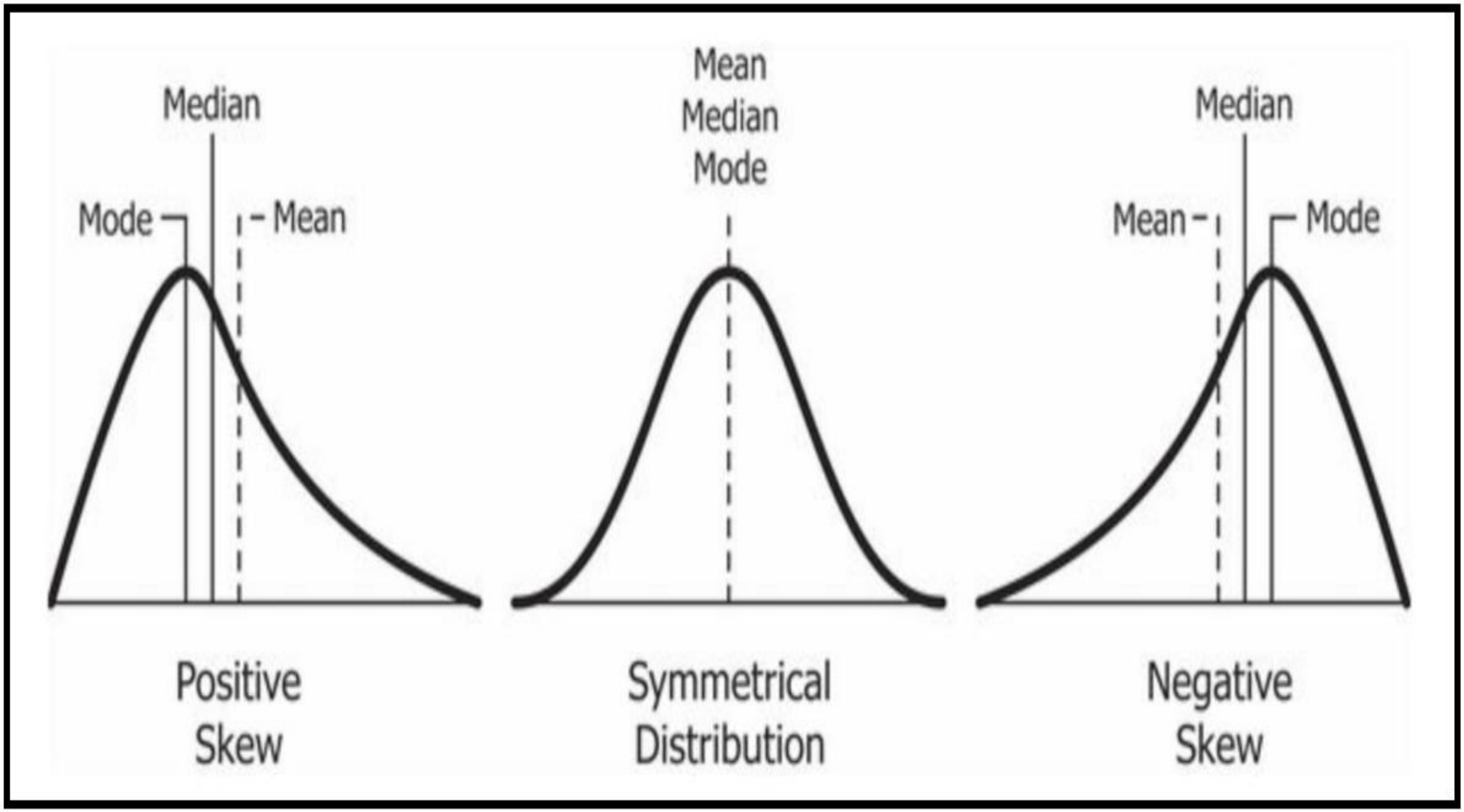
SKEWNESS
&
KURTOSIS

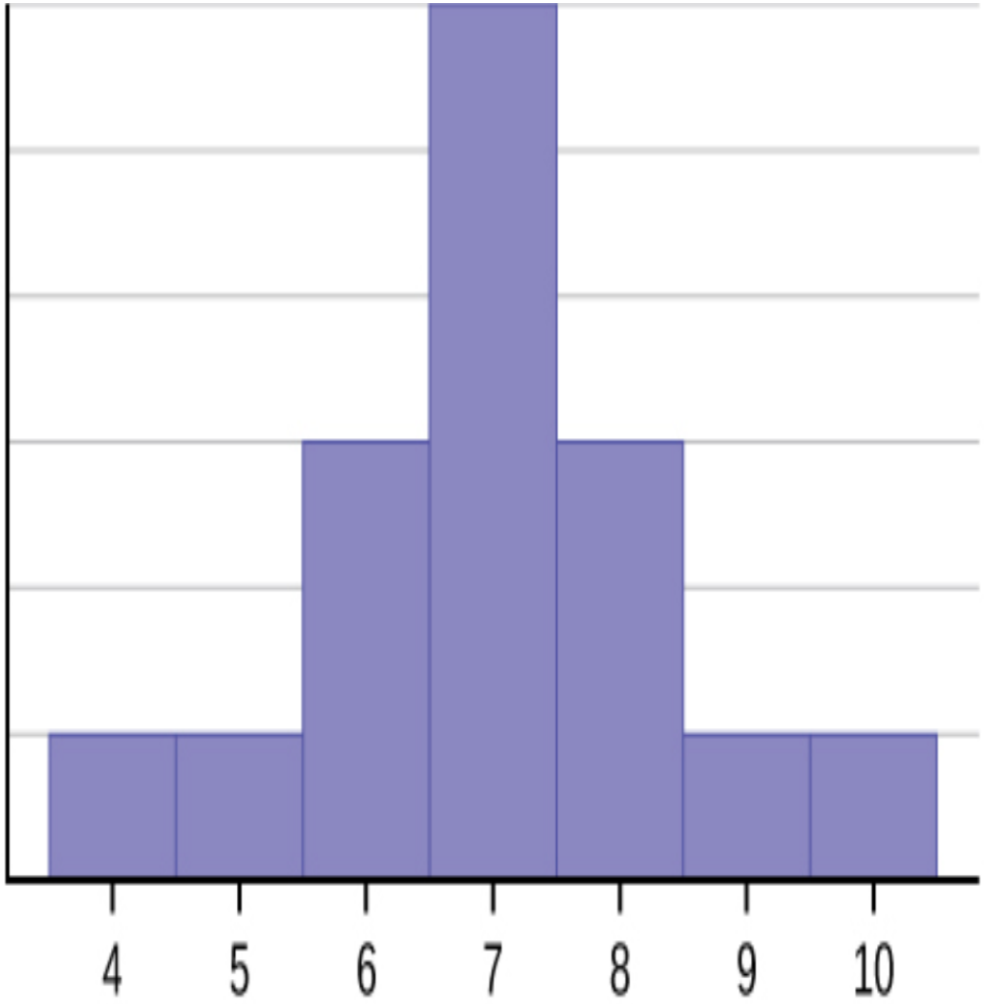
What is Skewness ?

- Skewness is the measure of the asymmetry of an ideally symmetric probability distribution and is given by the third standardized moment.
- In simple words, skewness is the measure of how much the probability distribution of a random variable deviates from the **normal distribution.**
  - **Bell-shaped curve** a **Normal Distribution.** Carl Friedrich Gauss discovered it so sometimes we also call it a **Gaussian Distribution** as well.
  - We can simplify the Normal Distribution's Probability Density by using only two parameters: **μ** Mean and **σ** S.D.
  - This curve is **symmetric** around the Mean.
  - For this distribution, the **Mean, Median, and Mode are all the same.**
  - the normal distribution is the probability distribution without any skewness.


