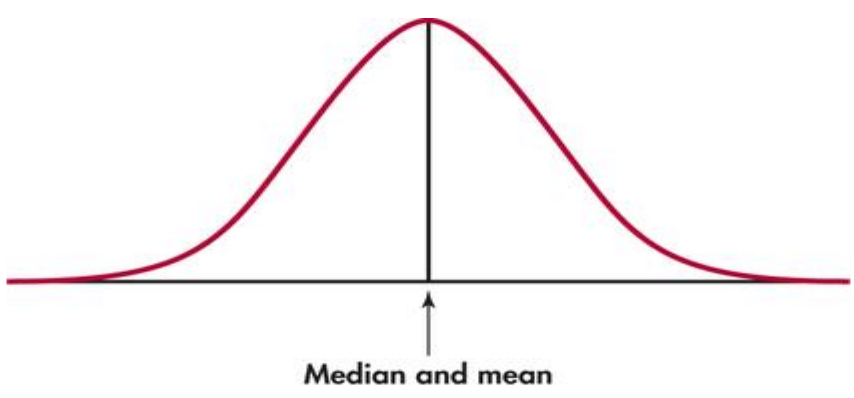
Lecture 5

SKEWNESS AND KURTOSIS

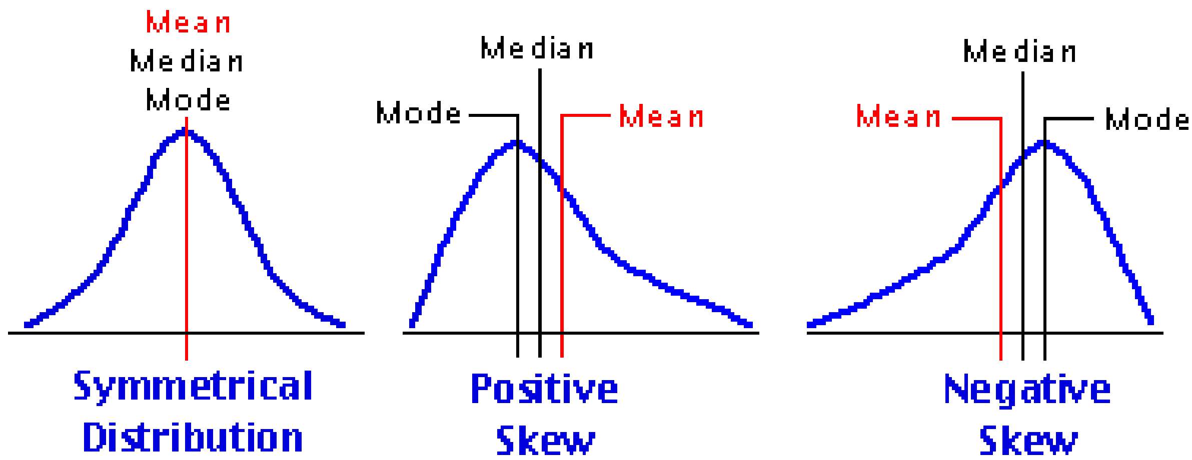
- In addition to measures of central tendency and dispersion, we also need to have an **idea about the shape of the distribution**.
- Skewness is a **measure of asymmetry**. Kurtosis is a **more intelligent measure of peakedness** compared to a **Gaussian (normal) distribution.**
- Measure of skewness gives the **direction and the magnitude of the lack of symmetry** whereas the kurtosis gives the idea of **flatness/peakedness**.
- If the distribution is not symmetric, the frequencies will not be uniformly distributed about the center of the distribution.

CONCEPT OF SKEWNESS

- Skewness means lack of symmetry.
- In mathematics, a figure is called symmetric if **there exists a point in it through which if a perpendicular is drawn on the X-axi**s, it **divides** the figure into **two congruent parts** i.e. identical in all respect or one part can be superimposed on the other i.e mirror images of each other.
- In Statistics, a distribution is called symmetric if **mean, median and mode coincide.** Otherwise, the **distribution becomes asymmetric**.
- If the **right tail is longer**, we get a **positively skewed distribution** for which **mean > median > mode** while **if the left tail is longer**, we get a **negatively skewed distribution** for which **mean < median < mode**.
- The example of the Symmetrical curve, Positive skewed curve and Negative skewed curve are given as follows:



Mean = Median = Mode



Symmetric	**Positive Skewed Curve**	**Negative Skewed Curve**
Mean = Median = Mode	**Mean > Median > Mode**	**Mean < Median < Mode**

VARIOUS 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.
- Although **positive or negative skewness** can be **detected** graphically depending on whether the right tail or the left tail is longer but, we don't get **idea of the magnitude**.
- Besides, borderline cases between symmetry and asymmetry may be difficult to detect graphically.
- Hence some **statistical measures are required** to find the **magnitude** of lack of symmetry.

