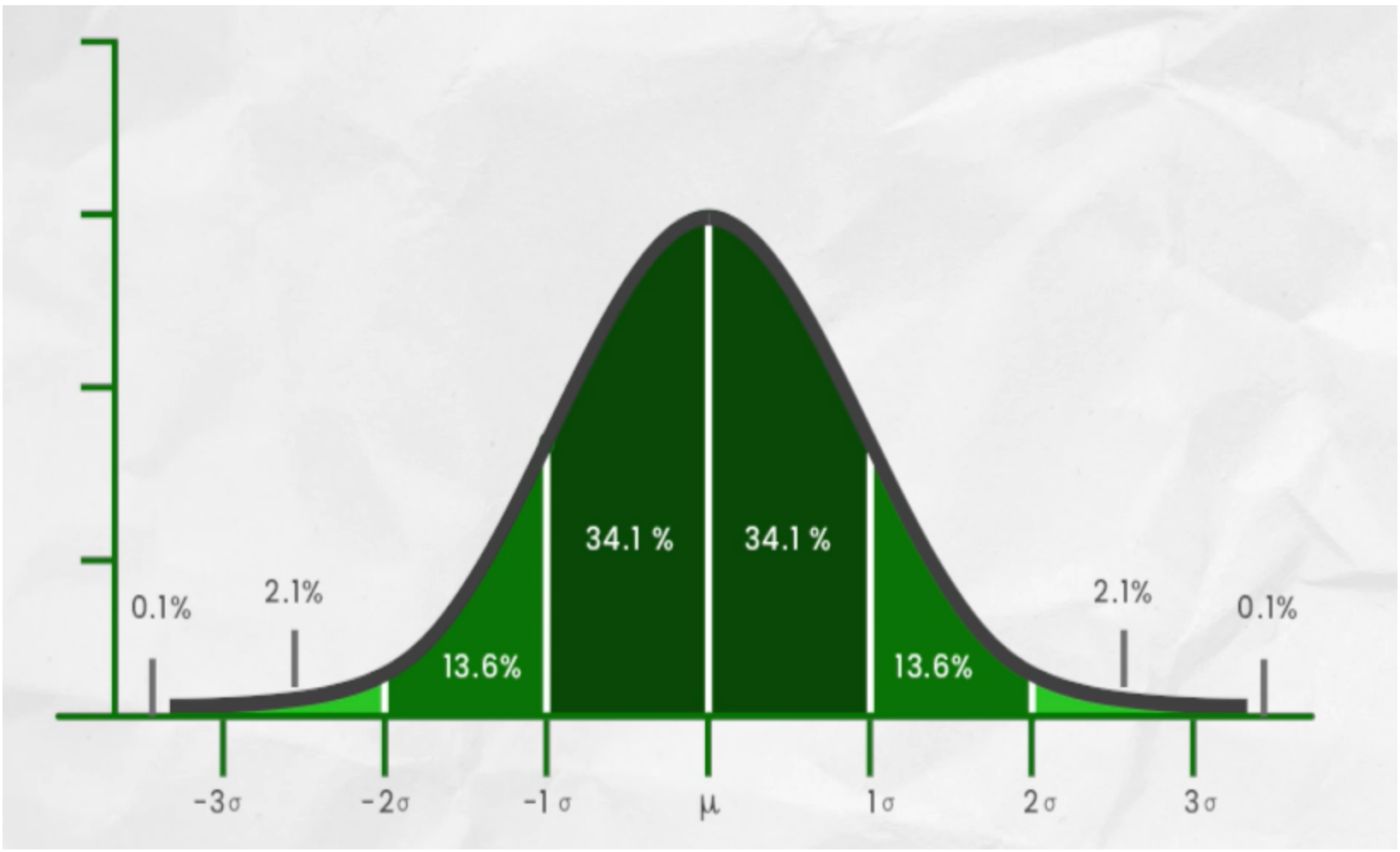
Measures of Shape

When analyzing data, it is essential to not only understand the central tendency and variability of the data but also its shape. The shape of the distribution tells us about the pattern of the data and how the values are distributed around the central tendency. Measures of shape are used to describe the distribution of data and help us to understand the patterns that may be present.