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

Symmetrical Distribution

- Consider the following data set:
  4,5,6,6,6,7,7,7,7,7,7,8,8,8,9, 10
- This data set can be represented by this histogram.
- Each interval has width one, and each value is located in the middle of an interval.



$$\text{Mean} = \frac{4+5+6+6+6+7+7+7+7+7+7+8+8+8+9+10}{16}$$

$$= \frac{112}{16}$$

Mean = 7

$$\text{Median} = \frac{N+1}{2} = \frac{16+1}{2} = \frac{17}{2} = 8.5 = \frac{7+7}{2} = \frac{14}{2} = 7$$

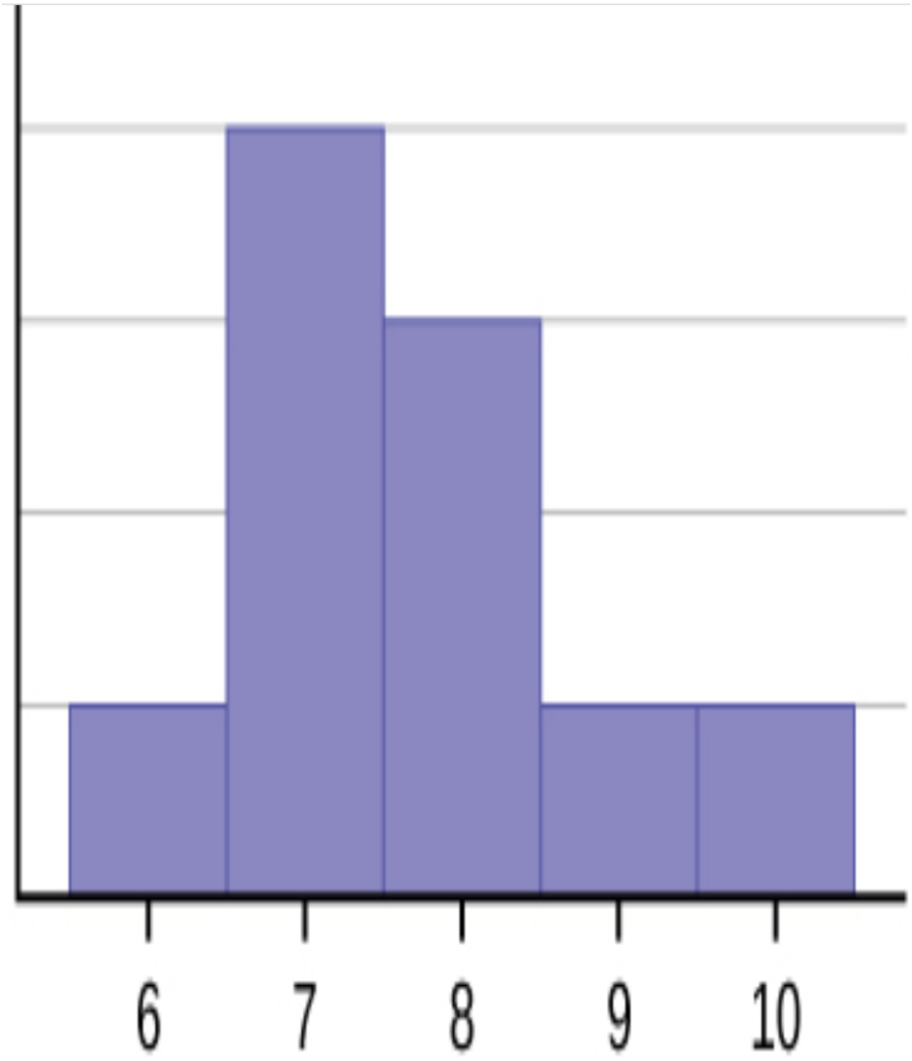
Median = 7

Mode = 7

- The histogram displays a **symmetrical** distribution of data.
- A distribution is symmetrical if a vertical line can be drawn at some point in the histogram such that the shape to the left and the right of the vertical line are mirror images of each other.
- The mean, the median, and the mode are each seven (7) for these data.
- **In a perfectly symmetrical distribution, the mean and the median are the same.**
- This example has one mode (unimodal), and the mode is the same as the mean and median.
- **In a symmetrical distribution that has two modes (bimodal), the two modes would be different from the mean and median.**

Positive Skewness

- The histogram for the data: 6,7,7,7,7,8,8,8,9,10 is not symmetrical.
- The left-hand side seems “chopped off” compared to the right side.
- A distribution of this type is called **skewed to the right** because it is pulled out to the right.



Positive Skewness

- Mean : $\frac{6+7+7+7+7+8+8+8+9+10}{10} = \frac{77}{10}$
- Mean = 7.7
- Median = $\frac{N+1}{2} = \frac{10+1}{2} = \frac{11}{2} = 5.5 = \frac{7+8}{2} = \frac{15}{2} = 7.5$
- Mode = 7
- Notice that the mean is greater than the median, and they are both greater than the mode.
- **Mean > Median > Mode**
- Of the three statistics, **the mean is the largest, while the mode is the smallest**. Again, the mean reflects the skewing the most.
- .

