Measures of Shape: Skewness and Kurtosis

The measure of central tendency and measure of dispersion can describe the distribution but they are not sufficient to describe the nature of the distribution.

For this purpose, we use other two statistical measures that compare the shape to the normal curve called Skewness and Kurtosis.

Skewness and Kurtosis are the two important characteristics of distribution that are studied in descriptive statistics

1-Skewness

Skewness is a statistical number that tells us if a distribution is symmetric or not. A distribution is symmetric if the right side of the distribution is similar to the left side of the distribution.

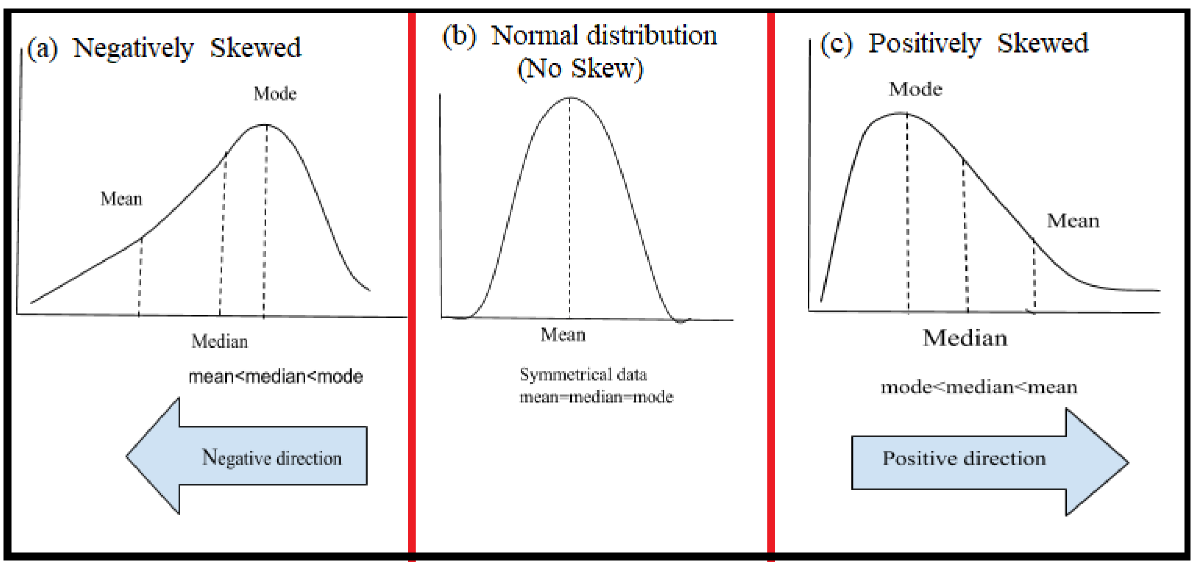
If a distribution is symmetric, then the Skewness value is 0.

i.e. If a distribution is Symmetric (normal distribution):

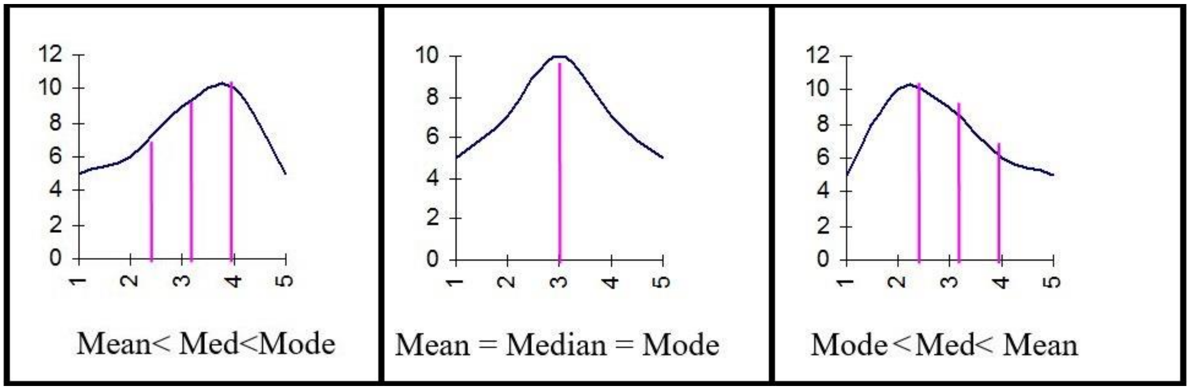
median= mean= mode, (Skewness value is 0)

If Skewness is greater than 0, then it is called right-skewed or that the right tail is longer than the left tail. If Skewness is less than 0, then it is called left-skewed or that the left tail is longer than the right tail.

