7 Skewness and Kurtosis

Introduction

In the last two units we discussed the most important characteristics (Central Tendency and Dispersion) of a statistical data. Knowledge of averages and measures of dispersion is of great importance in assessing the properties of a group of data. An average represents the central tendency of the distribution and the dispersion measures the degree of variation around the central value.

These measures of central tendency and dispersion do not reveal whether the dispersal of values on either side of an average is symmetrical or not. If the spread of the frequencies is the same on both sides of the centre point of the curve then that curve is called symmetric curve.

Mirror image is symmetric

In a study of a frequency distribution it would be of great help to know whether it would give a symmetric curve and if not, what extend it would deviate from symmetry. The word skewness indicates the absence of symmetry in a data set. We study skewness to have an idea about the shape of the frequency curve which we can draw with the help of given data.

Two or more distributions may differ in terms of flatness or peakedness of their frequency curve. The method of measuring this characteristic of a data set is termed as Kurtosis. Thus the word Kurtosis relates to the degree of flatness or peakedness of a data.

7.1 Skewness

Consider the following frequency distributions which give the scores obtained by the students who were studying in Commerce, Humanities and Science groups in a higher secondary school.

1.	Scores	:	0-10	10-20	20-30	30-40	40-50	50-60
	No. of students	:	5	8	15	15	8	5
2.	Scores	:	0-10	10-20	20-30	30-40	40-50	50-60
	No. of students	:	4	7	16	11	7	5
3.	Scores	:	0-10	10-20	20-30	30-40	40-50	50-60
	No. of students	:	5	7	11	16	7	4

Mean and variance of the above distributions are same but they differ widely in their overall appearance as we can seen from the following diagrams

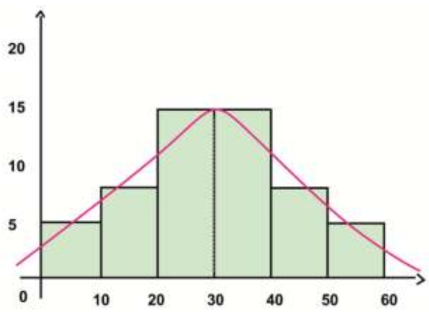


Figure (1)

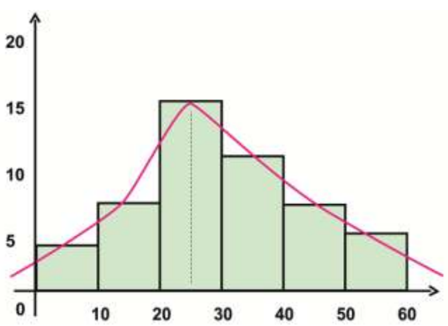


Figure (2)

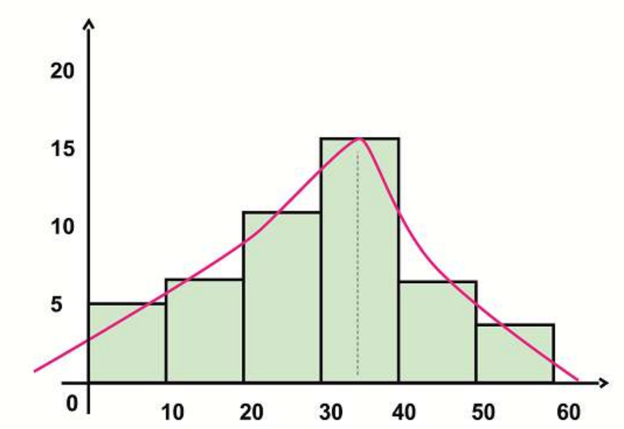


Figure (3)

In figure(1) the right and left of mode(highest ordinate) are perfect mirror images of one another. They are called as symmetric distributions.
In figure (2) you can see more items on the right side of mode and have a longer tail to the right side of mode. Similarly in figure (3), more items on the left of mode and have longer tail to the left of mode. When frequency curves are drawn for different frequency distributions, there is an apparent common characteristic, which is striking to the eye, called symmetry or lack of symmetry. The lack of symmetry of a distribution is known as skewness. Here it is clear that figures (2) and (3) are not symmetric or they are skewed. Thus, there are 2 types of skewness 1) Positive skewness 2) Negative skewness.

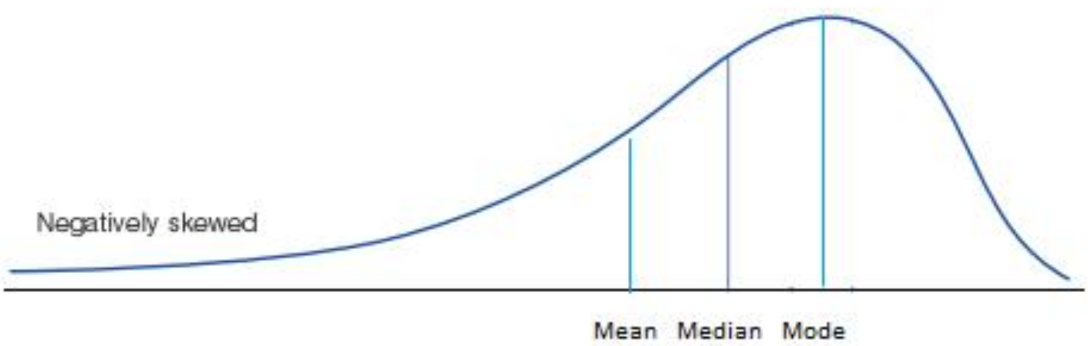
Positive Skewness

The frequency curve is said to be positively skewed if more items are found to the right side of the mode. In this case the frequency curve will have a longer tail to the right. Also Mode, Median and Mean are in the ascending order of their magnitude.

