Skewness and Kurtosis

Skewness and Kurtosis are coefficient that measure how different a distribution is from a normal distribution. Skewness measures the symmetry of a normal distribution while Kurtosis measure the thickness of the tail ends relative to the tails of a normal distribution.

A distribution is said to be skewed if;

- Mean, Median, Mode fall at different points ie, Mean $\neq$ Median $\neq$ Mode;
- Quartiles are not equidistant from median;
- The curve drawn with the help of given data is not symmetrical but stretched more to one side than to other.

Measures of skewness

Various measures of skewness are;

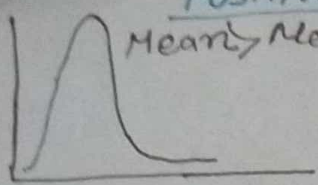
$$S_K = M - M_d$$

$$S_K = M - M_o$$

$$S_K = (Q_3 - M_d) - (M_d - Q_1)$$ where Q_1 and Q_3 are quartiles.

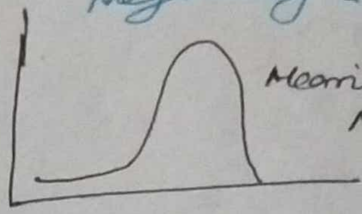
Graphs

1) Positively skewed



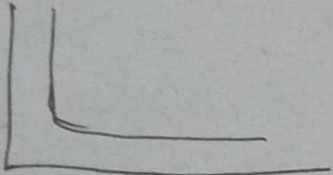
Mean > Median > Mode

2) Negatively skewed



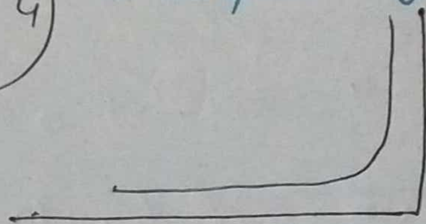
Mean < Median < Mode

3) 'L' Shaped Positively skewed



Mean < Mode
Mean < Median

4) 'J' Shaped Negatively skewed



Mean > Mode
Mean > Median.

Measures Of Skewness

There are four measures of skewness. The measures of skewness are —

Karl Pearson's Coefficient of skewness
Bowley's Coefficient of skewness
Kelly's Coefficient of skewness

Karl Pearson's ~~Coeffi~~ Coefficient of skewness:—

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

where, Sk_p = Karl Pearson's Coefficient of skewness.

σ = standard deviation.

In case the mode is indeterminate, the coefficient of skewness is

$$Sk_p = \frac{\text{Mean} - (3\text{Median} - 2\text{Mean})}{\sigma}$$

Now this formula is equal to —

$$Sk_p = \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

The value of coefficient of skewness is zero, when the distribution is symmetrical.

Normally, this coefficient of skewness lies between +1. If the mean is greater than the mode, then the coefficient of skewness will be positive, otherwise negative

Bowley's Coefficient of skewness

Bowley developed a measure of skewness, which is based on quartile values. The formula for measuring skewness is

$$SK_B = \frac{(Q_3 - Q_2)(Q_2 - Q_1)}{(Q_3 - Q_1)}$$

SK_B = Bowley's coefficient skewness

Q_1 = Quartile first

Q_2 = Quartile second

Q_3 = " third

To above formula can be converted to

$$SK_B = \frac{Q_3 + Q_1 - 2\,Median}{(Q_3 - Q_1)}$$

The value of coefficient of skewness is zero, if it is a symmetrical distribution. if the value is greater than zero, it is positively skewed and if the value is less than zero, it is negatively skewed distribution

MOMENTS

In mechanics, the term moment is used to denote the rotating effect of a force. In statistics, it is used to indicate peculiarities of a frequency distribution. The utility of moments lies in the sense that they indicate different aspects of a given distribution. Thus, by using moments, we can measure the central tendency of a series, dispersion or variability, skewness and the peakedness of the curve. The moments about the actual arithmetic mean are denoted by μ. The first four moments about mean or central moments are as follows:

<u>In case of ungrouped data</u>

First moment $\mu_1 = \frac{1}{N} \Sigma (x_1 - \bar{x})$

Second moment $\mu_2 = \frac{1}{N} \Sigma (x_1 - \bar{x})^2$ [Similar as variance]

Third moment $\mu_3 = \frac{1}{N} \Sigma (x - \bar{x})^3$

Fourth moment $\mu_4 = \frac{1}{N} (x - \bar{x})^4$

<u>In case of grouped data</u>

First moment $\mu_1 = \frac{1 \Sigma f}{N} (x_1 - \bar{x})$

Second moment $\mu_2 = \frac{1}{N} \Sigma f (x_1 - \bar{x})^2$

Third moment $\mu_3 = \frac{1}{N} \Sigma f (x - \bar{x})^3$

Fourth moment $\mu_4 = \frac{1}{N} \Sigma f (x - \bar{x})^4$

Two important constants calculated from μ_2, μ_3 and μ_4 are-

β_1 (read as beta one)

β_2 (read as beta two)

β_1 (Beta one) – Beta one is defined as

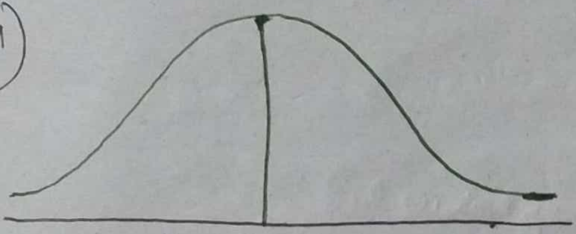
β_1 is used to measures of skewness. In symmetrical distribution β_1 shall be zero. However, the coefficient β_1 as a measure of skewness has serious limitations. β_1 as a measure of skewness cannot tell us about the direction of skewness that is whether it is positive or negative.