A good measure of skewness should possess three criteria:

1. It should be **a unit free number** so that the shapes of different distributions, so far as symmetry is concerned, **can be compared** even if the unit of the underlying variables are different;
2. If the distribution is symmetric, the **value of the measure should be zero.** Similarly, the measure should give **positive or negative** values according as the distribution has **positive or negative skewness** respectively; and

3. As we move from **extreme negative skewness to extreme positive skewness**, the value of the measure should **vary accordingly**.

Measure of Skewness:

The degree of skewness in a distribution can be classified as follows:

i. Absolute measure of skewness
ii. Relative measure of skewness

Absolute Measures of Skewness

Following are the absolute measures of skewness:

1. Skewness (Sk) = Mean – Median
2. Skewness (Sk) = Mean – Mode
3. Skewness (Sk) = $(Q_3 - Q_2) - (Q_2 - Q_1)$

- For comparing two series, we do not calculate these absolute measures, we calculate the relative measures which are called **coefficient of skewness**.
- Coefficient of skewness are pure numbers, **independent of units** of measurements.

Relative Measures of Skewness

- In order to make valid comparison between the skewness of two or more distributions we have to eliminate the distributing influence of variation.
- Such elimination can be done by dividing the absolute skewness by standard deviation.

The following are the important methods of measuring relative skewness:

Karl Pearson's Coefficient of Skewness

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

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

- The value of this coefficient would be zero in a symmetrical distribution.
- If mean is greater than mode, coefficient of skewness would be positive otherwise negative.

- As a general rule of thumb:
  - ✓ If skewness is less than -1 or greater than 1, the distribution is highly skewed.
  - ✓ If skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed.
  - ✓ If skewness is between -0.5 and 0.5, the distribution is approximately symmetric.

If mode is not well defined, we use the formula

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

By using the relationship Mode = (3 Median – 2 Mean). Here, $-3 \leq S_k \leq 3$.

Bowleys's Coefficient of Skewness

This method is based on quartiles. The formula for calculating coefficient of skewness is given by

$$S_k = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \frac{(Q_3 - 2Q_2 + Q_1)}{(Q_3 - Q_1)}$$

where,

$$Q_i = l + \frac{h}{f}\left(\frac{in}{4} - c\right)$$

l = Lower limit of the ith quartile class,
h = width of the ith quartile class,
f = frequency of the ith quartile class
c = cumulative frequency of the class preceding the ith quartile class
n = total frequency

- ✓ The value of Sk would be 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.
- ✓ It will take value between +1 and -1.

EXAMPLE: Calculate arithmetic measures of skewness of data of days to maturity 40 short-term investments.

Table: Data of days to maturity 40 short-term investments.

Class interval	No. of investments
30—39	3
40—49	1
50—59	8
60—69	10
70—79	7
80—89	7
90—99	4
Total	40

Solution:

Mean = 68
Mode = 64
Standard deviation = 16.42
$Q_1 \quad = 57.5$
$Q_2 \quad = 68$
$Q_3 \quad = 81.4$

Karl Pearson's Coefficient of Skewness is given by

$$S_k = \frac{\text{Mean – Mode}}{\text{Standrd dviation}} = \frac{68 - 64}{16.42} = \frac{4}{16.42} = 0.24$$

The value of $S_k = 0.12$ (between -.5 to 0,5) indicates that the distribution is approximately symmetric.

Bowleys's Coefficient of Skewness

$$S_k = \frac{(Q_3 - 2Q_2 + Q_1)}{(Q_3 - Q_1)} = \frac{81.4 - 2(68) + 57.5}{81.4 - 57.5} = \frac{138.9 - 136}{23.9} = \frac{2.9}{23.9} = 0.12$$

The value of $S_k = 0.12$ (between -.5 to 0,5) indicates that the distribution is approximately symmetric.

Inferring

- Only one sample from a population was used to create our data set.
- Even though the population is symmetric, it's possible that this sample is skewed due to normal sample variability.
- However, one can infer that the population is skewed if the sample is too skewed for random chance to be the explanation.
- But what do we mean by "too much for random chance to be the explanation"? To answer that, you need to divide the sample skewness S_k by the **standard error of skewness (SES)** to get the **test statistic**, which measures how many standard errors separate the sample skewness from zero:

$$SES = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

$$Test\ Statistic: Z_{S_k} = S_k/SES$$

The critical value of Z_{S_k} is approximately 2. (This is a two-tailed test of skewness $\neq 0$ at roughly the 0.05 significance level.)