,]

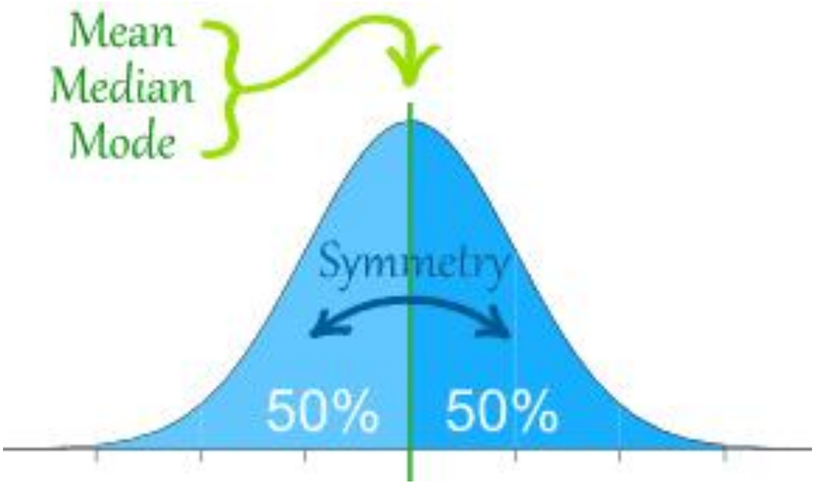
Negative Skewness

The frequency curve is said to be negatively skewed if more items are found to the left side of the mode. In this case the curve will have a longer tail to the left and Mode, Median and Mean are in the descending order of their magnitude.



For a symmetric frequency curve

- Mean = Median = Mode
- The highest ordinate (mode) divides the total area under the curve into two equal parts
- Quartiles are equidistant from median. Ie. $Q_3 -$ Median = Median $- Q_1$.



For a positively skewed data, $Q_3 - \text{Median} > \text{Median} - Q_1$
For a negatively skewed data, $Q_3 - \text{Median} < \text{Median} - Q_1$

In a symmetric distribution the mean, median, mode coincides. The symmetry of such a curve can be changed by adding some items to one side or other side of the mode. If there are more items to the right of the highest ordinate of the frequency curve, the curve is said to be positively skewed. Addition of more items on right side pulls AM away from mode towards the right. As there are some more items on the right side of the mode, the median is also slightly pulled from mode towards right. Similar thing will happen if more items are added to the left. Thus skewness has got the effect of pulling the arithmetic mean and the median away from the mode, sometimes towards the right and sometimes towards the left.

Activity

Calculate Mean, Median, Mode and Quartiles of three distributions given below and Compare the results.

1.	Scores	:	0-10	10-20	20-30	30-40	40-50
	No. of students	:	5	12	16	12	5
2.	Scores	:	0-10	10-20	20-30	30-40	40-50
	No. of students	:	5	20	12	8	5
3.	Scores	:	0-10	10-20	20-30	30-40	40-50
	No. of students	:	5	8	12	20	5

Also draw frequency curves of above three distributions and compare them.

Activity

Collect data regarding

(a) Marks in various subjects in the class.
(b) Heights of students in the class.
(c) Data on income of parents
(d) Consumption of electricity

Compute Mean, Median and Mode. Comment about the skewness of the distributions.

7.2 Measures of Skewness

Measures of skewness indicate to what extend and in what direction the distribution of variable departs from symmetry of a frequency curve. It gives the information about the shape of the distribution and the degree of variation on either side of the central value.

There are 3 important measures of Skewness

1. Karl Pearsons coefficient of Skewness
2. Bowleys coefficient of Skewness
3. Coefficient of Skewness based on Moments.

Karl Pearsons Coefficient of Skewness

For a symmetrical data we have mean = mode. If there is skewness, the mean is separated from the mode. If mean - mode is positive then there is positive Skweness and if mean - mode is negative then there is negative skewness. This measure is a natural way of measuring skewness.

The above measure is inadequate for comparison since they may be same for two distributions with different dispersions. So we use coefficient of skewness which is numerical figures independent of units of measurement.

Karl Pearson derived the coefficient of skewness denoted by S_k and is defined as

Karl Pearsons Coefficient of Skewness,

$$\begin{aligned} S_k &= \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}} \\ &= \frac{\overline{X} - \text{Mode}}{\sigma} \end{aligned}$$

For positive skewness, $S_k > 0$ (since Mean> Mode).

For negative skewness, $S_k < 0$ (since Mean < Mode).

For symmetry $S_k = 0$ (since Mean= Mode).

Illustration 7.1

For a distribution Mean=30, Mode=26.8 and variance=64. Find the coefficient of skewness. Interpret the result

Solution. *Given* $\overline{X} = 30$, *Mode* $= 26.8, \sigma^2 = 64, \sigma = 8.$

Karl Pearsons Coefficient of Skewness,

$$\begin{aligned} S_K &= \frac{\overline{x} - \textit{Mode}}{\sigma} \\ &= \frac{30 - 26.8}{8} \\ &= 0.40 \end{aligned}$$

Since $S_K > 0$, *the distribution is positively Skewed.*

Illustration 7.2

For a group of 20 items, $\sum x = 1452, \sum x^2 = 144280$ and mode = 63.7. Obtain Karl Pearsons coefficient of skewness.