For example, the symmetrical and skewed distributions are shown by curves as:



And other example



The Formula of Skewness is:

$$\text{Skewness} = \frac{\sum (x - \bar{x})^3}{(n - 1) \cdot S^3}$$

Where:

S: standard deviation

$\overline{X}$: Mean

Difference between Variance and Skewness

The following two points of difference between variance and skewness should be carefully noted.

1. Variance tells us about the amount of variability while skewness gives the direction of variability.

2. In business and economic series, measures of variation have greater practical application than measures of skewness. However, in the medical and life science field measures of skewness have greater practical applications than the variance.

<u>Karl Pearson's Coefficient of Skewness</u>

Karl Pearson developed two methods to find skewness in a sample.

This method is most frequently used for measuring skewness. The formula for measuring coefficient of skewness is given by

1. Pearson's Coefficient of Skewness #1 uses the mode. The formula is:

$$SK = \frac{(\bar{X} - Mo)}{SD}$$

Where $\bar{X}$:the mean,
Mo : the mode
SD: the standard deviation for the sample.

If mode is not well defined, we use the formula

Pearson's Coefficient of Skewness #2 uses the median. The formula is

$$SK = \frac{3(\bar{X} - Md)}{SD}$$

Where $\bar{X}$:the mean,

Mo = the mode,

and SD = the standard deviation for the sample.

It is generally used when you don't know the mode.

The value of this coefficient would be zero in a symmetrical distribution.

If the mean is greater than the mode, the coefficient of skewness would be positive otherwise negative.

The value of Karl Pearson's coefficient of skewness usually lies between ±1 for moderately skewed destitution.

Example:

Use Pearson's Coefficient 1 and 2 to find the skewness for data with the following characteristics:

- Mean = 70.5
- Median = 80
- Mode = 85

- Standard deviation = 19.33

Pearson's Coefficient of Skewness 1 (Mode):

Step 1: Subtract the mode from the mean: 70.5 – 85 = -14.5

Step 2: Divide by the standard deviation: -14.5 / 19.33 = -0.75

Pearson's Coefficient of Skewness 2 (Median):

Step 1: Subtract the median from the mean: 70.5 – 80 = -9.5

Step 2: Multiply Step 1 by 3: -9.5(3) = -28.5

Step 2: Divide by the standard deviation: -28.5 / 19.33 = -1.47

Remarks about Skewness

1. If the value of mean, median, and mode are the same in any distribution, then the skewness does not exist in that distribution. Larger the difference in these values, the larger the skewness;

2. If sum of the frequencies are equal on both sides of the mode then skewness does not exist;

3. If the distance of the first quartile and third quartile are the same from the median then a skewness does not exist. Similarly, if deciles (first and ninth) and percentiles (first and ninety-nine) are at equal distance from the median. Then there is no asymmetry;

4. If the sums of positive and negative deviations obtained from mean, median, or mode are equal then there is no asymmetry; and

5. If a graph of data become a normal curve and when it is folded at the middle and one-part overlap fully on the other one then there is no asymmetry

2-Kurtosis

Kurtosis is a statistical number that tells us if a distribution is taller or shorter than a normal distribution. If a distribution is similar to the normal distribution, the Kurtosis value is 0. If Kurtosis is greater than 0, then it has a higher peak compared to the normal distribution. If Kurtosis is less than 0, then it is flatter than a normal distribution.

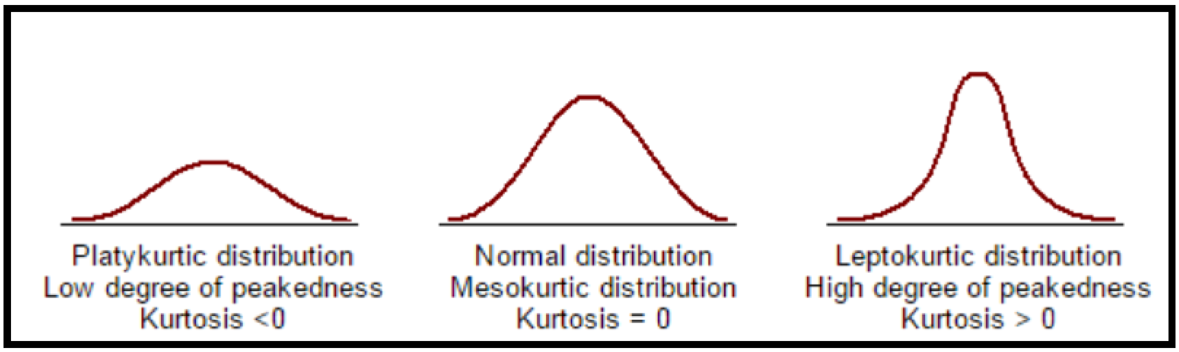
There are three types of distributions:

