SKEWNESS
&
KURTOSIS

Munmun Akter
Lecturer
Dept of NFE, DIU

Concept of Skewness

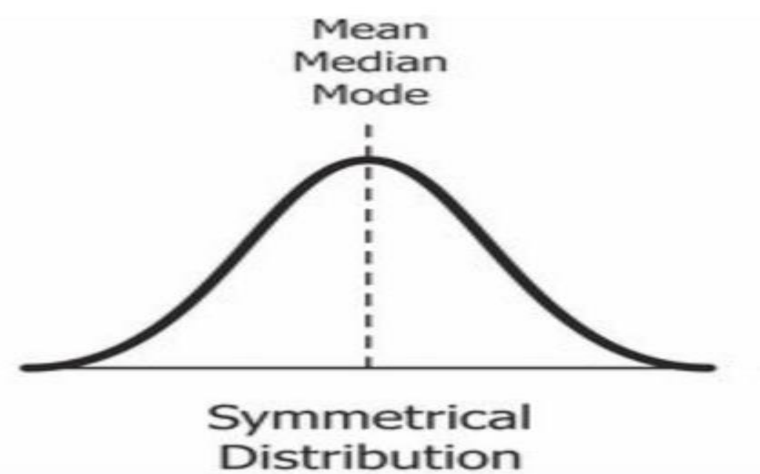
A distribution is said to be skewed-when the mean, median and mode fall at different position in the distribution and the balance (or center of gravity) is shifted to one side or the other i.e. to the left or to the right.

Therefore, the concept of skewness helps us to understand the relationship between three measures-

- **Mean.**
- **Median.**
- **Mode.**

Symmetrical Distribution

- A frequency distribution is said to be symmetrical if the frequencies are equally distributed on both the sides of central value.
- A symmetrical distribution may be either bell – shaped or U shaped.
- In symmetrical distribution, the values of mean, median and mode are equal i.e. **Mean=Median=Mode**



Skewed Distribution

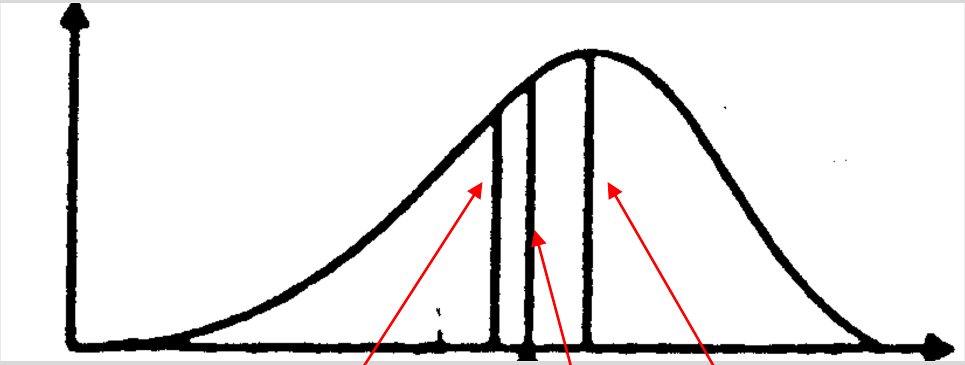
- A frequency distribution is said to be skewed if the frequencies are not equally distributed on both the sides of the central value.