Solution. Given Mode=63.7,

$$\begin{aligned}\overline{x} &= \frac{\sum x}{n} \\ &= \frac{1452}{20} \\ &= 72.6 \\ \sigma &= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \\ &= \sqrt{\frac{144280}{20} - \left(\frac{1452}{20}\right)^2} \\ &= 44.08\end{aligned}$$

Karl Pearsons coefficient of Skewness,

$$\begin{aligned}S_K &= \frac{\overline{X} - \text{Mode}}{\sigma} \\ &= \frac{72.6 - 63.7}{44.08} \\ &= 0.2019\end{aligned}$$

Since $S_K > 0$, the distribution is positively skewed

Illustration 7.3

The number of accidents reported at city hospital in a week as follows 40,62,40,25, 40,34 and 60. Calculate Karl Pearsons coefficient of skewness.

Solution. Given

$$\begin{aligned}
\overline{x} &= \frac{\sum x}{n} \\
&= \frac{301}{7} \\
&= 43 \\
\sigma &= \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} \\
&= \sqrt{\frac{14025}{7} - \left(\frac{301}{7}\right)^2} \\
&= 12.43 \\
\text{Mode} &= 40 (\text{Most frequently occurred item})
\end{aligned}$$

Karl Pearsons coefficient of skewness,

$$\begin{aligned}
S_K &= \frac{\overline{X} - \text{Mode}}{\sigma} \\
&= \frac{43 - 40}{12.43} \\
&= 0.24
\end{aligned}$$

Since $S_K > 0$, the distribution is positively skewed.

Illustration 7.4

Find the coefficient of Skewness by Karl Pearsons method for the following data and comment upon the nature of skewness.

Value	:	*6*	*12*	*18*	*24*	*30*	*36*	*42*
Frequency	:	*4*	*7*	*9*	*18*	*15*	*10*	*3*

Solution.

$$\bar{x} = \frac{\sum fx}{n} = \frac{1638}{66} = 24.82$$

$$\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2} = \sqrt{\frac{46188}{66} - \left(\frac{1638}{66}\right)^2} = 9.16$$

$$\text{Mode} = \text{observation having highest frequency} = 24$$

Karl Pearsons coefficient of Skewness,

$$S_K = \frac{\overline{X} - \text{Mode}}{\sigma} = \frac{24.82 - 24}{9.16} = 0.089$$

Since $S_K > 0$, the given distribution is positively skewed.

Illustration 7.5

The monthly income distribution of 100 persons living in a village is attached

Income(In '000s)	0-10	10-20	20-30	30-40	40-50	50-60
No. of Persons	12	18	27	20	17	6

Determine (a) Mode (b) Standard Deviation (c) Coefficient of Skewness of this distribution.

Solution. The highest frequency = 27, hence the model class is 20-30

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_2 - f_0}\right)c$$

Where $l = 20$, $c = 10$, $f_0 = 18$, $f_1 = 27$ and $f_2 = 20$

$$\begin{aligned} Mode &= 20 + \frac{90}{60} \\ &= 25.625 \\ \text{Mean } &= \frac{\sum fx}{N} \\ &= \frac{2800}{100} \\ &= 28 \\ \text{Standard deviation,}\sigma &= \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2} \\ &= \sqrt{\frac{98300}{100} - (28)^2} \\ &= \sqrt{199} = 14.11 \end{aligned}$$

Karl Pearsons coefficient of Skewness

$$\begin{aligned} S_K &= \frac{\overline{X} - \text{Mode}}{\sigma} \\ &= \frac{28 - 25.625}{14.11} \\ &= 0.16832 \end{aligned}$$

Since $S_K > 0$, the given distribution is positively skewed.

Know your progress

1. *A sample distribution of 200 employees according to their gross monthly salary drawn by them shows an average salary Rs. 3590, mode Rs. 3660 and Variance Rs. 625. Obtain Karl Pearsons coefficient of skewness and interpret it.*

2. *The sum of 20 observations is 300, the sum of their squares is 5000 and Mode is 15. Find the coefficient of skewness and coefficient of variations.*

Bowleys Coefficient of Skewness

For symmetrical data we have Median $-Q_1 = Q_3 -$ Median. It will not be equal for skewed data , hence the difference ($Q_3 -$ Median) – (Median – Q_1) can be used to measure skewness.

Sir Arthur Bowley derived another measure of skewness based on Quartiles which is known as Bowleys coefficient of skewness denoted as S_B and is defined as

Bowleys coefficient of skewness, $S_B = \dfrac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1}$

The value of S_B lies between -1 to +1.
If $S_B > 0$ the distribution is positively skewed.
If $S_B < 0$ the distribution is negatively skewed.
If $S_B = 0$ the distribution is symmetric.

Note:

1. Bowleys coefficient of skewness is very useful especially in the case of open-end classes
2. Results of S_B and S_k are not to be compared with each other.

Illustration 7.6

For a certain distribution the upper and lower quartiles are 56 and 44 respectively. If the median for the same data is 55 then identify the nature of skewness.