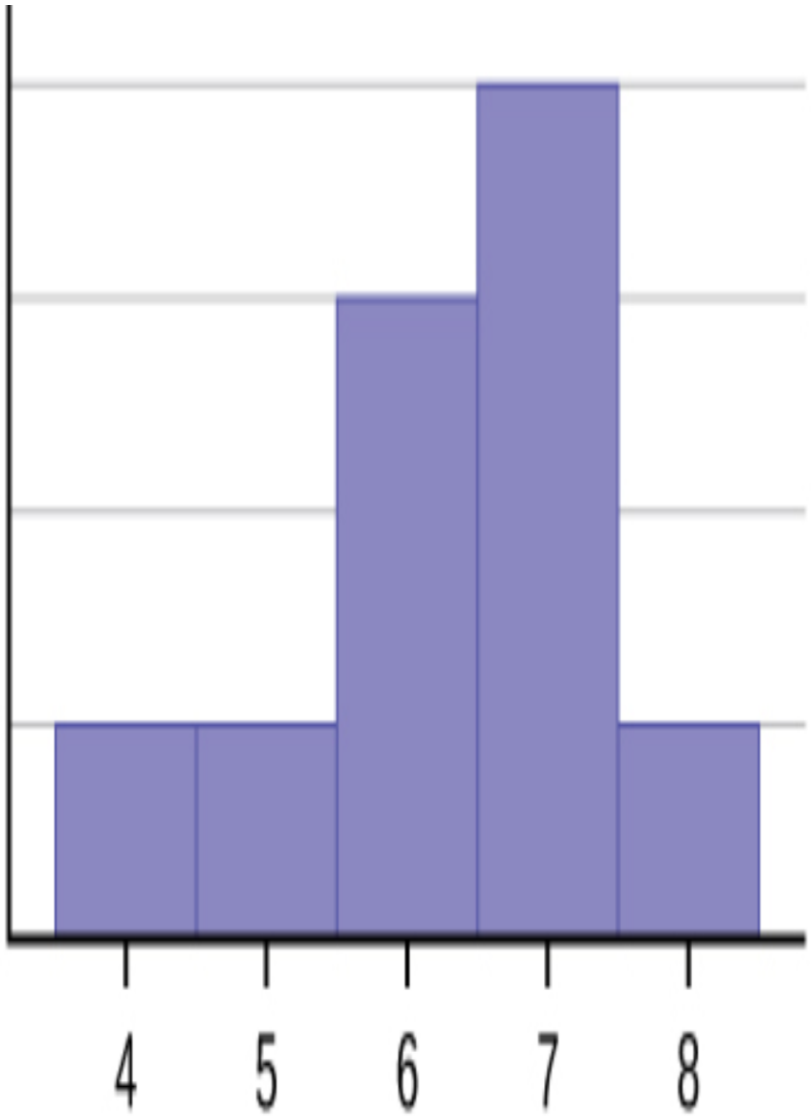
Positive Skewness

- A positively skewed distribution is the distribution with the tail on its right side. The value of skewness for a positively skewed distribution is greater than zero.
- The skewness of the distribution is on the right; it causes the mean to be greater than the median and eventually move to the right. Also, the mode occurs at the highest frequency of the distribution which is on the left side of the median. Therefore, **mode < median < mean**.
- Q2 is present nearer to Q1. This represents a positively skewed distribution.

$$Q_3 - Q_2 > Q_2 - Q_1$$

Negative Skewness

- The histogram for the data: 4,5,6,6,6,7,7,7,7,8 is not symmetrical.
- The right-hand side seems “chopped off” compared to the left side.
- A distribution of this type is called **skewed to the left** because it is pulled out to the left.



Negative Skewness

- **Mean** = 4+5+6+6+6+7+7+7+7+8 = 63 / 10
- The mean is 6.3.
- **Median** = N+ 1 / 2 = 10+1/2 = 11/2 = 5.5 = 6+7/2
- The median is 6.5
- The **mode** is 7.
- Notice that the mean is less than the median, and they are both less than the mode.
- **Mean < Median < Mode**
- The mean and the median both reflect the skewing, but the mean reflects it more so.