Normal Distribution:

When we plot a dataset such as a histogram, the shape of that charted plot is what we call its distribution. The most commonly observed shape of continuous values is the bell curve, also called the Gaussian or normal distribution. Most of the data in normal distribution is clustered around the center while the outliers are present towards the end. Here about 68% of the values lie within one standard deviation away from the mean. About 95% of the values lie within 2 standard deviations and 99.7% within 3 standard deviations. This is known as the empirical rule or the 3-sigma rule

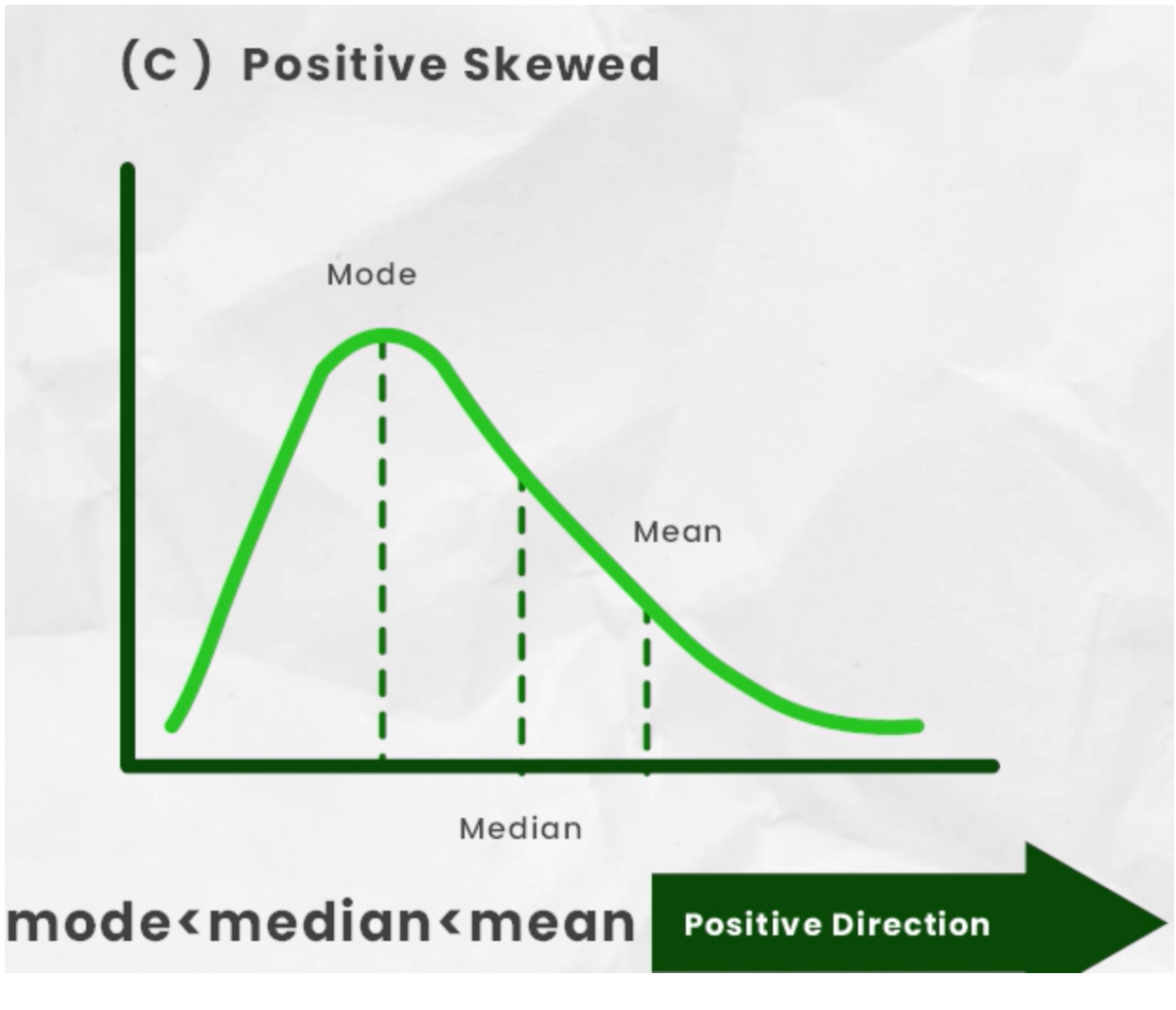


Skewness:

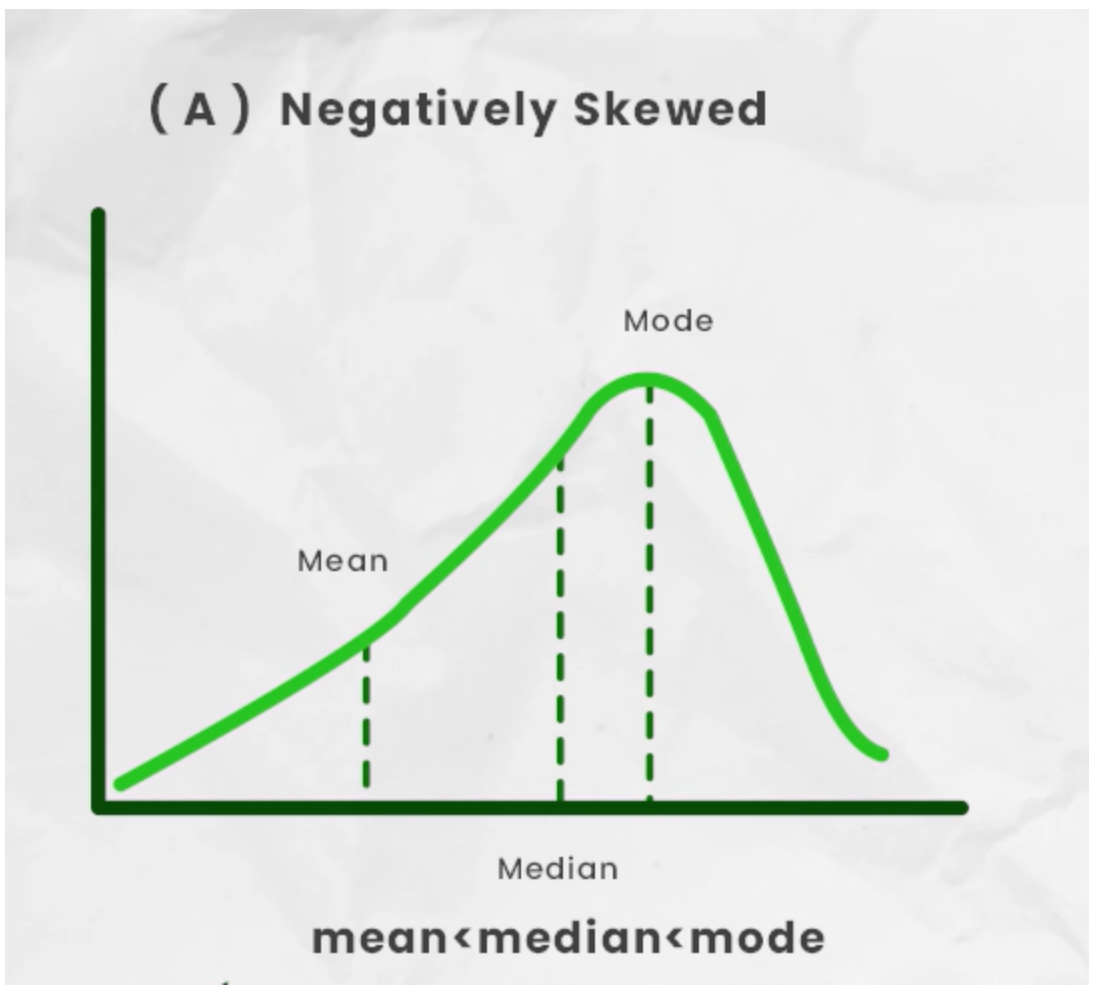
Skewness is a statistical measure that describes the degree of asymmetry in a dataset's distribution. It is used to understand how the data points are distributed around the mean value. A distribution can be positively skewed, negatively skewed, or have no skewness. In this article, we will discuss the different types of skewness and Pearson's coefficient of skewness