Solution. Given $Q_1 = 44$, $Q_3 = 56$, median $= 55$
Bowleys coefficient of skewness,

$$S_B = \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1}$$
$$= \frac{56 + 44 - 2 \times 55}{56 - 44}$$
$$= -0.83$$

Since $S_B < 0$, the distribution is negatively skewed.

Illustration 7.7

If 25% of the total observations lie above 70 and 50% of the total observations are less than 38 and 75% of the total observations are greater than 30, then find coefficient of skewness and interpret the result.

Solution. Since 25% observations are above 70, we have $Q_3 = 70.$,
Since 50% observations are below 38, we have $Q_2 = 38.$
Since 75% observations are greater than 30, we have $Q_1 = 30.$

Bowleys coefficient of skewness,

$$\begin{aligned} S_B &= \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1} \\ &= \frac{70 + 30 - 2 \times 38}{70 - 30} \\ &= 0.6 \end{aligned}$$

Since $S_B > 0$, the distribution is positively skewed.

Illustration 7.8

The Bowleys coefficient of skewness is 0.8, the sum of two quartiles is 80 and median is 30. Find the values of upper and lower quartiles.

Solution. Given $S_B = 0.8$, $Q_1 + Q_3 = 80$, Median $= 30$
We have, Bowleys coefficient of skewness,

$$\begin{aligned} S_B &= \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1} \\ 0.8 &= \frac{80 - 2 \times 30}{(80 - Q_1) - Q_1} \\ 0.8 &= \frac{20}{80 - 2Q_1} \\ \text{i.e.,} Q_1 &= 27.5, \; Q3 = 52.5 \end{aligned}$$

Illustration 7.9

Following marks were obtained in Statistics by 15 students. 15, 20, 20, 21, 22, 22, 24, 28, 28, 29, 30, 32, 25, 33 and 35. Calculate quartile coefficient of skewness.

Solution. Arranging the data in ascending order we have

$$15, 20, 20, 21, 22, 22, 24, 25, 28, 28, 29, 30, 32, 33, 35$$

$$Q_1 = \left[\frac{n+1}{4}\right]^{\text{th}} \text{item} = 4^{\text{th}} \text{ item} = 21$$

$$\text{Median} = \left[\frac{n+1}{2}\right]^{\text{th}} \text{item} = 8^{\text{th}} \text{ item} = 25$$

$$Q_3 = \left[\frac{3(n+1)}{4}\right]^{\text{th}} \text{item} = 12^{\text{th}} \text{value} = 30$$

Bowleys coefficient of skewness,

$$S_B = \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1}$$

$$= \frac{30 + 21 - 2 \times 25}{30 - 21}$$

$$= 0.11$$

Since $S_B > 0$, the distribution is positively skewed.

Illustration 7.10

The following data shows daily wages of 124 workers of a factory.

Wages(Rs.)	*200*	*250*	*300*	*350*	*400*	*450*	*500*	*550*
No. of workers	*10*	*15*	*18*	*30*	*26*	*15*	*8*	*2*

Find (1) Quartiles (2) Coefficient of Skewness

Solution.

X	C.f
200	10
250	25
300	43
350	73
400	99
450	114
500	122
550	124

$$Q_1 = \left[\frac{n+1}{4}\right]^{\text{th}} \text{item} = 31.25^{\text{th}} \text{ item}$$

$$= 300$$

$$\text{Median} = \left[\frac{n+1}{2}\right]^{\text{th}} \text{item} = 62.5^{\text{th}} \text{ item}$$

$$= 350$$

$$Q_3 = \left[\frac{3(n+1)}{4}\right]^{\text{th}} \text{item} = 93.75^{\text{th}} \text{ item}$$

$$= 400$$

Bowleys coefficient of skewness,

$$S_B = \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1}$$
$$= \frac{400 + 300 - 2 \times 350}{400 - 300}$$
$$= 0$$

Since $S_B = 0$, the distribution is symmetric.

Illustration 7.11

Following data show the ages of family members in a colony consisting of 20 families

Ages(Years)	*0-20*	*20-40*	*40-60*	*60-80*	*80-100*
No. of members	*4*	*10*	*15*	*20*	*11*

Find out the nature of skewness for the above data using quartile coefficient.

Solution.

Age	No. of members	c.f
0 -20	4	4
20-40	10	14
40-60	15	29
60-80	20	49
80-100	11	60

Find Quartiles as in chapter 5.

$$Q_1 = 40 + \left(\frac{15 - 14}{15}\right)20 = 41.33$$
$$\text{Median} = 60 + \left(\frac{30 - 29}{20}\right)20 = 61$$
$$Q_3 = 60 + \left(\frac{45 - 29}{15}\right)20 = 76$$

Bowleys coefficient of skewness,

$$S_B = \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1}$$
$$= \frac{76 + 41.33 - 2 \times 61}{76 - 41.33}$$
$$= -0.1347$$

Since $S_B < 0$, the distribution is negatively skewed.

Illustration 7.12

Find Quartile coefficient of Skewness of the two groups given below and point out which distribution is more skewed.

Scores	*Group A*	*Group B*
55-58	*12*	*20*
58-61	*17*	*22*
61-64	*23*	*25*
64-67	*18*	*13*
67-70	*11*	*7*

Solution.

