SKEWNESS AND KURTOSIS

Yash Sarang	47
Madhusudhana Naidu	36
Om Gaydhane	15
Manav Pahilwani	37
Parth Suryavanshi	58
Akshat Tiwari	62
Surabhi Tambe	59

ABOUT SKEWNESS AND KURTOSIS

"Skewness essentially measures the symmetry of the distribution, while kurtosis determines the heaviness of the distribution tails."

The understanding shape of data is a crucial action.
It helps to understand where the most information is lying and analyze the outliers in a given data.

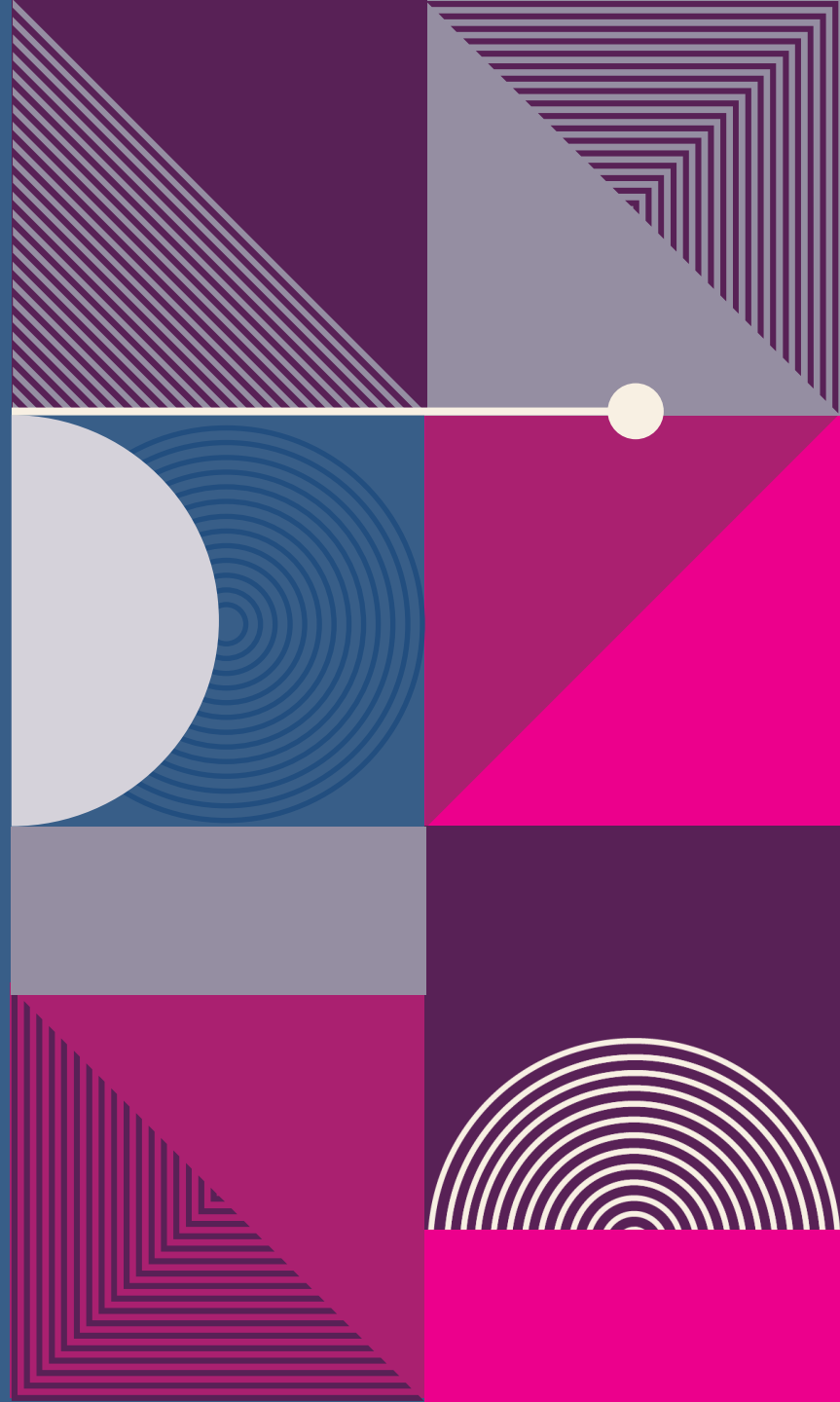
Skewness

