Statistics

Central Tendency

- Mean
- Median
- Mode
- Skewness
- Normal Distribution

Central Tendency

- Central tendency refers to the statistical measures that represent the center or typical value of a dataset. It summarizes the data with a single value that best describes the middle or center of its distribution

Mean:

The mean is the sum of all the data points divided by the number of points. It is sensitive to extreme values (outliers).

Example:

Consider the dataset:

$10, 20, 30, 40, 50$

The mean is calculated as:

$$\text{Mean} = \frac{10 + 20 + 30 + 40 + 50}{5} = \frac{150}{5} = 30$$

To find the mean, there are 2 steps:

1) add up all the data values
2) divide by the number of data values

Example: using our student scores:

1) 5+8+3+5+2+9+7+5+4=48
2) 48 ÷ 9 = 5.333

So the mean average of these scores, or the score that most represents the group is 5.333.

Student	Score
1	5
2	8
3	3
4	5
5	2
6	9
7	7
8	5
9	4

Median

The median is the middle value when the data is arranged in ascending or descending order. If there is an even number of data points, the median is the average of the two middle numbers.

Example:

For the dataset:

$10, 20, 30, 40, 50$

The median is 30 because it is the middle value.

For an even number of data points, say:

$10, 20, 30, 40$

The median is the average of 20 and 30, which is:

$$\text{Median} = \frac{20 + 30}{2} = 25$$

Mode

The mode is the value that appears most frequently in the dataset. A dataset may have more than one mode or no mode at all if all values occur with the same frequency.

Example:

For the dataset:

$10, 20, 20, 30, 40$

The mode is 20 because it appears twice, more than any other number.

- **Mean** is the arithmetic average.
- **Median** is the middle value.
- **Mode** is the most frequent value.
- These measures help summarize data, with each having different properties and being suitable for different types of analysis.

Skewness

- **Skewness** is a measure of the asymmetry of the probability distribution of a dataset.
- It indicates whether the data points in a dataset are more concentrated on one side of the distribution or whether the distribution has a tail on one side.
- Skewness helps understand the direction and extent to which the data deviates from a normal (symmetrical) distribution.

Types of Skewness

1. Positive Skewness (Right-Skewed Distribution):

1. In a positively skewed distribution, the tail on the right side of the distribution is longer or fatter than the left side.
2. The majority of the data values lie to the left of the mean, while the larger values (outliers) stretch out toward the right.
3. The mean is greater than the median, and the mode is less than the median.

Example:
Incomes in a population are often positively skewed. Most people earn a modest income, while a few earn extremely high salaries, pulling the distribution's right tail.