- A skewed distribution may be-

  - **Positively Skewed**
  - **Negatively Skewed**

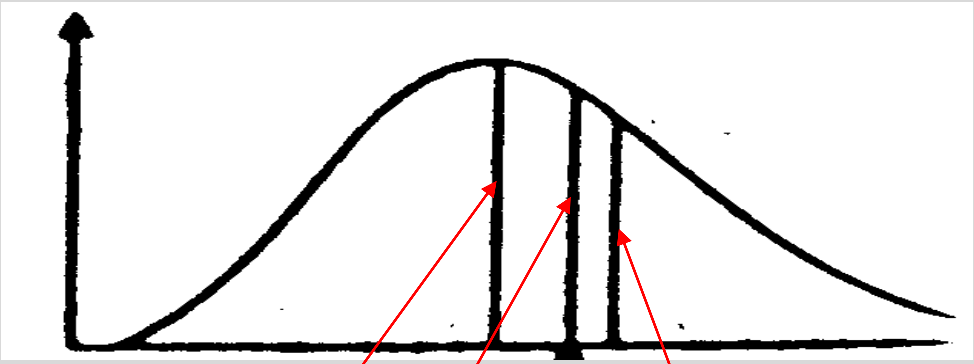
Skewed Distribution

- **<u>Negatively Skewed</u>**
- In this, the distribution is skewed to **the left (negative)**
- Here, **Mode** exceeds Mean and Median.



Mean<Median<Mode

- **<u>Positively Skewed</u>**
- In this, the distribution is skewed to **the right (positive)**
- Here, **Mean** exceeds Mode and Median.



Mode<Median<Mean

Tests of Skewness

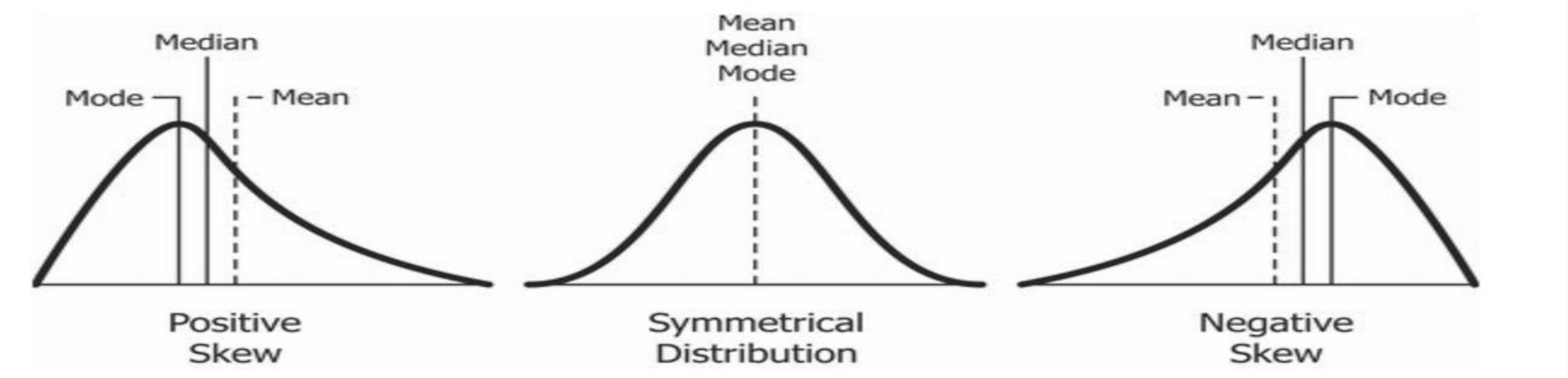
In order to ascertain whether a distribution is skewed or not the following tests may be applied. Skewness is present if:

- The values of mean, median and mode do not coincide.
- When the data are plotted on a graph they do not give the normal bell shaped form i.e. when cut along a vertical line through the center the two halves are not equal.
- The sum of the positive deviations from the median is not equal to the sum of the negative deviations.
- Quartiles are not equidistant from the median.
- Frequencies are not equally distributed at points of equal deviation from the mode.