If the values of a specific independent variable (feature) are skewed, depending on the model, skewness may violate model assumptions or may reduce the interpretation of feature importance.

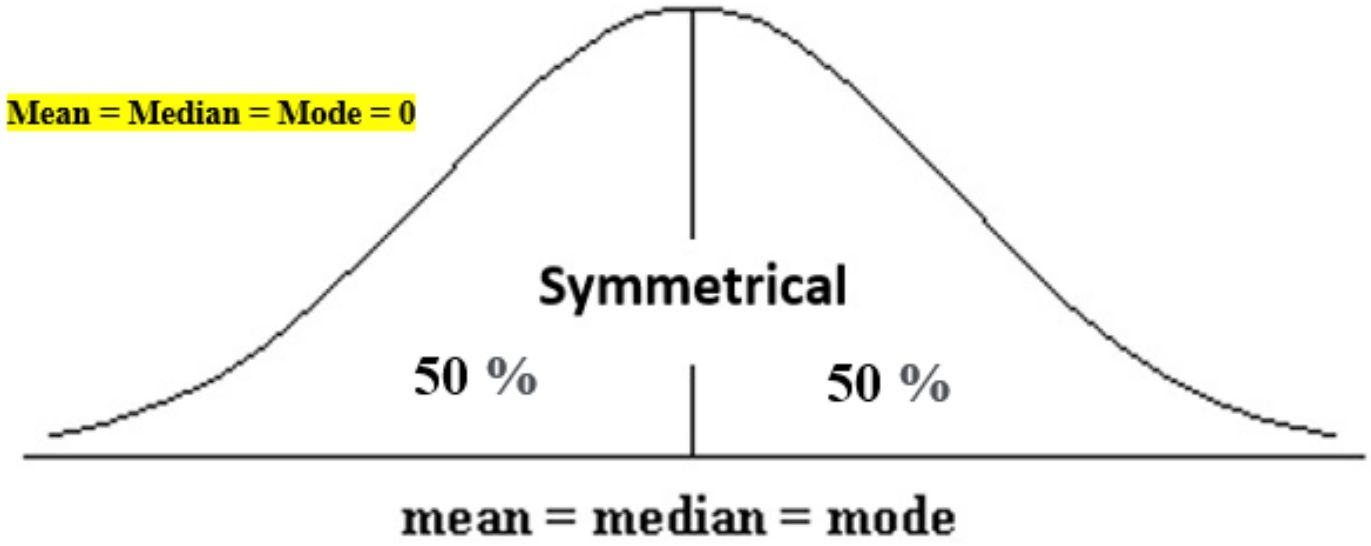
In statistics, skewness is a degree of asymmetry observed in a probability distribution that deviates from the symmetrical normal distribution (bell curve) in a given set of data.

The normal distribution helps to know a skewness. When we talk about normal distribution, data symmetrically distributed.

The symmetrical distribution has zero skewness as all measures of a central tendency lies in the middle.