Types of Skewness

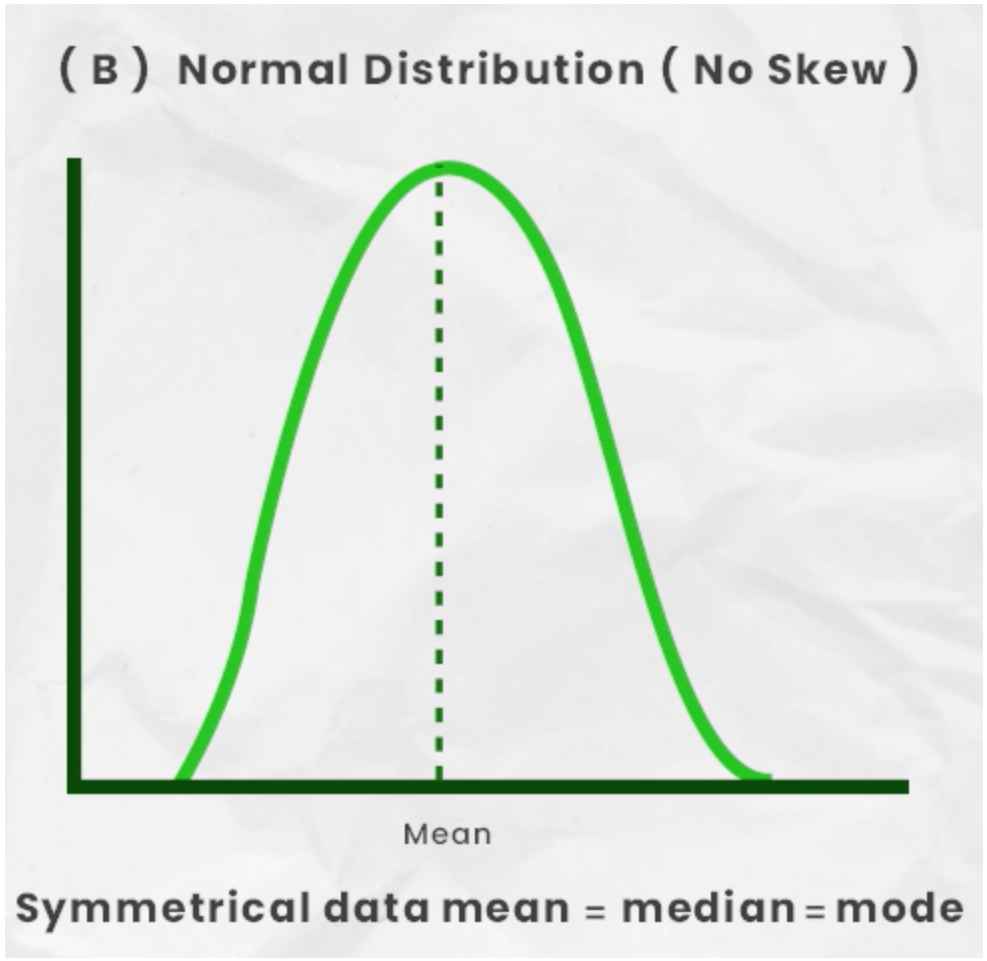
- **Positive Skewness:** A distribution is said to be positively skewed when the tail of the distribution extends towards the right-hand side of the curve. The mean of a positively skewed distribution is greater than the mode and median. This indicates that there are more data points on the left-hand side of the distribution and fewer data points on the right-hand side.