Graphical Measures of Skewness

- Measures of skewness help us to know to what degree and in which direction (positive or negative) the frequency distribution has a departure from symmetry.
- Positive or negative skewness can be detected graphically (as below) depending on whether the right tail or the left tail is longer but, we don’t get idea of the magnitude
- Hence some statistical measures are required to find the magnitude of lack of symmetry

Mean> Median> Mode **Mean=Median=Mode** **Mean<Median<Mode**



Statistical Measures of Skewness

Absolute Measures of Skewness	Relative Measures of Skewness
Following are the absolute measures of skewness: • Skewness (Sk) = Mean – Median • Skewness (Sk) = Mean – Mode • Skewness (Sk) = (Q3 - Q2) - (Q2 - Q1)	There are four measures of skewness: • β and γ Coefficient of skewness • Karl Pearson's Coefficient of skewness • Bowley’s Coefficient of skewness • Kelly’s Coefficient of skewness

β and γ Coefficient of Skewness

- Karl Pearson defined the following **β and γ coefficients** of skewness, based upon the second and third central moments:-

$$\boldsymbol{\beta_1} = \frac{\mu_3^2}{\mu_2^3}$$

- $\boldsymbol{\beta_1}$ as a measure of skewness does not tell about the direction of skewness, i.e. positive or negative.
- This drawback is removed if we calculate Karl Pearson's Gamma coefficient $\boldsymbol{\gamma_1}$ which is the square root of $\boldsymbol{\beta_1}$ ie

$$\boldsymbol{\gamma_1} = \pm\sqrt{\boldsymbol{\beta_1}} = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{\mu_3}{\sigma^3}$$

Karl Pearson's Coefficient of Skewness……01

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

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

Where,

SK_P = Karl Pearson's Coefficient of skewness,

σ = standard deviation.

Normally, this coefficient of skewness lies between -3 to +3.

Karl Pearson's Coefficient of Skewness…..02

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**.
The value of coefficient of skewness is **positive**, when the distribution is **positively skewed**.
The value of coefficient of skewness is **negative**, when the distribution is **negatively skewed**.

Bowley's Coefficient of Skewness……01

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

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

Where,

SK_B = Bowley's Coefficient of skewness,

Q_1 = Quartile first Q_2 = Quartile second

Q_3 = Quartile Third

Bowley's Coefficient of Skewness…..02

The above formula can be converted to-

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

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** distribution.
And if the value is **less than zero**, it is **negatively skewed** distribution.

Kelly's Coefficient of Skewness…..01

Kelly developed another measure of skewness, which is based on percentiles and deciles.

The formula for measuring skewness is based on percentile as follows:

$$SK_k = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

Where,

SK_K = Kelly's Coefficient of skewness,

P_{90} = Percentile Ninety.

P_{50} = Percentile Fifty.

P_{10} = Percentile Ten.

Kelly’s Coefficient of Skewness…..02

This formula for measuring skewness is based on percentile are as follows:

$$SK_k = \frac{D_9 - 2D_5 + D_1}{D_9 - D_1}$$

Where,

SK_K = Kelly’s Coefficient of skewness,

D_9 = Deciles Nine.

D_5 = Deciles Five. D_1 = Deciles one.

Example:

Question: For a distribution Karl Pearson's coefficient of skewness is 0.64, standard deviation is 13 and mean is 59.2 Find mode and median.

Solution: We have given

S_k = 0.64, σ = 13 and Mean = 59.2

Therefore by using formula

$$S_k = \frac{Mean - Mode}{\sigma}$$

$$0.64 = \frac{59.2 - Mode}{13}$$

Mode = 59.20 – 8.32 = 50.88
Mode = 3 Median – 2 Mean
50.88 = 3 Median - 2 (59.2)

$$\text{Median} = \frac{50.88+118.4}{3} = \frac{169.28}{3} = 56.42$$

Homework:

- **Ques:** The following are the marks of 150 students in an examination. Calculate Karl Pearson's coefficient of skewness.