Skewness



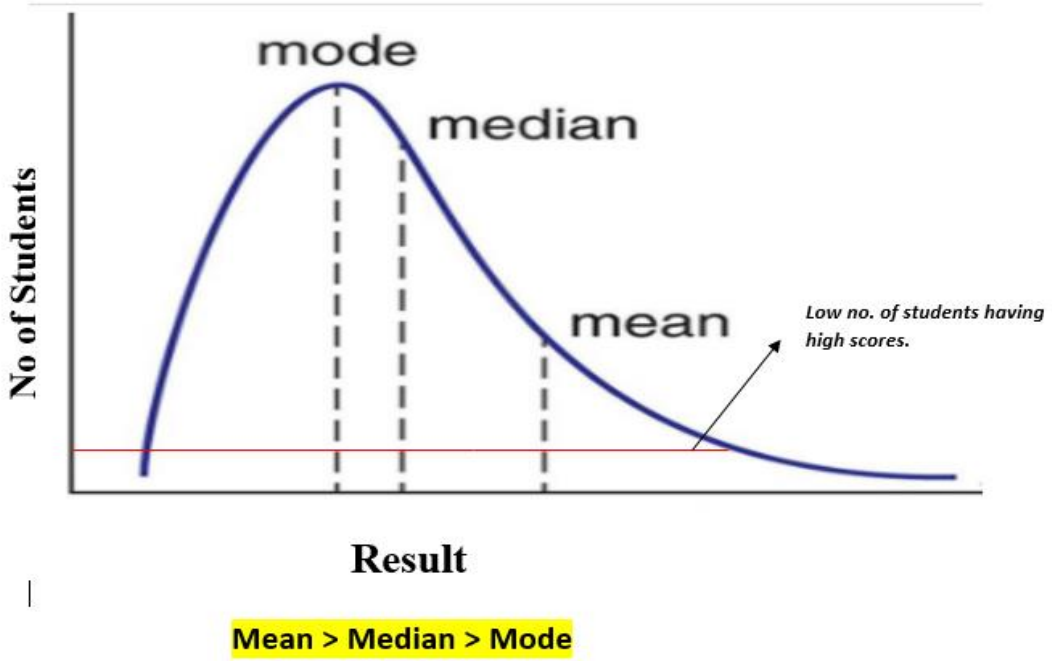
When data is symmetrically distributed,
the left-hand side, and right-hand side, contain the same number of observations.
(If the dataset has 90 values, then the left-hand side has 45 observations, and the right-hand side has 45 observations.).

But, what if not symmetrical distributed?
That data is called asymmetrical data, and that time skewness comes into the picture.

Types of skewness

1. Positive skewed or right-skewed

In statistics, a positively skewed distribution is a sort of distribution where, u*nlike symmetrically distributed data where all measures of the central tendency (mean, median, and mode) equal each other,* with positively skewed data, the measures are dispersing, which means Positively Skewed Distribution is a type of distribution where the mean, median, and mode of the distribution are positive rather than negative or zero.

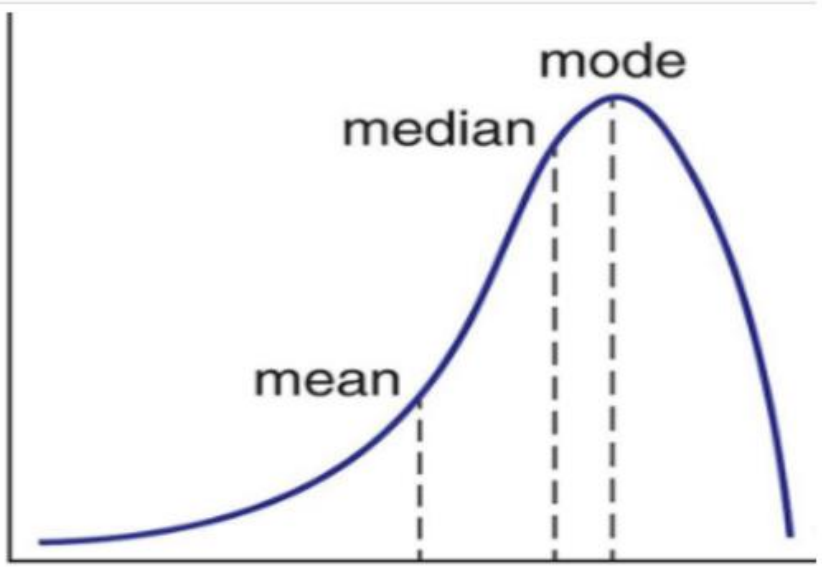


2. Negative skewed or left-skewed

A negatively skewed distribution is the straight reverse of a positively skewed distribution. In statistics, negatively skewed distribution refers to the distribution model where more values are plots on the right side of the graph, and the tail of the distribution is spreading on the left side.