For Group A

$$\begin{aligned} Q_1 &= 58 + \left(\frac{20.25 - 12}{17}\right)3 = 59.46 \\ \text{Median} &= 61 + \left(\frac{40.5 - 29}{23}\right)3 = 62.5 \\ Q_3 &= 64 + \left(\frac{60.75 - 52}{18}\right)3 = 65.46 \end{aligned}$$

Quartile coefficient of skewness for group A,

$$\begin{aligned} S_B &= \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1} \\ &= \frac{65.46 + 59.46 - 2 \times 62.5}{65.46 - 59.46} \\ &= -0.013 \end{aligned}$$

Similarly for Group B, $Q_1 = 58.24$, Median $= 61.18$ and $Q_3 = 63.79$
Quartile coefficient of skewness for group B,

$$\begin{aligned} S_B &= \frac{Q_3 + Q_1 - 2\text{Median}}{Q_3 - Q_1} \\ &= \frac{63.79 + 58.24 - 2 \times 61.18}{63.79 - 58.24} \\ &= -0.0595 \end{aligned}$$

S_B for group B is greater than that of group A in magnitude. So B is more skewed than A.

7.3 Moments

Moments are statistical constants used to describe the various characteristics of a frequency distribution like central tendency, variation, skewness and kurtosis. It is a convenient and unifying method for summarising descriptive statistical measures.

Moments are calculated using the arithmetic mean. The arithmetic mean of the various powers of deviations of observations in any distribution is called the moments of the distribution. If the deviations are taken from arithmetic mean then the moments are termed as central moments and it is denoted by μ (pronounced as 'mu'), a Greek letter. The first four central moments are defined below. Let $x_1, x_2, x_3, \ldots, x_n$ be n observations then

First central moment

$$\mu_1 = \frac{\sum(x-\overline{x})}{n}$$

$$= 0 \text{(Since sum of deviations of items from the mean is always zero}$$

Second central moment

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

$$= \text{variance}$$

Third central moment

$$\mu_3 = \frac{\sum(x-\overline{x})^3}{n} \text{ and}$$

Fourth central moment

$$\mu_4 = \frac{\sum(x-\overline{x})^4}{n}$$

Coefficient of Skewness based on Moments.

Coefficient of skewness Based on moments, β_1 (pronounced as beeta one) is defined as

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

If $\beta_1 = 0$ the curve is symmetric. The greater the values of β_1 the more skewed the distribution. β_1 does not gives the direction of skewness, since μ_3^2 and μ_2^3 are always positive. Hence Karl Pearson defined γ_1(pronounced as Gamma one) as

$$\gamma_1 = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{\mu_3}{\sigma^3} = \sqrt{\beta_1}$$

if $\mu_3 > 0$ then $\gamma_1 > 0$ the distribution is positively skewed. If $\mu_3 < 0$ then $\gamma_1 < 0$ the distribution is negatively skewed. If $\mu_3 = 0$ then $\gamma_1 = 0$ the distribution is symmetric. Thus μ_3 determines the nature of skewness.

Illustration 7.13

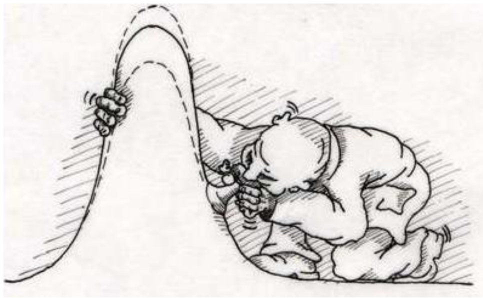
The first four central moments of a distribution are 0, 14.75, 39.75 and 142.31. Find coefficient of Skewness and state the nature of Skewness.

Solution. Given$\mu_2 = 14.75$, $\mu_3 = 39.75$

$$\text{Coefficient of Skewness } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{39.75^2}{14.75^3} = 0.4924$$

Since $\mu_3 > 0$, the distribution is positively skewed.

7.4 Kurtosis



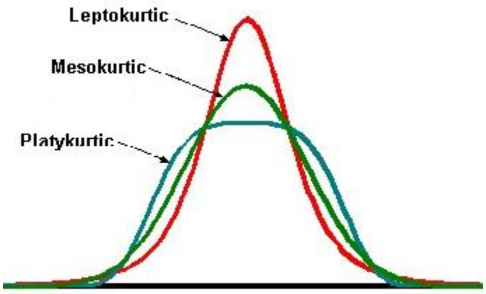
Two or more distributions may have identical average, variation and skewness, but they show different degrees of concentration of values of observation around mode and hence they may show different degrees of peakedness of the distributions. Kurtosis is the measure of Peakedness or flatness of the frequency distribution.