- **If $Z_{S_k} < -2$**, the population is very likely skewed negatively (though you don't know by how much).
- **If Z_{S_k} is between -2 and $+2$**, you can't reach any conclusion about the skewness of the population: it might be symmetric, or it might be skewed in either direction.
- **If $Z_{S_k} > 2$**, the population is very likely skewed positively (though you don't know by how much).

- Don't mix up the meanings of this test statistic and the amount of skewness.
- The amount of skewness tells you how highly skewed your sample is: the bigger the number, the bigger the skew.
- The test statistic tells you whether the whole population is probably skewed, but not by how much: the bigger the number, the higher the probability.

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:

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

where

$$\mu_2 = \frac{\sum(x_i - \bar{x})^2}{n}$$

and

$$\mu_3 = \frac{\sum(x_i - \bar{x})^3}{n}$$

- It is used as measure of skewness.
- For a symmetrical distribution, β_1 shall be zero.
- β_1 as a measure of skewness does not tell about the direction of skewness, i.e. positive or negative. Because μ_3 being the sum of cubes of the deviations from mean may be positive or negative but μ_3^2 is always positive. Also, μ_2 being the variance always positive. Hence, β_1 would be always positive. This drawback is removed if we calculate Karl Pearson's Gamma coefficient $\gamma_1$which is the square root of β_1 i. e.

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

- Then the sign of skewness would depend upon the value of μ_3 whether it is positive or negative. It is advisable to use γ_1 as measure of skewness.

Example: 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

Sk = 0.64, σ = 13 and Mean = 59.2

Therefore, by using formulae

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

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

$$Mode = 59.2 - (0.64)(13) = 59.2 - 8.32 = 50.88$$

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

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

CONCEPT OF KURTOSIS

- If we have the knowledge of the measures of central tendency, dispersion and skewness, even then we cannot get a complete idea of a distribution.
- In addition to these measures, we need to know another measure to get the complete idea about the shape of the distribution which can be studied with the help of Kurtosis.
- Kurtosis gives a measure of flatness of distribution.

The degree of kurtosis of a distribution is measured relative to that of a normal curve.

- The curves with greater peakedness than the normal curve are called **"Leptokurtic".**
- The curves which are flatter than the normal curve are called **"Platykurtic"**.
- The normal curve is called "**Mesokurtic**."
- The Fig.1 describes the three different curves mentioned above:

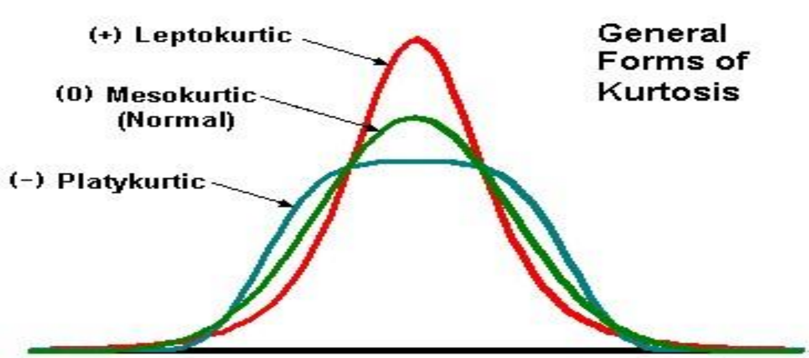


Fig. 1: Platykurtic Curve, Mesokurtic Curve and Leptokurtic Curve

Measures of Kurtosis

Karl Pearson's Measures of Kurtosis

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

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$
$$\gamma_2 = \beta_2 - 3$$

where, μ_2 = Second order central moment of distribution
μ_4 = Fourth order central moment of distribution

Description:
1. If β_2 = 3 or γ_2 = 0, then curve is said to be mesokurtic;
2. If β_2 < 3 or γ_2 < 0, then curve is said to be platykurtic;
3. If β_2 > 3 or γ_2 > 0, then curve is said to be leptokurtic;

Example: First four moments about mean of a distribution are 0, 2.5, 0.7 and 18.75. Find coefficient of skewness and kurtosis.

Solution: We have μ_1 = 0, μ_2 = 2.5, μ_3 = 0.7 and μ_4 = 18.75

Therefore, Skewness, $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(0.7)^2}{(2.5)^3} = \frac{0.49}{15.625} = 0.031$

$$Kurtosis, \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{2.5^2} = \frac{18.75}{6.25} = 3$$

As β_2 is equal to 3, so the curve is mesokurtic.