
UNIT 4 SKEWNESS AND KURTOSIS

Structure

- 4.1 Introduction
 - Objectives
- 4.2 Concept of Skewness
- 4.3 Various Measures of Skewness
- 4.4 Concept of Kurtosis
 - Measures of Kurtosis
- 4.5 Summary
- 4.6 Solutions/Answers

4.1 INTRODUCTION

In Units 1 and 2, we have talked about average and dispersion. They give the location and scale of the distribution. In addition to measures of central tendency and dispersion, we also need to have an idea about the shape of the distribution. Measure of skewness gives the direction and the magnitude of the lack of symmetry whereas the kurtosis gives the idea of flatness.

Lack of symmetry is called skewness for a frequency distribution. If the distribution is not symmetric, the frequencies will not be uniformly distributed about the centre of the distribution. Here, we shall study various measures of skewness and kurtosis.

In this unit, the concepts of skewness are described in Section 4.2 whereas the various measures of skewness are given with examples in Section 4.3. In Section 4.4, the concepts and the measures of kurtosis are described.

Objectives

On studying this unit, you would be able to

- describe the concepts of skewness;
- explain the different measures of skewness;
- describe the concepts of kurtosis;
- explain the different measures of kurtosis; and
- explain how skewness and kurtosis describe the shape of a distribution.

4.2 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-axis, 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 $\text{mean} > \text{median} > \text{mode}$ while if the left tail is longer, we get a negatively skewed distribution for which $\text{mean} < \text{median} < \text{mode}$.

The example of the Symmetrical curve, Positive skewed curve and Negative skewed curve are given as follows:

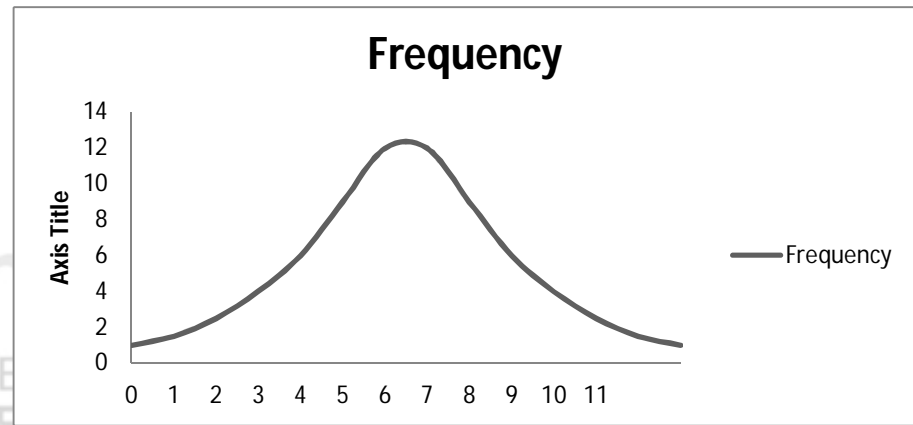


Fig. 4.1: Symmetrical Curve

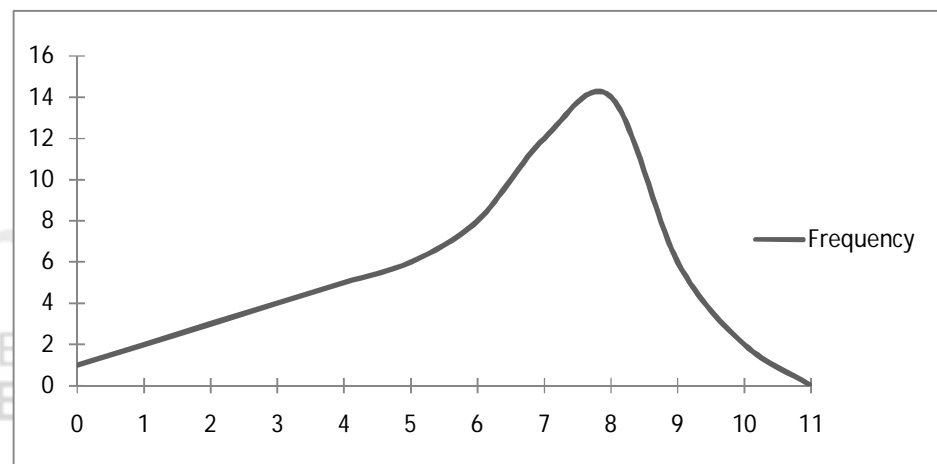


Fig. 4.2: Negative Skewed Curve

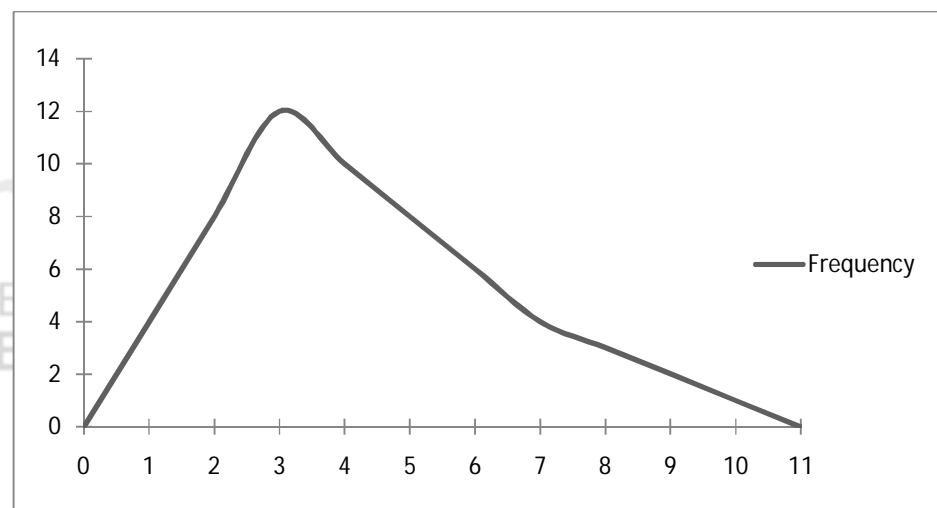


Fig. 4.3: Positive Skewed Curve

4.2.1 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 medical and life science field measures of skewness have greater practical applications than the variance.

4.3 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.

Measures of skewness can be both absolute as well as relative. Since in a symmetrical distribution mean, median and mode are identical more the mean moves away from the mode, the larger the asymmetry or skewness. An absolute measure of skewness can not be used for purposes of comparison because of the same amount of skewness has different meanings in distribution with small variation and in distribution with large variation.

4.3.1 Absolute Measures of Skewness

Following are the absolute measures of skewness:

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

For comparing to 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.

4.3.2 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:

1. 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:

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

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 γ_1 which is the square root of β_1 i. e.

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

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.

2. 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. The value of the Karl Pearson's coefficient of skewness usually lies between ± 1 for moderately skewed distribution. If mode is not well defined, we use the formula

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

By using the relationship

$$\text{Mode} = (3 \text{ Median} - 2 \text{ Mean})$$

Here, $-3 \leq S_k \leq 3$. In practice it is rarely obtained.

3. Bowleys's Coefficient of Skewness

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

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

The value of S_k 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.

4. Kelly's Coefficient of Skewness

The coefficient of skewness proposed by Kelly is based on percentiles and deciles. The formula for calculating the coefficient of skewness is given by

Based on Percentiles

$$S_k = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{(P_{90} - P_{10})} \\ = \frac{(P_{90} - 2P_{50} + P_{10})}{(P_{90} - P_{10})}$$

where, P_{90} , P_{50} and P_{10} are 90th, 50th and 10th Percentiles.

Based on Deciles

$$S_k = \frac{(D_9 - 2D_5 + D_1)}{D_9 - D_1}$$

where, D_9 , D_5 and D_1 are 9th, 5th and 1st Decile.

Example1: 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

$$S_k = 0.64, \sigma = 13 \text{ and Mean} = 59.2$$

Therefore by using formulae

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

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

$$\text{Mode} = 59.20 - 8.32 = 50.88$$

$$\text{Mode} = 3 \text{ Median} - 2 \text{ Mean}$$

$$50.88 = 3 \text{ Median} - 2 (59.2)$$

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

E1) Karl Pearson's coefficient of skewness is 1.28, its mean is 164 and mode 100, find the standard deviation.

E2) For a frequency distribution the Bowley's coefficient of skewness is 1.2. If the sum of the 1st and 3rd quarterlies is 200 and median is 76, find the value of third quartile.

E3) The following are the marks of 150 students in an examination. Calculate Karl Pearson's coefficient of skewness.

Marks	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	10	40	20	0	10	40	16	14

Remarks about Skewness

1. If the value of mean, median and mode are same in any distribution, then the skewness does not exist in that distribution. Larger the difference in these values, larger the skewness;
2. If sum of the frequencies are equal on the both sides of mode then skewness does not exist;
3. If the distance of first quartile and third quartile are 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 a data become a normal curve and when it is folded at middle and one part overlap fully on the other one then there is no asymmetry.

4.4 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. Prof. Karl Pearson has called it the "Convexity of a Curve". 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 more flat than the normal curve are called "**Platykurtic**". The normal curve is called "**Mesokurtic**." The Fig.4 describes the three different curves mentioned above:

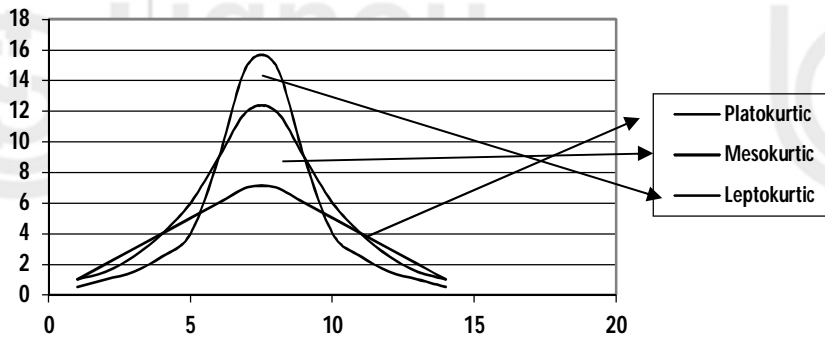


Fig.4.4: Platykurtic Curve, Mesokurtic Curve and Leptokurtic Curve

4.4.1 Measures of Kurtosis

1. 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:

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

or $\gamma_2 = \beta_2 - 3$

where, μ_2 = Second order central moment of distribution

μ_4 = Fourth order central moment of distribution

Description:

1. If $\beta_2 = 3$ or $\gamma_2 = 0$, then curve is said to be mesokurtic;
2. If $\beta_2 < 3$ or $\gamma_2 < 0$, then curve is said to be platykurtic;
3. If $\beta_2 > 3$ or $\gamma_2 > 0$, then curve is said to be leptokurtic;

2. Kelly's Measure of Kurtosis

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.

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

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

Example 2: 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 $\mu_1 = 0$, $\mu_2 = 2.5$, $\mu_3 = 0.7$ and $\mu_4 = 18.75$

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

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

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

E4) The first four raw moments of a distribution are 2, 136, 320, and 40,000. Find out coefficients of skewness and kurtosis.

4.5 SUMMARY

In this unit, we have discussed:

1. What is skewness;
2. The significance of skewness;
3. The types of skewness exists in different kind of distributions;
4. Different kinds of measures of skewness;
5. How to calculate the coefficients of skewness;
6. What is kurtosis;
7. The significance of kurtosis;
8. Different measures of kurtosis; and
9. How to calculate the coefficient of kurtosis.

4.6 SOLUTIONS/ANSWERS

E1) Using the formulae, we have

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

$$1.28 = \frac{164 - 100}{\sigma}$$

$$\sigma = \frac{64}{1.28} = 50$$

E2) We have given $S_k = 1.2$

$$Q_1 + Q_3 = 200$$

$$Q_2 = 76$$

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

$$1.2 = \frac{(200 - 2 \times 76)}{(Q_3 - Q_1)}$$

$$Q_3 - Q_1 = \frac{48}{1.2} = 40$$

$$Q_3 - Q_1 = 40$$

... (1)

and it is given $Q_1 + Q_3 = 200$

$$Q_1 = 200 - Q_3$$

Therefore, from equation (1)

$$Q_3 - (200 - Q_3) = 40$$

$$2Q_3 = 240$$

$$Q_3 = 120$$

- E3)** Let us calculate the mean and median from the given distribution because mode is not well defined.

Class	f	x	CF	$d' = \frac{x-35}{10}$	fd'	fd'^2
0-10	10	5	10	-3	-30	90
10-20	40	15	50	-2	-80	160
20-30	20	25	70	-1	-20	20
30-40	0	35	70	0	0	0
40-50	10	45	80	+1	10	10
50-60	40	55	120	+2	80	160
60-70	16	65	136	+3	48	144
70-80	14	75	150	+4	56	244
					$\sum fd' = 64$	$\sum fd'^2 = 828$

$$\begin{aligned} \text{Median} &= L + \frac{\left(\frac{N}{2} - C\right)}{f} \times h \\ &= 40 + \frac{75 - 70}{10} \times 10 = 45 \end{aligned}$$

$$\begin{aligned} \text{Mean } (\bar{x}) &= A + \frac{\sum fd'}{N} \times h \\ &= 35 + \frac{64}{150} \times 10 = 39.27 \end{aligned}$$

$$\begin{aligned} \text{Standard Deviation } (\sigma) &= h \times \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \\ &= 10 \times \sqrt{\frac{828}{150} - \left(\frac{64}{150}\right)^2} \end{aligned}$$

$$= 10 \times \sqrt{5.33} = 23.1$$

Therefore, coefficient of skewness:

$$S_k = \frac{3(\text{Mean} - \text{Median})}{\sigma} \\ = \frac{3(39.27 - 45)}{23.1} = -0.744$$

E4) Given that $\mu'_1 = 2$, $\mu'_2 = 136$, $\mu'_3 = 320$ and $\mu'_4 = 40,000$

First of all we have to calculate the first four central moments

$$\mu_1 = \mu'_1 - \mu'_1 = 0$$

$$\mu_2 = \mu'_2 - (\mu'_1)^2 = 136 - 2^2 = 132$$

$$\mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1^3 \\ = 320 - 3 \times 136 \times 2 + 2(2)^3 \\ = 320 - 792 + 16 = -456$$

$$\mu_4 = \mu'_4 - 4\mu'_1\mu'_3 + 6\mu_2^2\mu'_1 - 3\mu_1^4 \\ = 40,000 - 4 \times 2 \times 320 + 6 \times 2^2 \times 136 - 3 \times 2^4 \\ = 40,000 - 2560 + 3,264 - 48 = 40656$$

$$\text{Skewness, } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-456)^2}{(132)^3} = 0.0904$$

$$\text{Kurtosis, } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{40656}{(132)^2} = 2.333$$