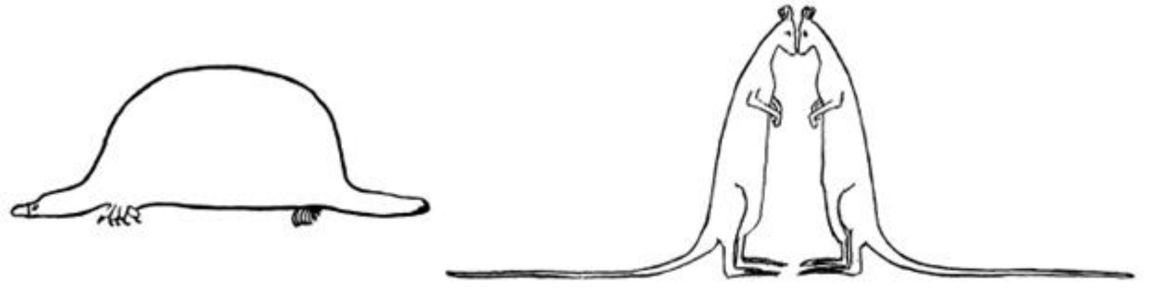
Types of Kurtosis

Look at the following frequency curves

Figure(1) Figure(2) Figure(3)

All of them have same centre of location, dispersion and are symmetrical, but they differ in peakedness. In figure 1 the curve is more peaked than others and is called as Lepto Kurtic. In figure 2 the curve is less peaked than others or it is more flat topped, is called as Platty Kurtic. In figure 3 curve is neither more peaked nor more flat topped or it is moderately peaked, is called Meso Kurtic. Meso Kurtic is also known as Natural Curve or Normal Curve.





The word "Kurtosis" is derived from the Greek word meaning "Humped" or "Bulginess". Famous British Statistician William S Gosset (Student) has very humorously pointed out the nature of kurtosis in his research paper "Errors of Routine Analysis" that platy kurtic curves, like the platypus, are squat with short tails; lepto kurtic curves are high with long tails like the Kangaroos noted for lepping. Gossets little sketch is reproduced as shown in the figure.

7.5 Measures of Kurtosis

Kurtosis is measured by coefficient β_2(pronounced as beeta two) is defined by

$$\begin{aligned}\beta_2 &= \frac{\mu_4}{\mu_2^2}\\ &= \frac{\mu_4}{\sigma^4}\end{aligned}$$

Karl Pearson defined one more coefficient of kurtosis as $\gamma_2 = \beta_2 - 3$
If $\beta_2 = 3$ (ie,$\gamma_2 = 0$) the curve is meso kurtic.
If $\beta_2 > 3$ (ie, $\gamma_2 > 0$) the curve is lepto kurtic.
If $\beta_2 < 3$ (ie, $\gamma_2 < 0$) the curve is platy kurtic.

Illustration 7.14

The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Test the kurtosis of the distribution

Solution. Given $\mu_1 = 0$, $\mu_2 = 2.5$, $\mu_3 = 0.7$, $\mu_4 = 18.75$

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

Since $\beta_2 = 3$ the distribution is meso kurtic.

Illustration 7.15

For the following data calculate coefficient of skewness and coefficient of kurtosis and comment on it 2, 3, 7, 8, 10

Solution.

X	$(x-6)$	$(x-6)^2$	$(x-6)^3$	$(x-6)^4$
2	-4	16	-64	256
3	-3	9	-27	81
7	1	2	1	1
8	4	4	8	16
10	4	16	64	256
30	0	46	-18	610

$$\bar{x} = \frac{30}{5} = 6$$

$$\mu_1 = \frac{\sum(x-\bar{x})}{n} = 0(\text{Always})$$

$$\mu_2 = \frac{\sum(x-\bar{x})^2}{n} = \frac{46}{5} = 9.2$$

$$\mu_3 = \frac{\sum(x-\bar{x})^3}{n} = \frac{-18}{5} = -3.6$$

$$\mu_4 = \frac{\sum(x-\bar{x})^4}{n} = \frac{610}{5} = 122$$

$$\text{Coefficient of skewness, } \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-3.6)^2}{9.2^3} = 0.0166$$

Since $\mu_3 < 0$, the distribution is negatively skewed

$$\text{Coefficient of kurtosis, } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{122}{9.2^2} = 1.44$$

Since $\beta_2 < 3$, the distribution is platy kurtic

Let us sum up

In this chapter the concept of skewness and kurtosis were introduced. Skewness means lack of symmetry where as kurtosis is the measure of peakedness. There are two types of skewness, positive skewness and negative skewness. If a frequency curve has longer tail towards right side of mode, then it is said to be positively skewed. If a frequency curve has longer tail towards left side of mode, it is said to be negatively skewed. If a curve is relatively narrow and peaked at the top, it is called leptokurtic. The curve which is more flat topped is called platy kurtic. The curve which is neither more peaked nor more flat topped is called meso kurtic. Various measures of skewness and kurtosis were also discussed here. Measures of skewness indicate to what extend and in what direction the distribution of variable differs from symmetry of a frequency curve. Measures of kurtosis denote the shape of top of a frequency curve. For a symmetric curve Karl Persons coefficient of skewness , Bowleys coefficient of skewness and moment coefficient of skewness are all equal to zero.

Learning outcomes

After transaction of this unit, the learner:-

- distinguishes symmetric and asymmetric distributions.
- recognises skewness of distributions.
- evaluates and interprets nature of skewness.
- explains kurtosis of distributions.
- evaluates and interprets types of kurtosis.