In negatively skewed, the mean of the data is less than the median (a large number of data-pushed on the left-hand side).

Negatively Skewed Distribution is a type of distribution where the mean, median, and mode of the distribution are negative rather than positive or zero.



- ***Pearson's first coefficient of skewness***

Subtract a mode from a mean, then divides the difference by standard deviation.

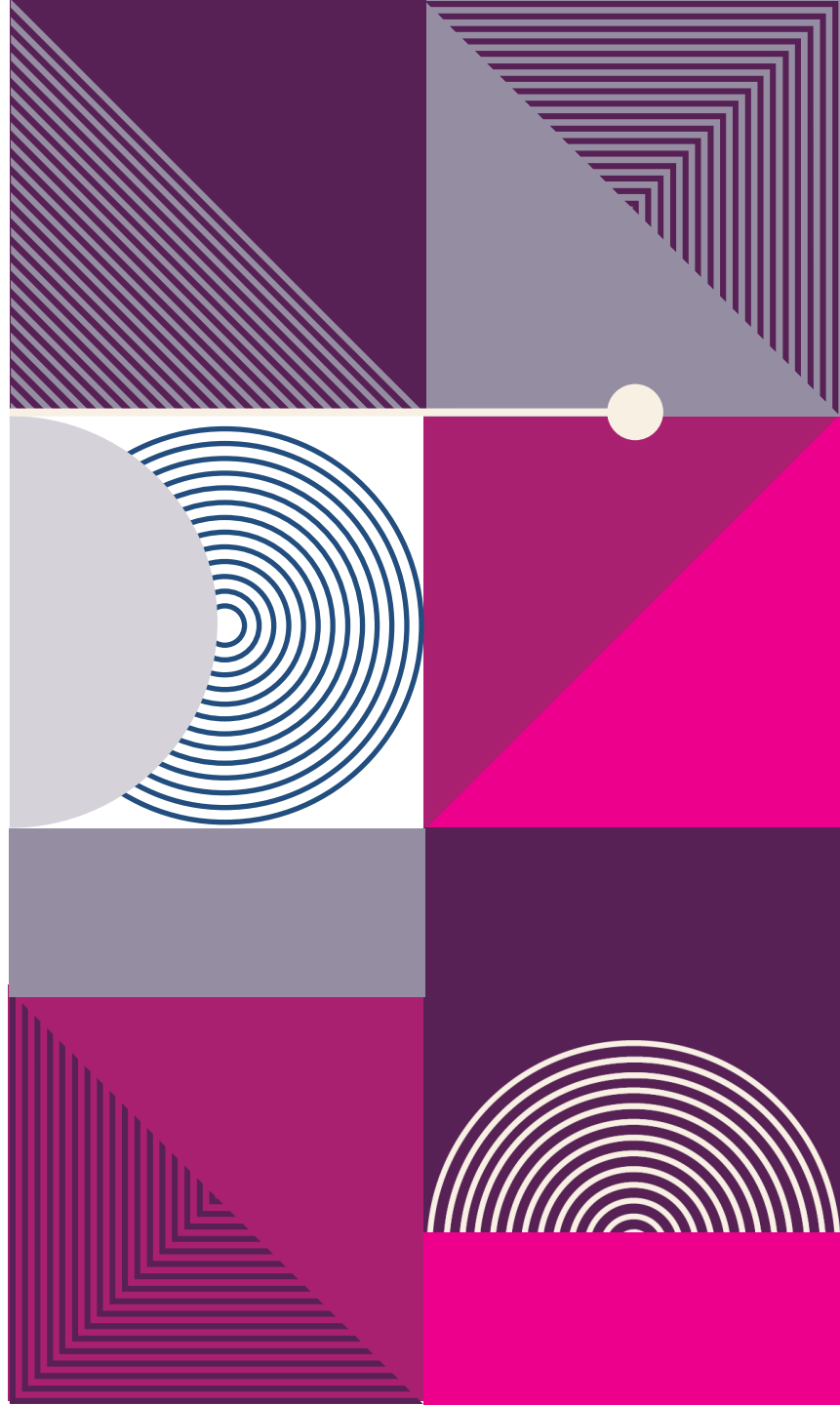
$$\text{Pearson's first coefficient} = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$$