Marks	No. of Students
0-10	20
10-20	10
20-30	40
30-40	0
40-50	15
50-60	20
60-70	15
70-80	10
80-90	30

Moments:

- In Statistics, moments 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 following:-

Moments:

Moments around Mean

For ungrouped data, $\mu_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^r$

For grouped data, $\mu_r = \frac{1}{n}\sum_{i=1}^{n} f_i(x_i - \overline{x})^r$

where $n = \sum_{i=1}^{k} f_i$ and $\overline{x} = \sum_{i=1}^{k} f_i x_i$

Moments around any Arbitrary No

For ungrouped data, $\mu_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r$

For grouped data, $\mu_r' = \frac{1}{n}\sum_{i=1}^{n} f_i x_i^r$

where, $n = \sum_{i=1}^{k} f_i$

Conversion formula for Moments

1st moment: $\mu_1 = 0$ **(Mean)**

2nd moment: $\mu_2 = \mu_2' - \mu_1'^2$ **(Variance)**

3rd moment: $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$ **(Skewness)**

4th moment: $\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$ **(Kurtosis)**

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 is used to measures of skewness.
- It is defined as:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

- β_2 is used to measures Kurtosis.
- It is defined as:

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

Kurtosis

- Kurtosis is another measure of the shape of a frequency curve. It is a 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:-
  - **Leptokurtic**
  - **Mesokurtic**
  - **Platykurtic**