This is for the simple reason that μ_3 being the sum of the cubes of the deviation from the mean may be positive or negative but μ_3^2 is always positive.

Hence,

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ is always positive.}$$

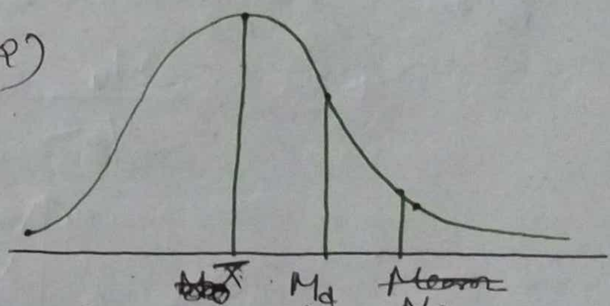
1)



$\bar{x}$ (Mean) = Mo = Md

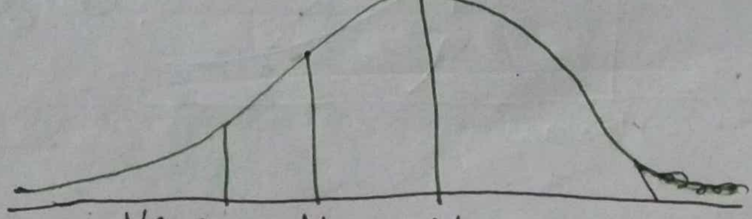
(Symmetric Distribution)

2)



(Positively skewed Distribution)

3.)



Negatively skewed Distribution

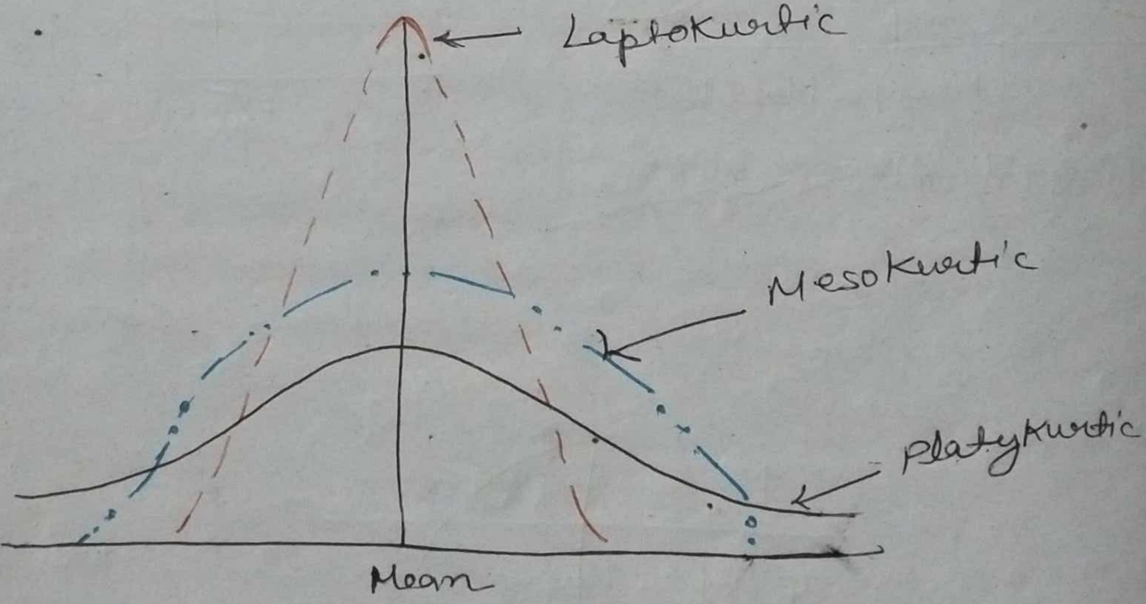
Kurtosis

Kurtosis is another measure of the shape of a frequency curve. It is a ~~Gre~~ Greek word, which means bulginess.
While skewness signifies the extent of asymmetry, kurtosis measures the degree of peakedness of a frequency distribution. Karl Pearson classified curves into three types on the basis of the shape of their peaks.
These are Mesokurtic, leptokurtic and ~~platy~~ platykurtic. These three types of curves are shown in figure below:

β_2 (Beta two)

Beta two measures Kurtosis and is defined as—

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$



Measures of Kurtosis

Kurtosis is measured by β_2, or its its ~~deri~~ derivative γ_2. Beta two measures Kurtosis and is ~~too~~ defined as:

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

And

$$\gamma_2 = \beta_2 - 3$$

In case of a normal distribution, that is, mesokurtic curve, the value of $\beta_2 = 3$

If β_2 turn out to be greater than 3, the curve is called a leptokurtic curve and is more peaked than the normal curve.

when β_2 is less than 3, the curve is called a platykurtic curve and is less peaked than the normal curve.

The measure of kurtosis is very helpful in the selection of an appropriate average. For example, for normal distribution, median is most appropriate; and for ~~platy~~ platykurtic distribution, the quartile range is most appropriate

```
from scipy.stats import skew
data = [88, 85, 82, 97, 67, 77, 74, 86, 81, 95, 77,
        88, 85, 76, 81]
skew (data, bias = False)
  0.0326967

from scipy.stats import kurtosis
kurtosis (data, bias = False)
  0.11757
```