- ***Pearson's second coefficient of skewness***

Multiply the difference by 3, and divide the product by standard deviation.

$$\text{Pearson's second coefficient} = \frac{3\,(\text{Mean} - \text{Median})}{\text{Standard Deviation}}$$

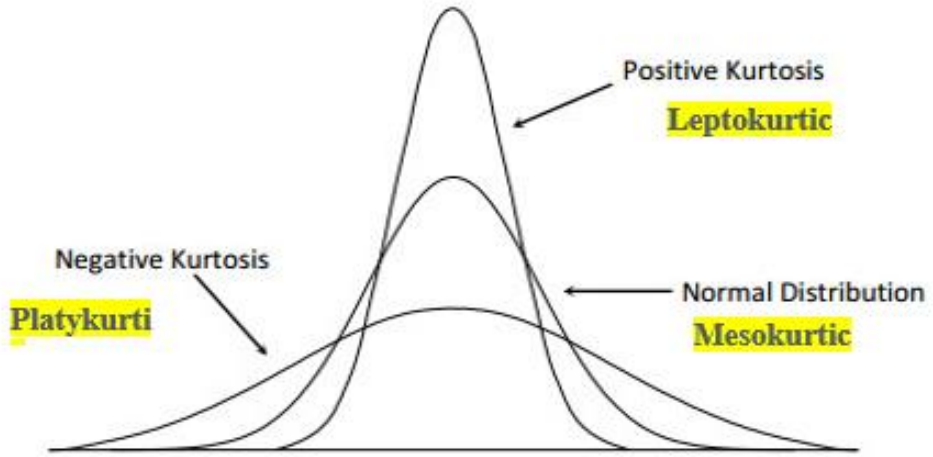
$$\text{Mean} - \text{Mode} \approx 3\,(\text{Mean} - \text{Median})$$



KURTOSIS

Kurtosis

Kurtosis refers to the degree of presence of outliers in the distribution. Kurtosis is a statistical measure, whether the data is heavy-tailed or light-tailed in a normal distribution.



Excess Kurtosis

The excess kurtosis is used in statistics and probability theory to compare the kurtosis coefficient with that normal distribution.

EXCESS KURTOSIS

Excess kurtosis can be positive (Leptokurtic distribution),
negative (Platykurtic distribution),
or near to zero (Mesokurtic distribution).
Since normal distributions have a kurtosis of 3, excess kurtosis is calculating by subtracting kurtosis by 3.