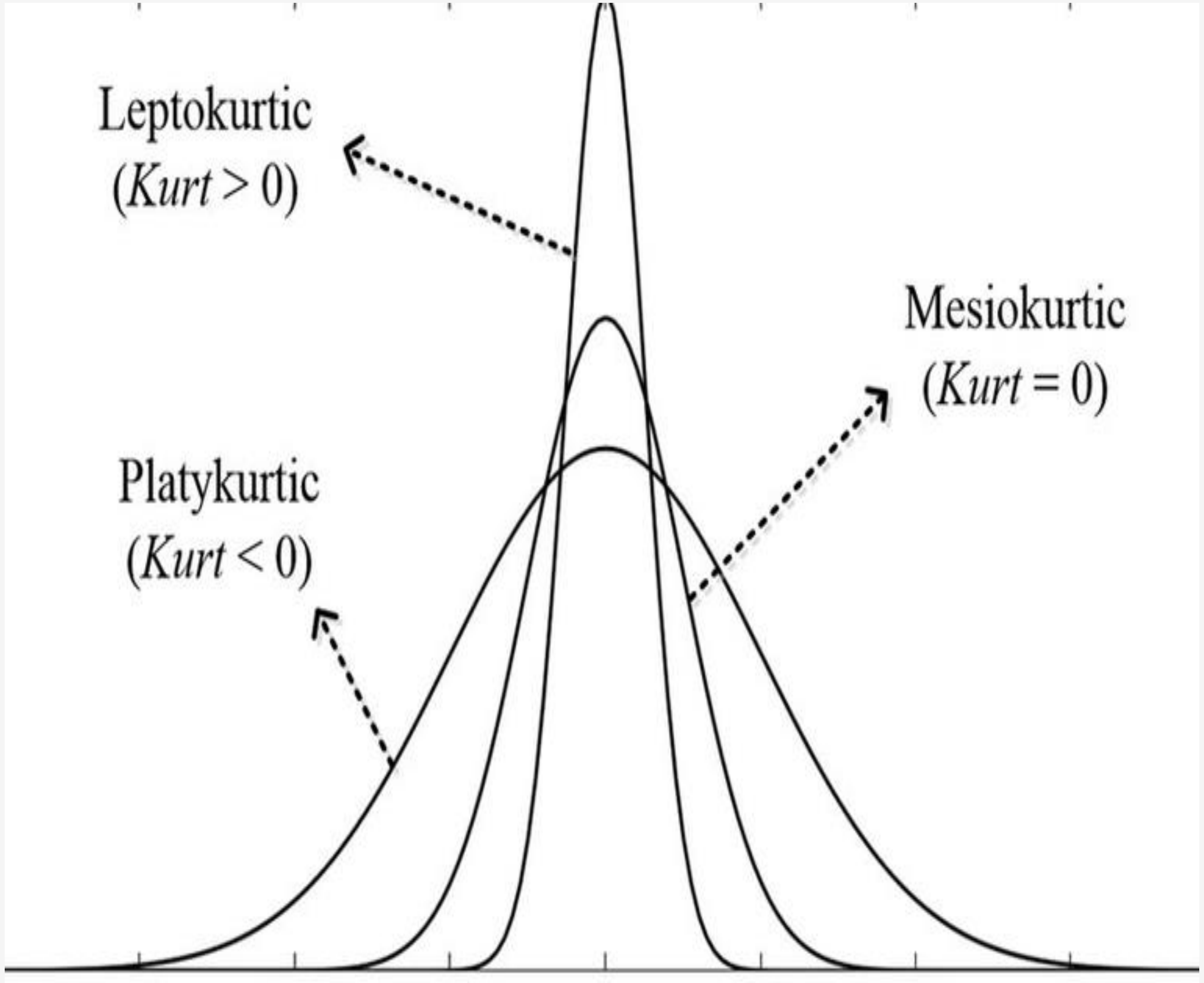
Kurtosis

- When the peak of a curve becomes relatively high then that curve is called **Leptokurtic.**

- When the curve is flat-topped, then it is called **Platykurtic**.

- Since normal curve is neither very peaked nor very flat topped, so it is taken as a basis for comparison.

- This normal curve is called **Mesokurtic**.



Measure of Kurtosis

- There are two measure of Kurtosis:

- **Karl Pearson’s Measures of Kurtosis**

- **Kelly’s Measure of Kurtosis**

Karl Pearson's Measures of Kurtosis

Formula

- For calculating the kurtosis, the second and fourth central moments of variable are used
- For this, following formula given by Karl Pearson is used:

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

- Or

$$\textbf{Kurtosis } (\gamma 2) = \left(\frac{\mu_4}{\mu_2^2}\right) - 3$$

where,

$\boldsymbol{\mu_2}$ = Second order central moment of distribution
$\boldsymbol{\mu_4}$ = Fourth order central moment of distribution

Result:

- If $\boldsymbol{\beta_2} = 3$ or $\boldsymbol{\gamma 2} = 0$, then curve is said to be mesokurtic;
- If $\boldsymbol{\beta_2} < 3$ or $\boldsymbol{\gamma 2} < 0$, then curve is said to be platykurtic;
- If $\boldsymbol{\beta_2} > 3$ or $\boldsymbol{\gamma 2} > 0$, then curve is said to be leptokurtic;

Kelly's Measure of Kurtosis

Formula	Result:

- Kelly has given a measure of kurtosis based on percentiles.
- The formula is given by :-

$$\beta_2 = \frac{P_{75} - P_{25}}{P_{90} - P_{10}}$$

where,

P_{75}, P_{25}, P_{90}, and P_{10} are 75th, 25th , 90th and 10th percentiles of dispersion respectively.

Result:

- If $\beta_2 > 0.26315$, then the distribution is platykurtic.
- If $\beta_2 < 0.26315$, then the distribution is leptokurtic.

Example:

- **Ques:** First four moments about mean of a distribution are 0, 2.5, 0.7 and 18.75. Find coefficient of skewness and kurtosis.
- **Sol:** We have $\boldsymbol{\mu_1} = 0$, $\boldsymbol{\mu_2} = 2.5$, $\boldsymbol{\mu_3} = 0.7$ and $\boldsymbol{\mu_4} = 18.75$

Therefore, Skewness,

$$\boldsymbol{\beta_1} = \frac{\mu_3^2}{\mu_2^3} = \frac{\mathbf{0.7^2}}{\mathbf{2.5^3}} = 0.031$$

and Kurtosis is

$$\boldsymbol{\beta_2} = \frac{\mu_4}{\mu_2^2} = \frac{\mathbf{18.75}}{\mathbf{2.5^2}} = \frac{\mathbf{18.75}}{\mathbf{6.25}} = 3$$

As $\boldsymbol{\beta_2}$ is equal to 3, so the curve is mesokurtic

Homework:

- **Ques:** The first four raw moments of a distribution are 2, 136, 320, and 40,000. Find out coefficients of skewness and kurtosis.