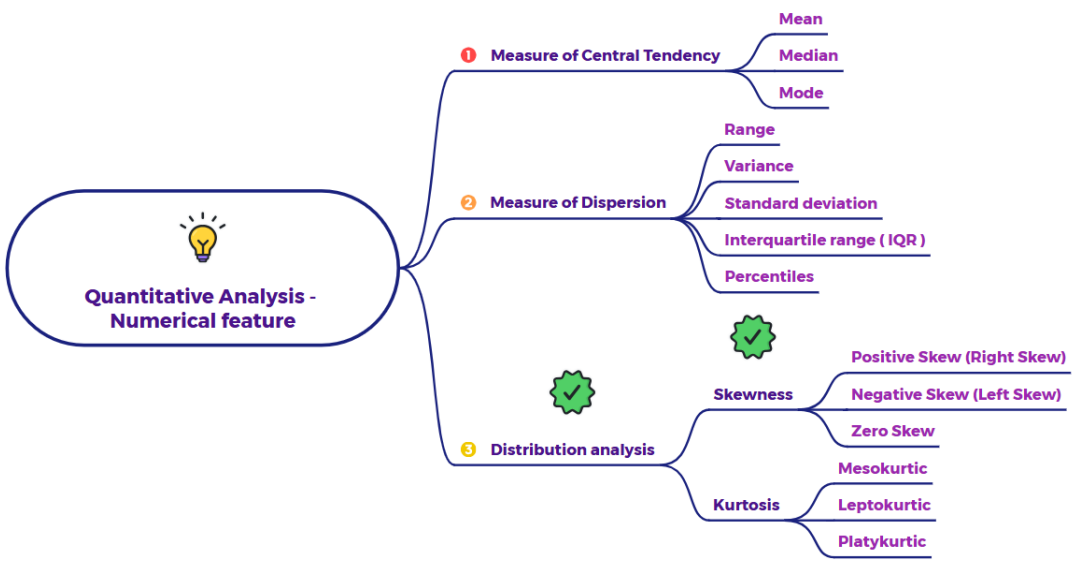
Explain Distribution Analysis – Skewness



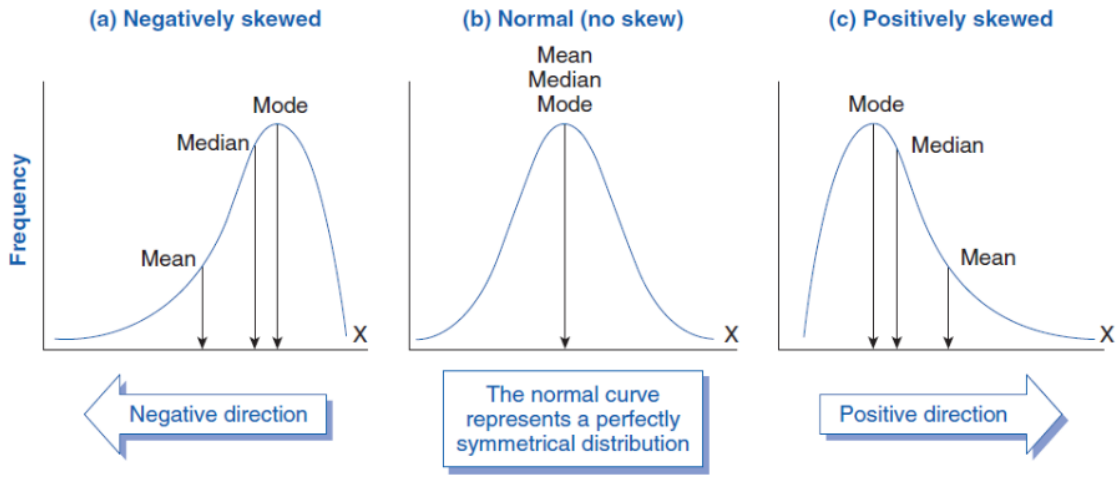
Concept:

A distribution is considered skewed if it's not symmetrical around its mean. Skewness indicates whether the tail of the distribution is heavier on the left or the right side.