Excess kurtosis = Kurt – 3

Types of excess kurtosis

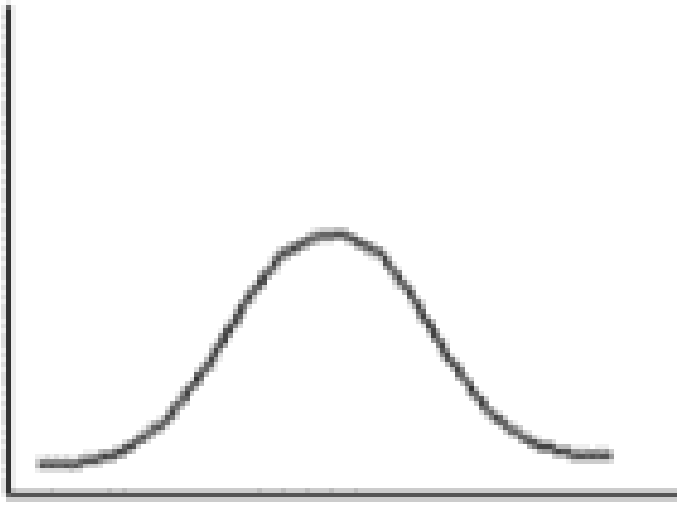
1. Leptokurtic or heavy-tailed distribution
   (kurtosis more than normal distribution).

2. Mesokurtic
   (kurtosis same as the normal distribution).

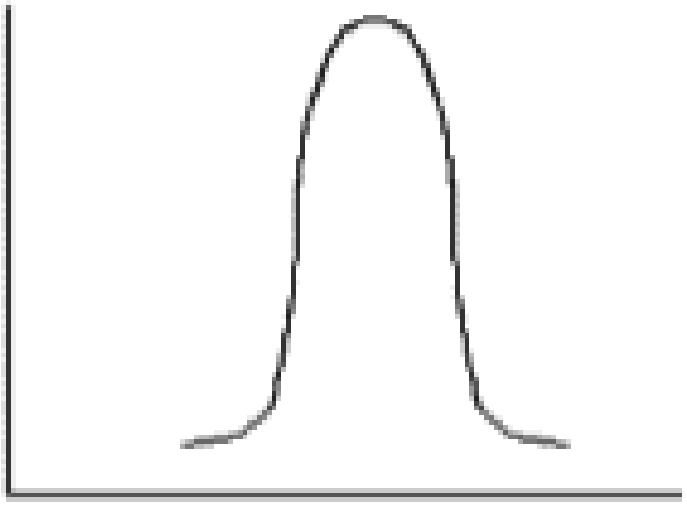
3. Platykurtic or short-tailed distribution
   (kurtosis less than normal distribution).

Types of excess kurtosis

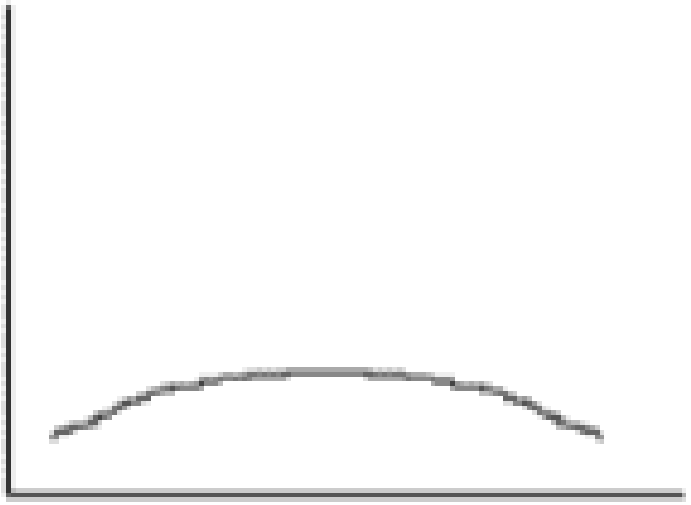
MESOKURTIC (KURTOSIS = 3)