Leptokurtic: Sharply peaked with fat tails, and less variable.

Mesokurtic: Medium peaked

Platykurtic: Flattest peak and highly dispersed.

For example, The different types of Kurtosis:



The Formula of kurtosis is:

$$\text{Kurtosis} = \frac{\sum (x - \bar{x})^4}{(n - 1) \cdot S^4}$$

Where:

S: standard deviation $\bar{X}$: Mean

Examples: Calculate Sample Skewness and Sample Kurtosis from the following grouped data

Class	Frequency
2 - 4	3
4 - 6	4
6 - 8	2
8 - 10	1

Solution:

Classes	Mid value (x)	f	$f{\cdot}x$	$(x-\bar{x})$	$f{\cdot}(x-\bar{x})^2$	$f{\cdot}(x-\bar{x})^3$	$f{\cdot}(x-\bar{x})^4$
2 - 4	**3**	**3**	**3×3= 9**	**3-5.2=-2.2**	**3×-2.2×-2.2=14.52**	**14.52×-2.2= -31.944**	**70.27**
4 - 6	**5**	**4**	**4×5= 20**	**5-5.2=-0.2**	**4×-0.2×-0.2=0.16**	**0.16×-0.2= -0.032**	**0.0064**
6 - 8	**7**	**2**	**2×7= 14**	**7-5.2=1.8**	**2×1.8×1.8=6.48**	**6.48×1.8=11.664**	**20.98**
8 - 10	**9**	**1**	**1×9= 9**	**9-5.2=3.8**	**1×3.8×3.8=14.44**	**14.44×3.8= 54.872**	**208.5**
---	**---**	**---**	**---**	**---**	**---**	**---**	
-TOTAL-	**--**	***n*=10**	$\sum f{\cdot}x$**=52**	**`--**	**=35.6**	**=34.56**	**=299.79**

$$\text{Mean} = \sum \boldsymbol{f} \cdot \boldsymbol{x} = \frac{\sum f{\cdot}x}{\sum \mathrm{f}} = \frac{52}{10} = 5.2$$

Calculate Standard deviation (S.D)

$$S.D = \sqrt{\frac{\sum_{i=1}^{n} fi(Xi - \bar{X})^2}{\sum_{i=1}^{n} fi}}$$

$$S.D = \sqrt{\frac{35.6}{10}} = 1.88$$

Calculate the Skewness

$$\boxed{\text{Skewness} = \frac{\sum (x - \bar{x})^3}{(n - 1) \cdot S^3}}$$

$$\text{Skewness} = \frac{34.56}{9*(1.88)^{3}} = 0.48$$

Calculate the Kurtosis:

$$\boxed{\text{Kurtosis} = \frac{\sum (x - \bar{x})^4}{(n - 1) \cdot S^4}}$$

$$\text{Kurtosis} = \frac{299.79}{9*(1.88)^{4}} = 2.12$$

Key Differences Between Skewness and Kurtosis

This is the fundamental differences between skewness and kurtosis:

1- The characteristic of a frequency distribution that ascertains its symmetry about the mean is called skewness. On the other hand, Kurtosis means the relative pointedness of the standard bell curve, defined by the frequency distribution.
2- Skewness is a measure of the degree of lopsidedness in the frequency distribution. Conversely, kurtosis is a measure of degree of tailedness in the frequency distribution.
3- Skewness is an indicator of lack of symmetry, i.e. both left and right sides of the curve are unequal, with respect to the central point. As against this, kurtosis is a measure of data, that is either peaked or flat, with respect to the probability distribution.
4- Skewness shows how much and in which direction, the values deviate from the mean? In contrast, kurtosis explain how tall and sharp the central peak is.