Evaluation Items

1. For a positively skewed frequency distribution, which of the following is true always ?

   (A) $Q_1 + Q_3 > 2Q_2$ (B) $Q_1 + Q_2 > 2Q_3$ (C) $Q_1 - Q_3 > Q_2$
   (D) $Q_3 - Q_1 > Q_2$

2. For negatively skewed data, which of the following is true?

   (A) Mean = Median = Mode (B) Median < Mean < Mode
   (C) Mean < Median < Mode (D) Mode < Mean < Median

3. For a negatively skewed distribution, more observations clustered

   (A) on the left tail (B) on the right tail (C) in the middle
   (D) Anywhere

4. The limits of Bowleys coefficient of skewness are

   (A) ± 1 (B) ± 2 (C) ± 3 (D) 0 to 1

5. Given $\mu_2 = 7$, $\mu_4 = 98$, then the curve is

   (A) meso kurtic (B) platy kurtic (C) positively skewed
   (D) lepto kurtic

6. For a symmetric distribution

   (A) $\beta_1 = 0$ (B) $\beta_1 < 0$ (C) $\beta_1 > 0$ (D) $\beta_1 \neq 0$

7. For a set of data, third central moment is -1.6, then the coefficient of skewness is

   (A) positively skewed (B) negatively skewed (C) symmetric
   (D) Can't detertmined

8. In an economic study, a sample of persons earning up to Rs. 30,000 per month were considered. It is found that 30% are earning less that 5000 per month, 95% are earning less than Rs. 15,000 per month and 98% less than 24,000 per month. Then the frequency curve for the data will be

   (A) Symmetric (B) Positively skewed (C) Negatively skewed
   (D) Nothing can be inferred

9. Which of the following is not correct statement

   (A) For a symmetric distribution Mean = Median = Mode
   (B) If Median $= 24$ and Mean $= 26$ then the skewness is positive
   (C) $\beta_1 = 0$ for a positively skewed data
   (D) If $\beta_2 = 3$, the distribution is meso kurtic

10. Old age distribution is an example for ________ Skewed distribution

11. If $\beta_2 > 3$ then the curve is ________

12. Skewness is ________ if $(Q_3 - Q_2) < (Q_2 - Q_1)$

13. If Karl Pearsons coefficient of Skewness is 0.40, standard deviation is 8 and mean is 30. Find the mode of the distribution.
    Ans. Mode $= 26.8$

14. For a group of 10 times $\sum x = 452$, $\sum x^2 = 24270$ and mode $= 43.7$. Find the Pearsonian coefficient of Skewness.
    Ans. Mean $= 45.2$,SD $= 19.6$ and $S_K = 0.08)$

15. The income distribution of two villages revealed the following information.

	Mean	Mode	Standard deviation
Village I	500	475	10
Village II	600	590	5

What is the nature of Skewness of the two distributions ? Which distribution is more skewed?

Ans. Positively Skewed, Village 1 is more skewed

16. Tests were conducted for 3 subjects namely Economics, English and Statistics for 37 students in a class. The marks obtained by them are tabulated as follows

Marks	No.of students		
	Economics	English	Statistics
12	2	2	1
14	5	12	3
16	7	8	5
18	9	6	6
20	7	5	8
22	5	3	12
24	2	1	2
Mean	18	16.6	18.96
Median	18	16	20
Mode	18	14	22

(A) Identify the Positively skewed, Negatively Skewed and Symmetric distributions

(B) Indicate the position of Mean, Median and Mode in three cases by drawing frequency curves of the distribution.

Ans. Eco-Symmetric, English - Positive, Statistics- Negative

17. For a frequency distribution the Mean=100, coefficient of skewness is 0.2 and Pearsons coefficient of variance is 35. Find mode of the distribution.

Ans. (SD =35, Mode=93)

18. In a distribution of wages of workers in a factory, the difference of upper and lower quartiles is 15, their sum is 35 and the Median is 20. Find the coefficient of Skewness.

Ans. $(S_B = 0.33)$

19. Suppose the distribution of scores of some students is symmetric. If the values of Q1and Q3 are 20 and 40 respectively, what is the Median mark? If the Median mark is 35, what would be the skewness of the distribution ?

Ans. Median $= 30$, $S_k = -0.5$

20. The following data represent the monthly salary (in thousands)of seven Assistant Professors in the department Statistics of Presidency college 26, 30, 32, 26, 29, 28, 60. Find (a) Mean (b) Mode (c) Coefficient of Skewness.

Ans. Mean $= 32.57$,Mode $= 26$, $SD = 11.2$,$S_k = 0.59$

21. The following figures represent the weights (in Kg) of 10 new born babies in a hospital on a particular day 2, 3, 3, 3, 4, 2, 2.5, 3.5, 3.7, 3

    (A) Compute Mean, Mode and Standard deviation

    (B) Are these data Skewed? Give reason

Ans. Mean $= 2.97$, Mode $= 3$, $SD = 0.63$, $S_k = -0.048$

22. The following table gives the height (in inches) of 100 students in a higher secondary school.

Class interval	No. of Students
60 and up to 62	5
62 and up to 64	18
64 and up to 66	42
66 and up to 68	20
68 and up to 70	8
70 and up to 72	7

Calculate (a) Mean (b)Mode (c) Coefficient of Skewness

Ans. Mean $= 65.58$, Mode $= 65.04$, $SD = 2.41$, $S_k = 0.23$

23. The president of company, which employs 50 persons, wants to study the pattern of absenteeism of all employees. The distribution of the number of days these employees were absent is given as follows:

No. of days absent	No. of Employees
0 to 2	15
3 to 5	20
6 to 8	8
9 to 11	5
12 to 14	2

Calculate (a) Mean (b) Standard deviation (c) Karl Pearsons Coefficient of Skewness. **Ans.** Mean = 4.54, Mode= 3.59, SD= 3,27, $S_K = 0.29$