- **Negative Skewness**: A distribution is said to be negatively skewed when the tail of the distribution extends towards the left-hand side of the curve. The mean of a negatively skewed distribution is less than the mode and median. This indicates that there are more data points on the right-hand side of the distribution and fewer data points on the left-hand side.



- **Normal Skewness:** A normal distribution is a distribution that is perfectly symmetrical. The mean, mode, and median of a normal distribution are equal. This means that there are an equal number of data points on the left-hand and right-hand sides of the distribution.



Pearson’s coefficient of skewness

Pearson's coefficient of skewness is a statistical measure that determines the degree of asymmetry in a probability distribution. It is based on the difference between the mean, mode, and median of a distribution. If the mean is greater than the mode and median, then the distribution is positively skewed. If the mean is less than the mode and median, then the distribution is negatively skewed. If the mean, mode, and median are equal, then the distribution is said to be perfectly symmetrical.

Pearson's first coefficient of skewness (P1)

P1 = (mean - mode) / Standard Deviation

where mean is the arithmetic mean of the distribution, mode is the mode of the distribution, and standard deviation is the standard deviation of the distribution.

Pearson's second coefficient of skewness (P2)

P2 = 3 * (mean - median) / Standard Deviation

where mean is the arithmetic mean of the distribution, median is the median of the distribution, and standard deviation is the standard deviation of the distribution.

1. **Positive Skewness:** For a positively skewed distribution, Pearson's first coefficient of skewness (P1) will be greater than zero, indicating that the distribution is positively skewed. Similarly, Pearson's second coefficient of skewness (P2) will also be greater than zero.
2. **Negative Skewness:** For a negatively skewed distribution, Pearson's first coefficient of skewness (P1) will be less than zero, indicating that the distribution is negatively skewed. Similarly, Pearson's second coefficient of skewness (P2) will also be less than zero.
3. **Normal Distribution:** For a normal distribution, Pearson's first coefficient of skewness (P1) and Pearson's second coefficient of skewness (P2) will both be equal to zero.

In conclusion, Pearson's coefficient of skewness is a useful statistical measure that can be used to determine the degree of asymmetry in a probability distribution. A positive value of P1 or P2 indicates a positively skewed distribution, while a negative value of P1 or P2 indicates a negatively skewed distribution. A value of zero for P1 or P2 indicates a perfectly symmetrical distribution.

Kurtosis

Kurtosis is a statistical measure that determines the shape of a probability distribution. It measures the degree to which the distribution is peaked or flat relative to a normal distribution. Kurtosis can help to identify the presence of outliers or heavy tails in a distribution.

There are three main types of kurtosis:

- Mesokurtic
- Leptokurtic
- Platykurtic.

1. **Mesokurtic Distribution:** A mesokurtic distribution is a distribution that has the same degree of kurtosis as a normal distribution. A normal distribution has a kurtosis value of zero. A mesokurtic distribution has a peak that is similar to a normal distribution, with tails that are neither heavy nor light. The distribution is symmetrical around the mean.
2. **Leptokurtic Distribution:** A leptokurtic distribution is a distribution that has a higher degree of kurtosis than a normal distribution. The peak of a leptokurtic distribution is higher and sharper than a normal distribution, with heavier tails. This means that there are more data points in the tails of the distribution than a normal distribution. Leptokurtic distributions are often seen in financial data and in the stock market, where extreme events are more common.
3. **Platykurtic Distribution:** A platykurtic distribution is a distribution that has lower kurtosis than a normal distribution. The kurtosis value of a platykurtic distribution is less than 3. This type of distribution has shorter and flatter tails compared to a normal distribution. In a platykurtic distribution, the data is more spread out, and there are fewer outliers. This indicates that the data has fewer extreme values and a lower level of variability.