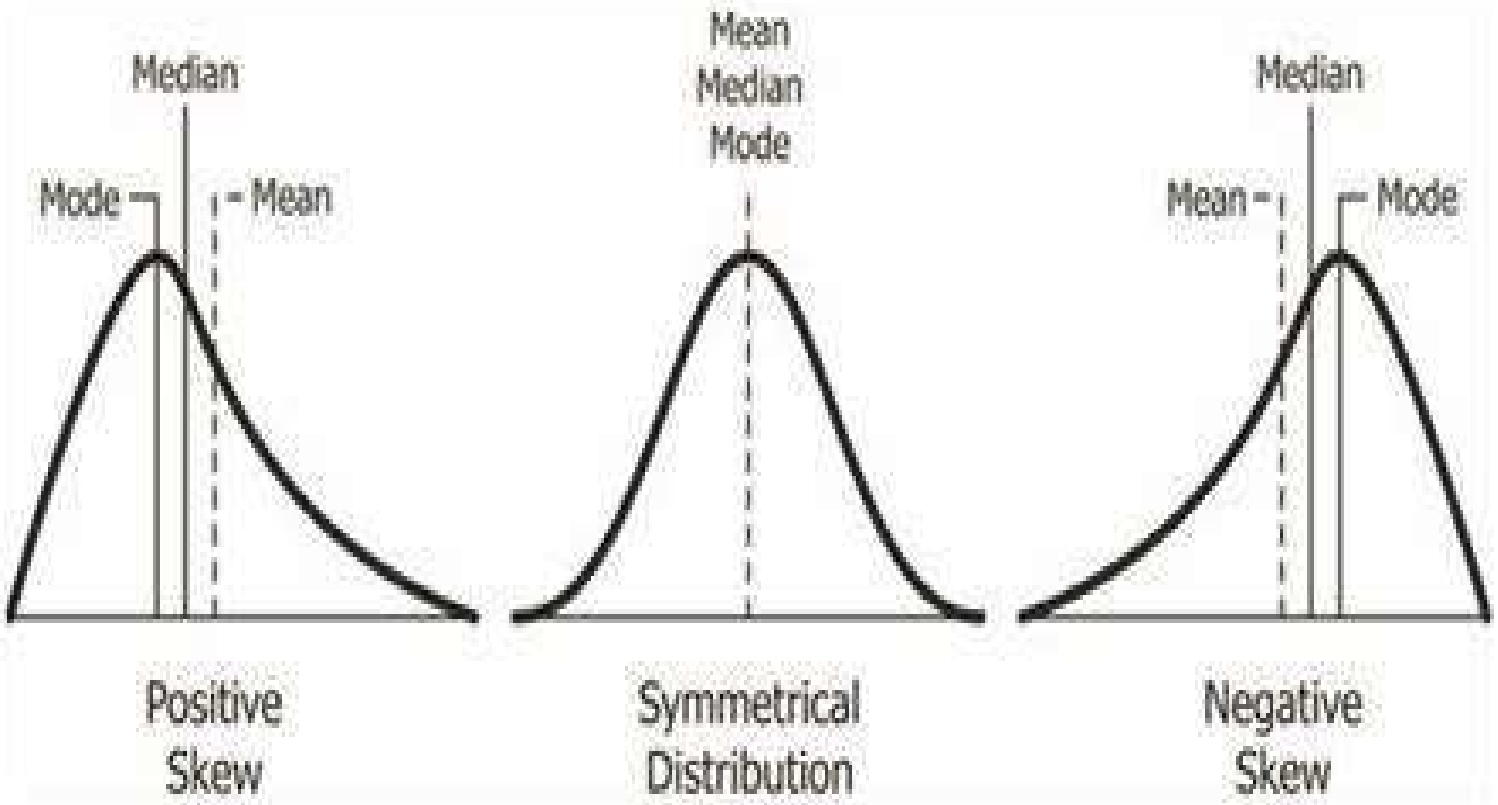
Mode < Median < Mean

2.Negative Skewness (Left-Skewed Distribution):

- In a negatively skewed distribution, the tail on the left side of the distribution is longer or fatter than the right side.
- Most of the data points lie to the right of the mean, while a few smaller values pull the distribution toward the left.
- The mean is less than the median, and the mode is greater than the median.
- **Example**:
  The age at retirement is often negatively skewed. Most people retire around the same age, but some retire earlier, pulling the left tail of the distribution.
- Mean < Median < Mode

- 3. **No Skewness (Symmetric Distribution)**:
- In a symmetric distribution, the data is evenly distributed on both sides of the mean, with no long tail on either side.
- The mean, median, and mode are approximately equal.
- **Example**:
  A perfect bell curve (normal distribution) is an example of no skewness, where the data is symmetrically distributed around the center.

Mean ≈ Median ≈ Mode


Median
Mode
Mean
Positive
Skew
Mean
Median
Mode
Symmetrical
Distribution
Median
Mean
Mode
Negative
Skew

- The representation of exam results forms a classic example of skewed distribution in real life. The distribution of scores obtained by the students of a class on any particularly difficult exam is generally positively skewed in nature. This is because due to the increased difficulty level of the exam, a majority of students tend to score low, and only a few of them manage to score high. Similarly, the distribution of scores obtained on an easy test is negatively skewed in nature because the reduced difficulty level of the exam helps more students score high, and only a few of them tend to score low.