24. Calculate Karl Pearsons coefficient of skewness for the following data

Life in Hrs.	:	80-160	160-240	240-320	320-400	400-480
No. of Tubes	:	24	90	45	12	30
Life in Hrs.	:	480-560	560-640	640-720		
No. of Tubes	:	120	39	30		

Ans. Mean=403.1, Mode=522.1,$SD = 174.2$, $S_K = -0.68$

25. Compute Karl Pearsons coefficient of skewness from the following data

Marks	:	Above 0	Above 10	Above 20	Above 30	Above 40
No. of Students	:	140	130	115	95	80
Marks	:	Above 50	Above 60	Above 70	Above 80	
No. of Students	:	70	30	14	0	

Ans. Mean=43.14,Mode=55.56,$SD = 20.96$, $S_k = -0.59$

26. Calculate coefficient of skewness using Quartiles

Mid value	:	15	20	25	30	35	40
Frequency	:	30	28	25	24	10	21

Ans. $Q_1 = 18.3$, $Q_2 = 24.7$, $Q_3 = 31.8$, $S_B = 0.052$

27. Calculate Bowleys coefficient of skewness for the following data

Life(in month) :	<87.5	<112.5	<137.5	<162.5
No. of bulbs :	35	75	123	223
Life(in month) :	<187.5	<212.5	<237.5	<262.5
No. of bulbs :	348	428	478	500

Ans. $Q_1 = 138$, $Q_2 = 167.9$, $Q_3 = 195.94$, $SB = -0.03$

28. Find Karl Pearsons coefficient of skewness for the two series and point out which one is more skewed

Age	No. of children	
(in Yrs)	School A	School B
6	3	1
8	9	10
9	15	9
10	8	7
11	5	3
Total	40	30

Ans. For A, Mean = 9, Mode = 9, $SD = 1.26$, $S_k = 0$; For B, Mean = 9, Mode = 8, $SD = 1.13$,$S_k = 0.88$. B is more Skewed

29. Calculate Bowleys coefficient of skewness for the marks of the students if

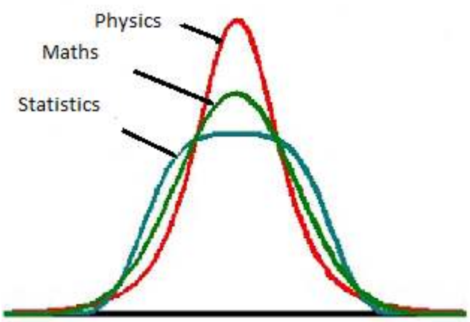
3 students get 3 marks each
5 students get 5 marks each
8 students get 7 marks each
6 students get 8 marks each
2 students get 10 marks each.

Ans. $Q_1 = 5$, $Q_2 = 7$, $Q_3 = 8$, $SB = -0.33$

30. If 2nd and 4th central moments of a frequency distribution are 2 and 16 respectively. What would be the nature of kurtosis of the distribution ?

Ans. $\beta_2 = 4$, lepto kurtic

31. The frequency curves of marks obtained in three subjects are given



Comment on the type of kurtosis.

Ans. Physics is lepto, Maths is meso and Statistics is platy kurtic

32. The first four central moments of a distribution are 0,9.56,-3.29 and 215.72. Compute the measures of skewness and kurtosis. State the nature of the curve.

Ans. $\beta_1 = 0.012$, skewness is negative,$\beta_2 = 2.36$ platy kurtic

33. The 1st four moments about mean of a distribution are 0, 16,-64 and 162. Calculate moment coefficients of skewness and kurtosis.

Ans. $\beta_1 = 1$, skewness is negative ,$\beta_2 = 0.633$ platy kurtic

34. For two distributions the second central moments are 9 and 16 and their third central moments are -8.1 and -12 respectively. Which of the two distributions is more skewed to the left ?

Ans. For I ,$\beta_1 = 0.9$, $\gamma_1 = -0.3$; For II $\beta_1 = 0.04$, $\gamma_1 = -0.2$; First one is More Skewed.

35. The first Four central moments of a distribution are 0, 9.2, -3.6 and 122. Calculate coefficient of kurtosis of the distribution.

Ans. $\beta_2 = 1.44$, platy kurtic

36. The values of β_1 and γ_1 for three distributions are given below

Identify the meso kurtic distribution.

Ans. Distribution II is meso kurtic

37. For a negatively skewed distribution, the values of β_1 and μ_2 are 9 and 4 respectively. Find the values of γ_1 and μ_3.

    Ans. $\gamma_1 = -3$, $\mu_3 = -24$ since skewness is negative

38. First two central moments of a meso kurtic distribution are 0 and 2.5 respectively. Find fourth central moment.

    Ans. $\mu_4 = 18.75$

39. Compute first four central moments for first four even numbers. Also calculate coefficients of skewness and kurtosis

    Ans. $\mu_2 = 5$, $\mu_3 = 0$, $\mu_4 = 41$, $\beta_1 = 0$, $\beta_2 = 1.64$

40. Determine the coefficient of kurtosis and comment on the nature of data 3, 6, 8, 10,18

    Ans. $\mu_2 = 25.6$, $\mu_3 = 97.2$, $\mu_4 = 1588$, $\beta_2 = 2.42$, platy kurtic

Answers:

1) a 2) c 3) a 4) a 5) b 6) a 7) b 8) c 9) c
10) positively skwed 11) lepto kurtic 12) negative