There are three main types of skewness:

- **Positive Skew (Right Skew):** The tail on the right side of the distribution is longer or fatter than the tail on the left side. In a positively skewed distribution, the mean is typically greater than the median, which is usually greater than the mode. This often occurs when there are some unusually high values pulling the mean to the right.
- **Negative Skew (Left Skew):** The tail on the left side of the distribution is longer or fatter than the tail on the right side. In a negatively skewed distribution, the mean is typically less than the median, which is usually less than the mode. This often happens when there are some unusually low values pulling the mean to the left.
- **Zero Skew (Symmetrical):** The distribution is symmetrical around its mean, resembling a bell curve (normal distribution) or a uniform distribution. In a perfectly symmetrical distribution, the mean, median, and mode are usually equal (or very close).

Visual Representation:



Imagine plotting the frequency of your numerical data (e.g., using a histogram or a density plot):

- **Positive Skew:** The peak of the distribution is towards the left, and the tail extends towards the right.
- **Negative Skew:** The peak of the distribution is towards the right, and the tail extends towards the left.
- **Zero Skew:** The distribution is balanced around the center.

Detailed Examples:

Let's consider different scenarios to illustrate skewness:

Example 1: Income Distribution

Suppose we look at the annual income distribution of people in a certain region. Income distributions are often positively skewed.

- Most people earn in the lower to middle-income range.
- A smaller number of people earn very high incomes, creating a long tail to the right of the distribution.

In this case, the mean income would likely be higher than the median income because the few very high incomes pull the average upwards, while the median (the middle income) is less affected by these extremes. The mode (the most

frequent income range) would likely be in the lower to middle range, further to the left.

Example 2: Exam Scores (Negatively Skewed)

Imagine a very easy exam where most students perform exceptionally well, but a few struggled. The distribution of scores might be negatively skewed.

- A large number of students score very high grades (e.g., 90s and 100s), forming the bulk of the distribution towards the right.
- A few students score much lower (e.g., 40s, 50s), creating a tail extending to the left.

In this scenario, the mean score would likely be lower than the median score because the few low scores pull the average downwards, while the median (the middle score) remains high. The mode (the most frequent score range) would be at the higher end.

Example 3: Heights of Adult Males

The distribution of heights of adult males tends to be approximately symmetrical (zero skew or very close to it).

- Most men fall within a certain average height range.
- There are roughly equal numbers of men who are significantly taller and significantly shorter than the average.

In this case, the mean, median, and mode of the height distribution would likely be very close to each other.

Quantifying Skewness:

While we can visually assess skewness, there are also numerical measures to quantify it, such as Pearson's coefficient of skewness and the skewness based on moments (often calculated by statistical software). These measures provide a numerical value indicating the degree and direction of skewness. A value close to zero indicates low skewness, a positive value indicates positive skew, and a negative value indicates negative skew. The magnitude of the value suggests the strength of the skew.

Importance of Understanding Skewness:

- **Choosing Appropriate Statistical Methods:** Many statistical tests and models assume a normal (symmetrical) distribution. Significant skewness can violate these assumptions and lead to inaccurate results. Transformations of the data might be needed to reduce skewness.
- **Interpretation of Data:** Understanding the skew helps in correctly interpreting the central tendency measures. For skewed data, the median might be a more representative measure of the "typical" value than the mean.
- **Identifying Potential Issues:** In some contexts, skewness can indicate underlying problems or phenomena. For example, a positive skew in waiting times might suggest bottlenecks in a system.

In summary, skewness is a crucial aspect of distribution analysis that describes the lack of symmetry in a dataset. Positive skew has a long right tail, negative skew has a long left tail, and zero skew indicates symmetry. Understanding the skewness of your data is important for choosing appropriate statistical methods and for correctly interpreting the characteristics of the distribution.