Negative Skewness

- A negatively skewed distribution is the distribution with the tail on its left side. The value of skewness for a negatively skewed distribution is less than zero.
- The skewness of the distribution is on the left, it causes the mean to be lesser than the median and eventually move to the left. Also, the mode occurs at the highest frequency of the distribution which is on the right side of the median. Therefore, **mean < median < mode**.
- Q2 is present nearer to Q3. This represents a negatively skewed distribution.

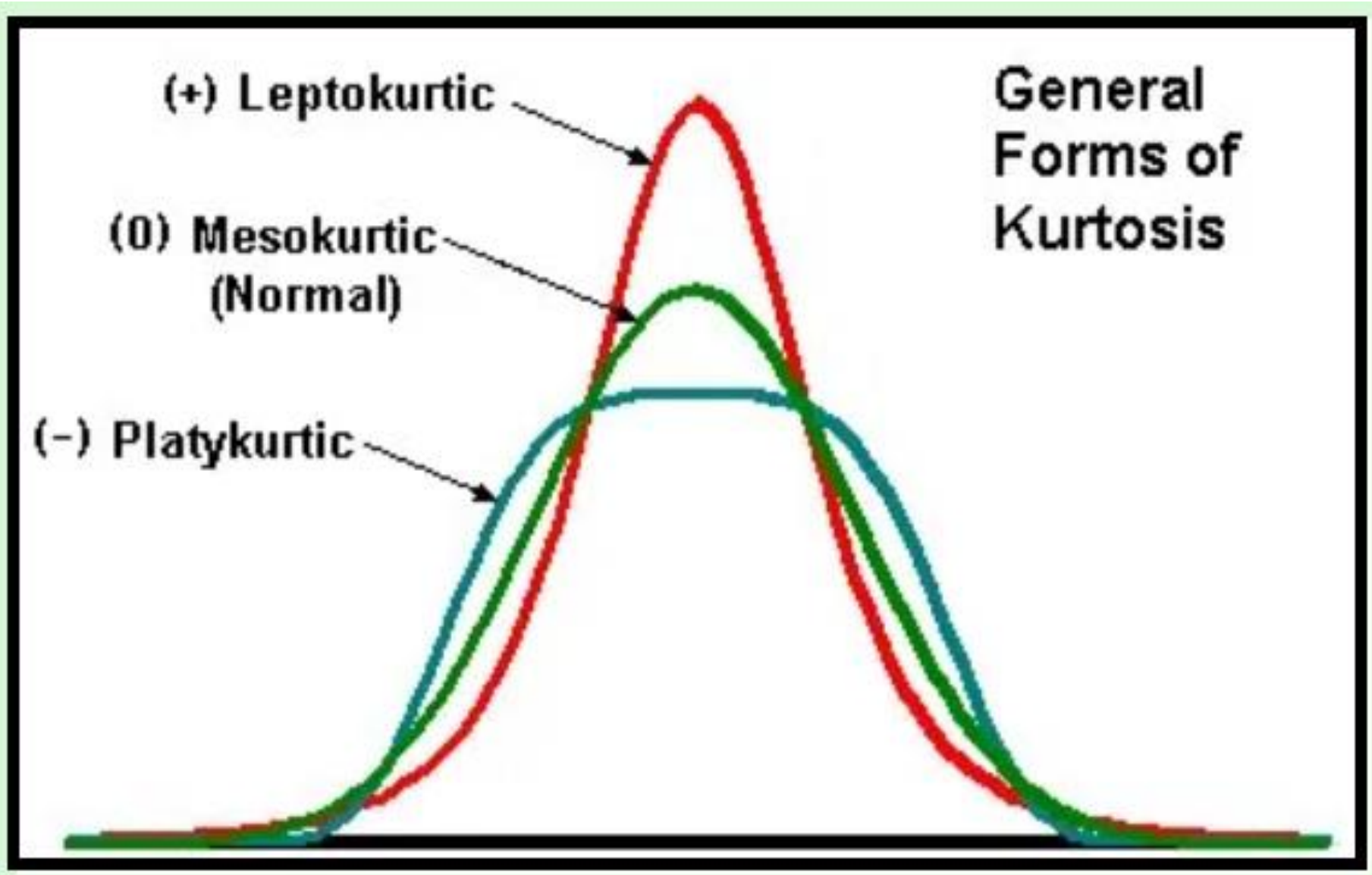
$$Q_3 - Q_2 < Q_2 - Q_1$$

KURTOSIS

- Kurtosis is a statistical measure used to describe the degree to which scores cluster in the tails or the peak of a frequency distribution. The peak is the tallest part of the distribution, and the tails are the ends of the distribution.
- Kurtosis is a statistical measure that is used to describe distribution.
