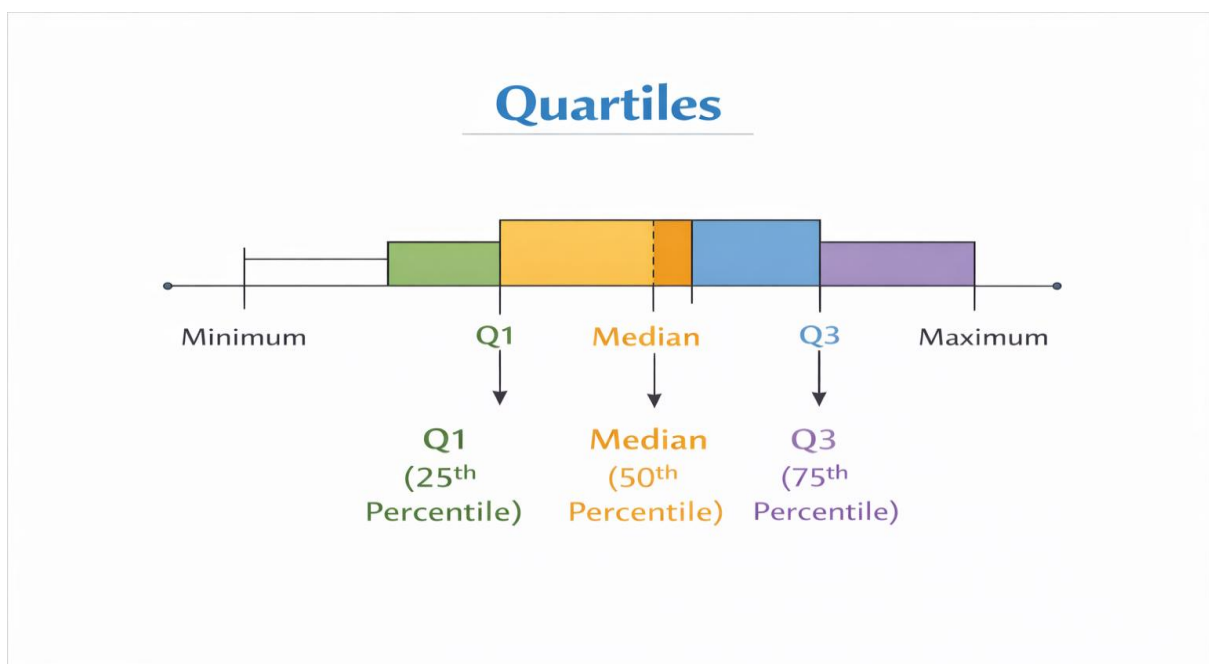
If a student is in the 90th percentile, it means he scored better than 90% of students.



## 5. Quartiles

Quartiles divide data into 4 equal parts:

- Q1 → 25%
- Q2 → 50% (Median)
- Q3 → 75%



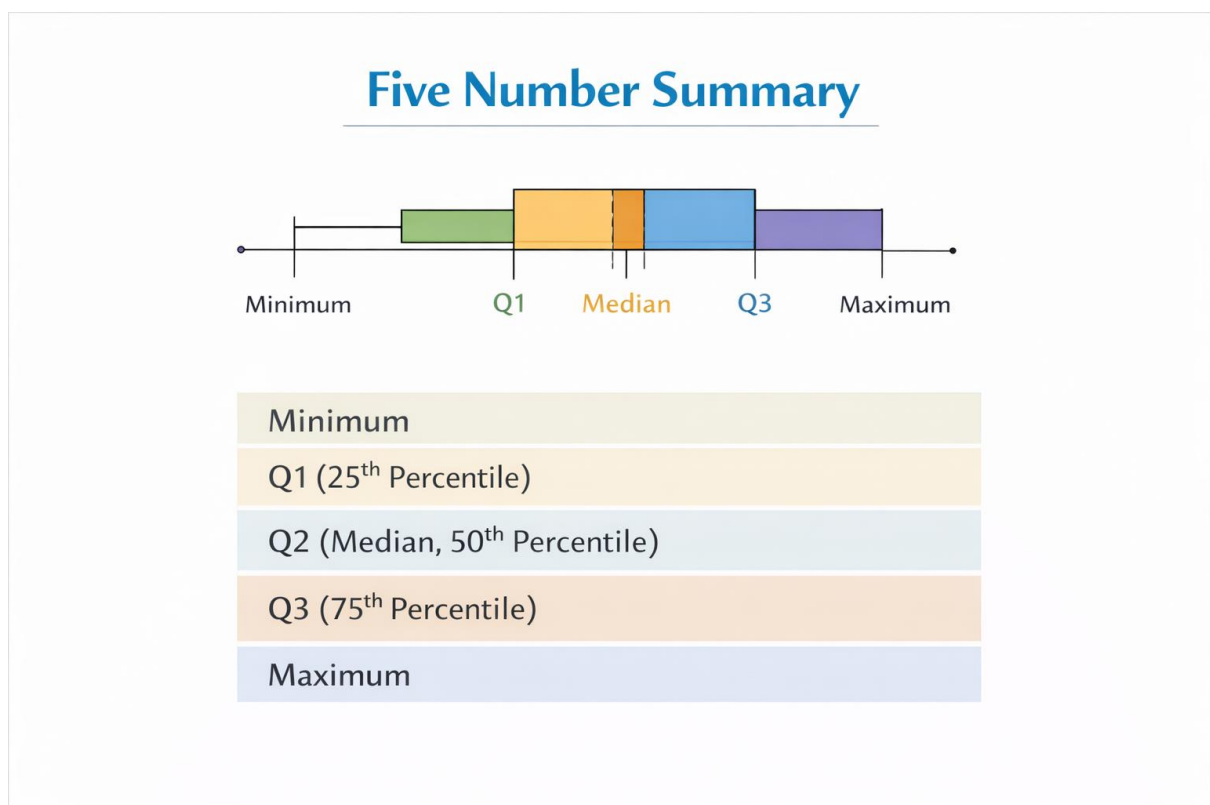


## 6. Five Number Summary

Five Number Summary includes:

1. Minimum
2. Q1
3. Median
4. Q3
5. Maximum

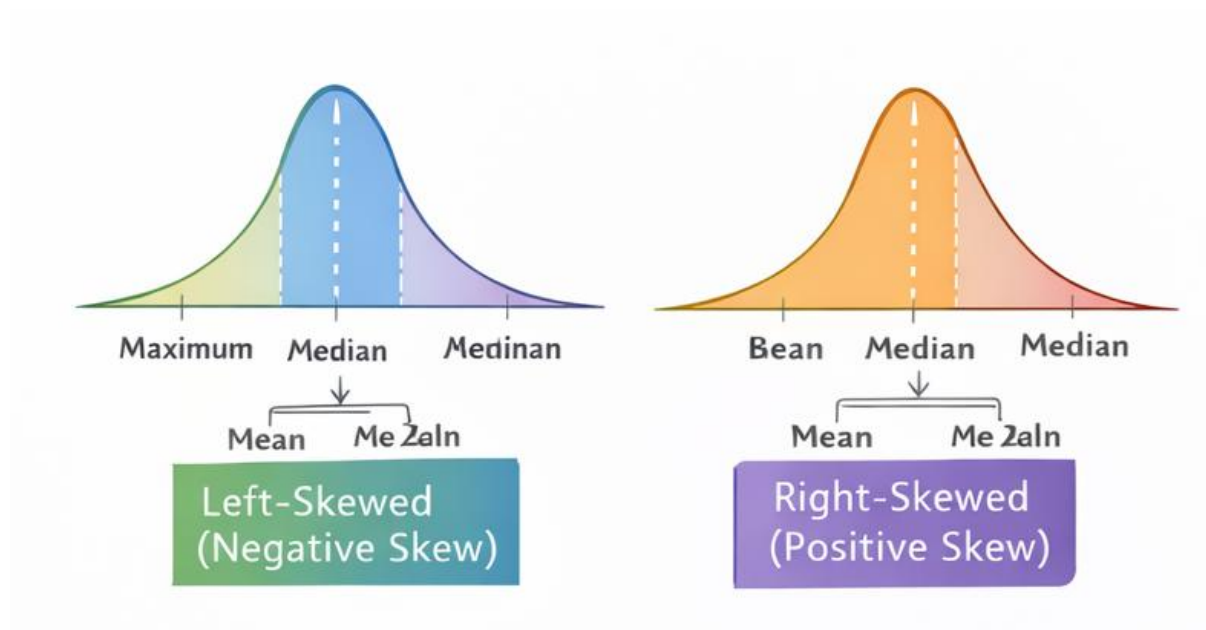
Used to create a Box Plot.



## 7. Skewness

Skewness measures the asymmetry of distribution.

- Positive skew → Right tail longer
- Negative skew → Left tail longer
- Zero skew → Symmetrical



## 8. Kurtosis

Kurtosis measures how peaked or flat a distribution is.

- Leptokurtic → Highly peaked
- Platykurtic → Flat
- Mesokurtic → Normal shape