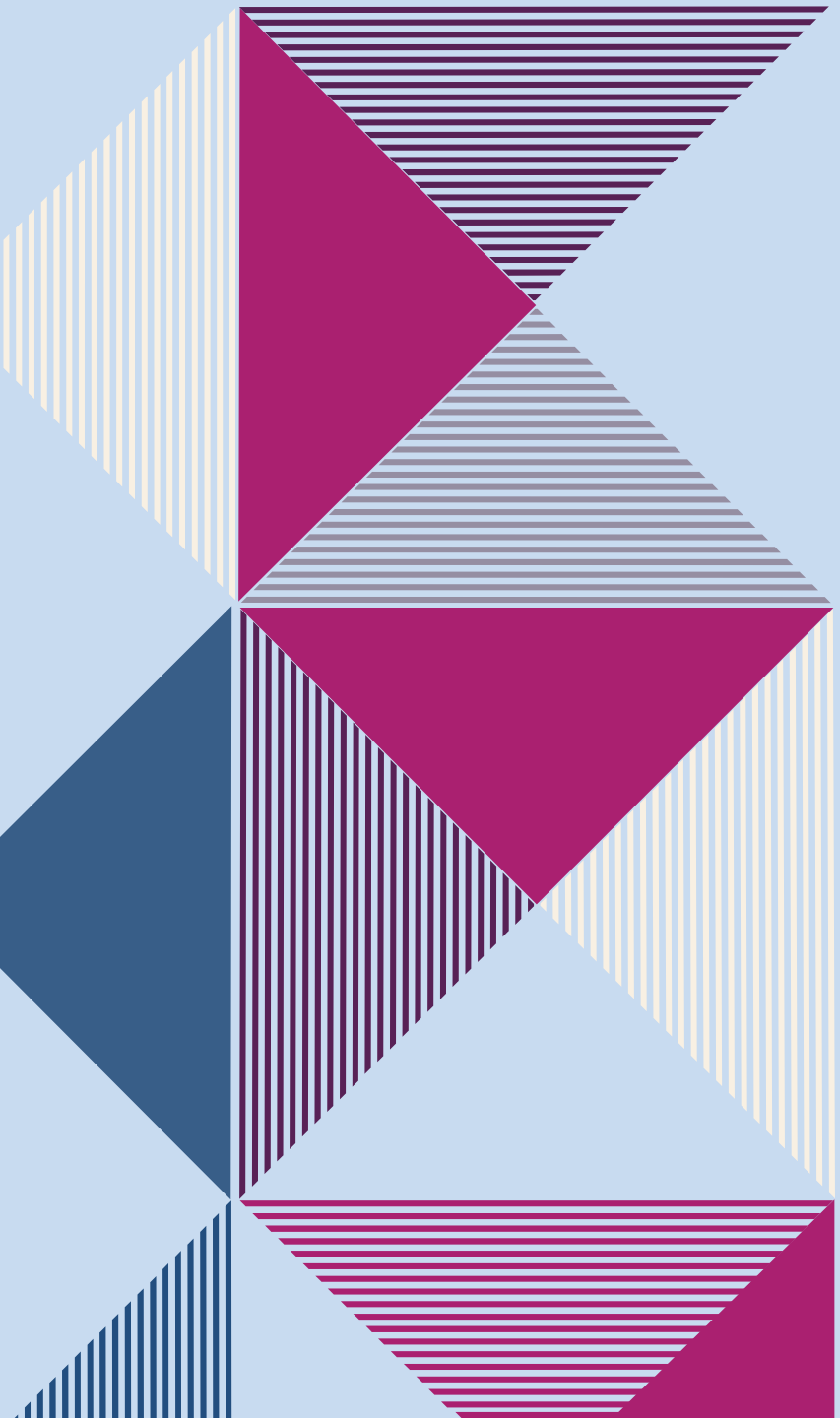


LEPTOKURTIC (KURTOSIS > 3)



PLATYKURTIC (KURTOSIS < 3)





SUMMARY

The skewness is a measure of symmetry or asymmetry of data distribution, and kurtosis measures whether data is heavy-tailed or light-tailed in a normal distribution.

When data skewed, the tail region may behave as an outlier for the statistical model, and outliers unsympathetically affect the model's performance especially regression-based models.

Some statistical models are hardy to outliers like Tree-based models, but it will limit the possibility to try other models. So there is a necessity to transform the skewed data to close enough to a Normal distribution.

THANK YOU