- Kurtosis is a measure of the combined weight of a distribution's tails relative to the center of the distribution.
- Kurtosis is a measure that describes the shape of a distribution's tails in relation to its overall shape.

Types of Kurtosis

- There are three categories of kurtosis that can be displayed by a set of data. All measures of kurtosis are compared against a standard normal distribution, or bell curve.
- 1. Mesokurtic
- 2. Leptokurtic
- 3. Platykurtic

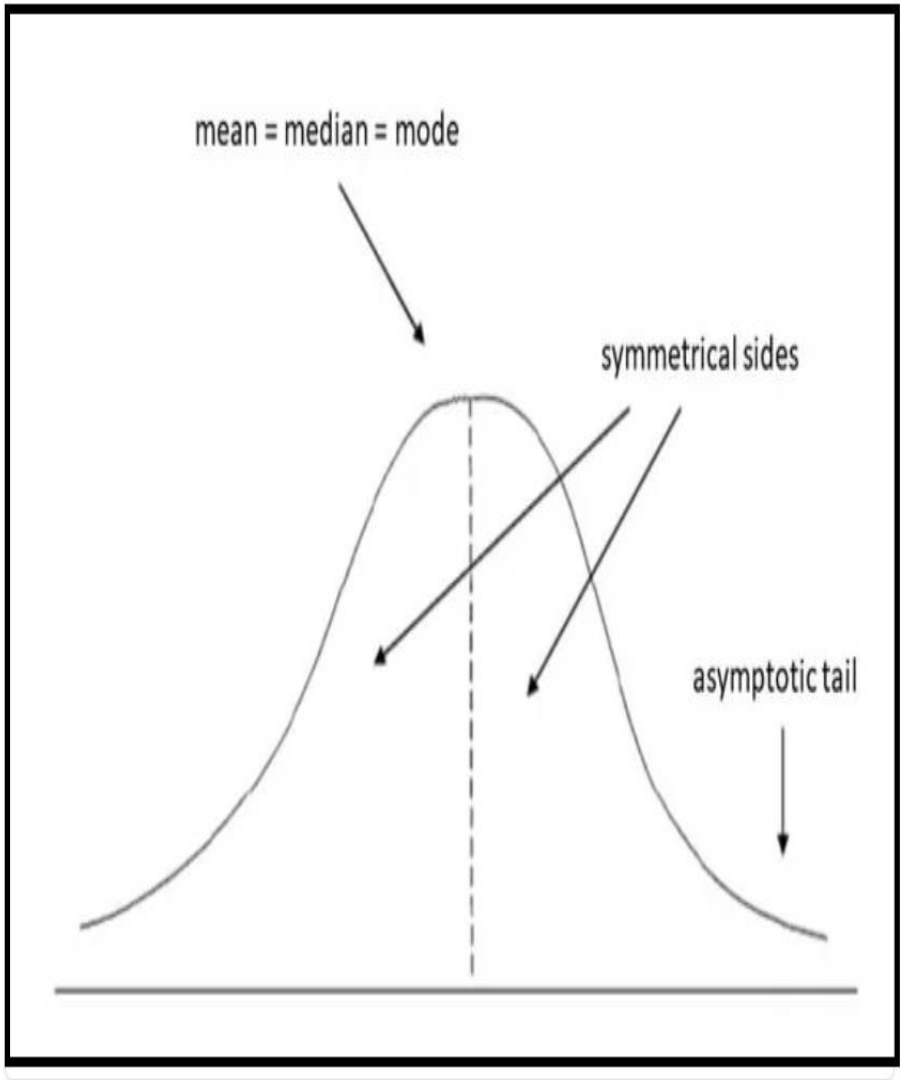

(+) Leptokurtic
(0) Mesokurtic
(Normal)
(-) Platykurtic
General
Forms of
Kurtosis

Mesokurtic

- Mesokurtic: Distributions that are moderate in breadth and curves with a medium peaked height.
- This distribution has a kurtosis statistic similar to that of the normal distribution, meaning the extreme value characteristic of the distribution is similar to that of a normal distribution.
- According to Carl Pearson the value of these distribution is equal to 3.

What does it mean when kurtosis is zero?

- When kurtosis is equal to 0, the distribution is mesokurtic. This means the kurtosis is the same as the normal distribution, it is mesokurtic (medium peak).
- The kurtosis of a mesokurtic distribution is neither high nor low, rather it is considered to be a baseline for the two other classifications.

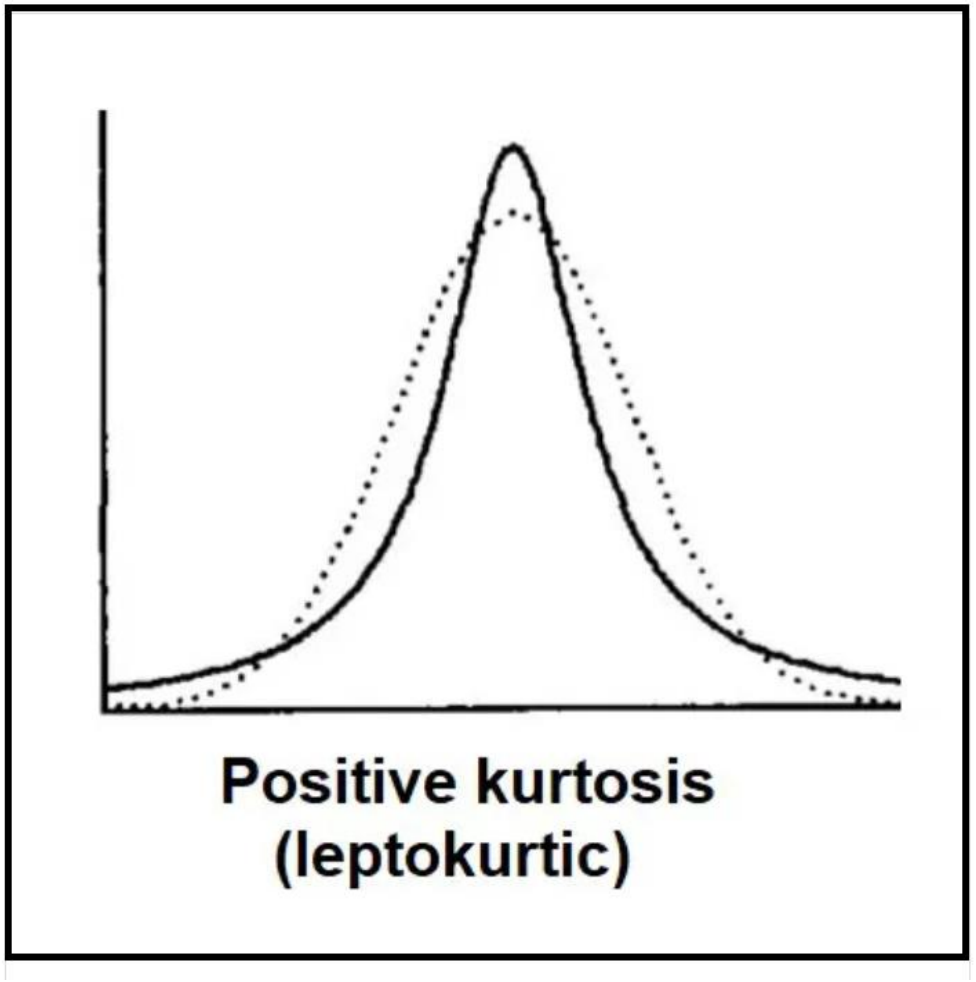


Leptokurtic

- Leptokurtic: leptokurtic distribution is a consequence of the outliers, which stretch the horizontal axis of the histogram graph, making the bulk of the data appear in a narrow ("skinny") vertical range.
- More values in the distribution tails and more values close to the mean (i.e. sharply peaked with heavy tails).
- According to Carl Pearson the value of these distribution is greater than 3.

What does it mean when kurtosis is positive?

- Positive values of kurtosis indicate that a distribution is peaked and possess thick tails. Leptokurtic distributions have positive kurtosis values.
- A leptokurtic distribution has a higher peak and taller (i.e. fatter and heavy) tails than a normal distribution.

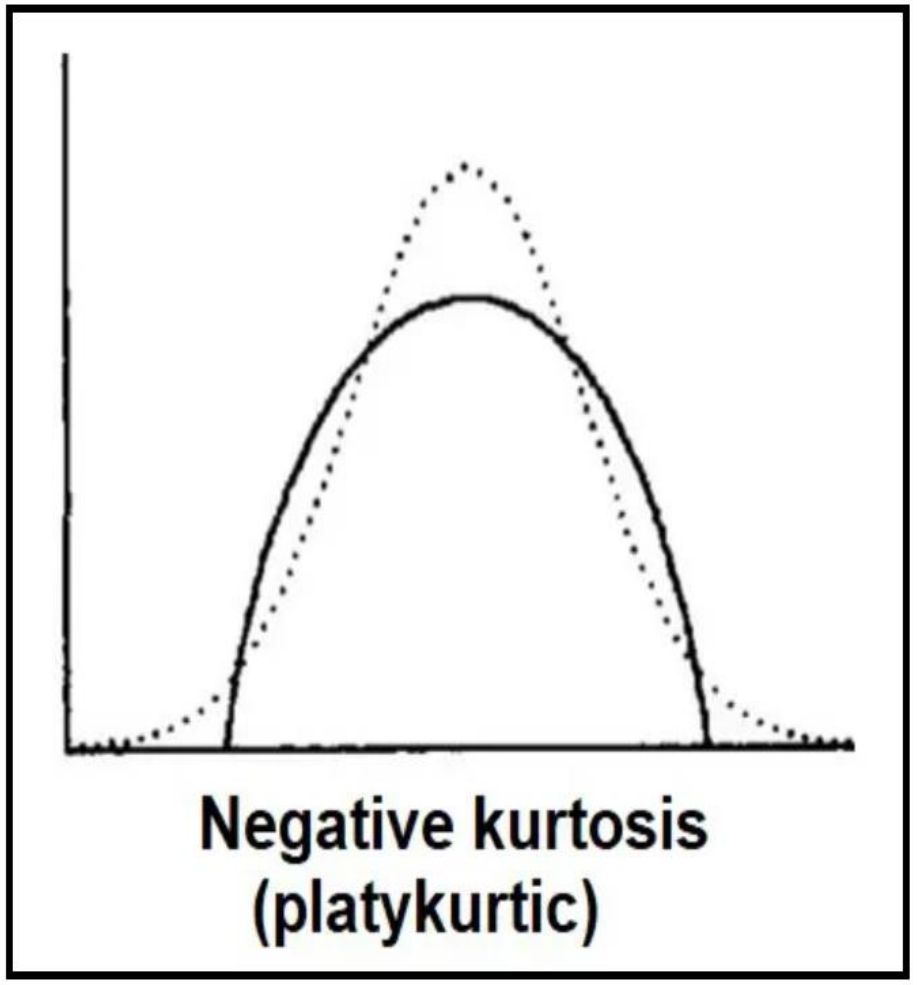


Platykurtic

- Platykurtic: Fewer values in the tails and fewer values close to the mean (i.e. the curve has a flat peak and has more dispersed scores with lighter tails).
- These types of distributions have short tails (paucity of outliers.) The prefix of "platy-" means "broad," and it is meant to describe a short and broad-looking peak. Uniform distributions are platykurtic and have broad peaks.
- According to Carl Pearson the value of these distribution is lesser than 3.

What does it mean when kurtosis is negative?

- Negative values of kurtosis indicate that a distribution is flat and has thin tails. Platykurtic distributions have negative kurtosis values.
- A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter (i.e. lighter and thinner) tails.



THANK YOU

SOURCE:
HTTPS://WWW.SIMPLYPSYCHOLOGY.ORG/KURTOSIS.HTML
HTTPS://WWW.INVESTOPEDIA.COM/TERMS/K/KURTOSIS.ASP