3. SKEWNESS

Let's Study

- Skewness
- Measures of skewness

Let's Learn

Two distributions having similar measures of central tendency and dispersion can still be different in terms of symmetry. This is necessarily important as it gives visual sense of data in numerical terms. Also these numbers help us indicate whether a distribution has higher frequencies (associated with the observations) accumulated in the left or towards the right end of the axis; this is a measure of symmetry called Skewness.

3.1 Skewness

Skewness indicates whether a distribution is symmetric or asymmetric and if asymmetric, it also indicates the extent of asymmetry with direction.

Zero Skewness: Say for instance, there is frequency distribution such that the **values of variables equidistant from the mean have equal frequencies**. In this case, the distribution is said to be symmetric (Zero Skewness) (fig. 3.3)

X (variable)	14	16	18	20	22	24	26
f (frequency)	2	5	10	14	10	5	2

Mean from the above table is 20

Note that Values of x equidistant from mean have equal frequencies. i.e. 18 and 22 are at a distance of 2 from 20 (mean) and have equal frequencies of 10 and 10 respectively. So on and so forth.

This is an indication of symmetric distribution.

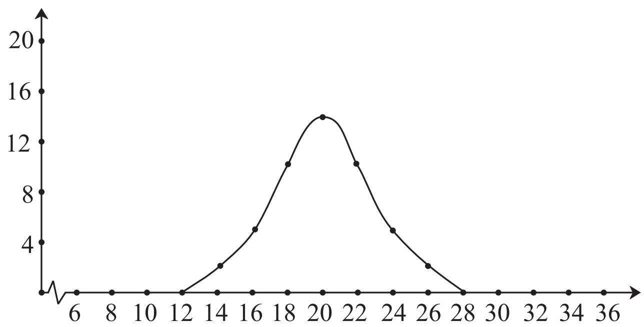

Fig 3.1

Symmetric distribution has mean = median = mode. For example length of rods, weight of individuals are usually approximately symmetric.

3.1.1 Asymmetric distribution (Positive skewness):

In cases where the values of x that are equidistant from mean have unequal frequencies, the distribution is said to be asymmetric . An asymmetric distribution is said to be positively skewed if the frequencies of the values of the variable lower than the mean have high frequencies compared to the frequencies associated with the values of variables higher than the mean.

In a visual graph, the right tail of such a distribution is longer and is called positively skewed distribution as the tail points to the positive end (higher side) of the x axis. (fig. 3.4).

It is quite logical that in such a distribution mean > Median> mode.

Salaries of employees are often positively skewed because of very few people getting high salaries. The high frequencies associated with the lower values of *x* are not offset by corresponding frequencies of the higher values of *x*.

Example and graph for positively skewed distribution:

X (variable)	14	16	18	20	22	24	26
f (frequency)	2	8	17	10	7	4	2

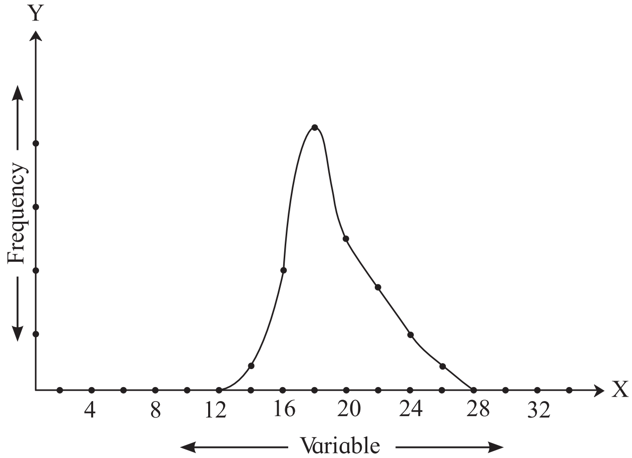

Fig 3.2

3.1.2 Asymmetric (Negative skewness):

An asymmetric distribution is said to be negatively skewed if the frequencies of the values of variables higher than mean have high frequencies compared to the frequencies associated with the variables lower than the mean.

The left tail of such a distribution is longer than the right tail (hence negatively skewed) as the tail points to the negative end of the *x* axis. (fig. 3.5).

In such a distribution mode> median > mean.

The lifespan of individuals are often negatively skewed because the frequencies associated with the higher values of *x* is high as compared to the frequencies associated with lower values of *x*.

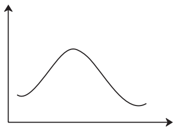

Fig 3.3

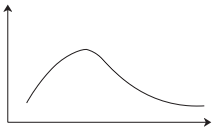

Fig 3.4

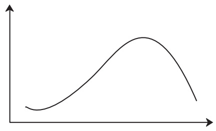

Fig 3.5

3.2 Measures of Skewness:

3.2.1 Karl Pearson's Coefficient of skewness (Pearsonian Coefficient of skewness):

The distance between mean and mode along with the sign (indicated by + or -) can be considered as the indicator of skewness .However with such a measure i.e. Mean – Mode, it is not possible to compare distributions of different nature.

For example : We cannot compare the skewness of weights of new born human babies and weights of new born lion cubs. These are two different things and hence cannot be compared.

To make comparison in such situations possible, we divide the difference of mean and mode by standard deviation. Karl Pearson suggested this method and hence it is known as Pearsonian coefficient of skewness.

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

3.2.2 Features of Pearsonian coefficient:

Sk_p is free from the units of the data values and also free from the size of values. Hence, comparison is possible across various distributions of different size.

For symmetric distribution, $Sk_p = 0$

For Positively skewed distribution $Sk_p > 0$

For negatively skewed distribution $Sk_p < 0$

Most Sk_p values lie between -1 and 1 ; all Sk_p values lie between -3 and 3.

In cases where mode is indeterminate we may use the following:

Mean – Mode = 3(mean-median);

Sk_p can be defined as:

$$Sk_p = 3\left(\frac{\text{Mean} - \text{Median}}{\text{S.D.}}\right)$$

SOLVED EXAMPLES

Ex.1 Following is the data for distribution of profits (in lakhs of rupees) of firms. Find Skp.

Profit	10-20	20-30	30-40	40-50	50-60
No. of firms	12	18	25	10	7

(Given $\sqrt{112.96} = 10.6283$)

Solution : The class indicated by 30-40 is modal class as it has maximum frequency

f_m is the frequency of modal class = 25

f_1 is frequency of pre-modal class = 18

f_2 is the frequency of post modal class = 10

h = class-width = 10

L = modal class lower limit = 30

$$\text{Mode} = L + \frac{(f_m - f_1).h}{2f_m - f_1 - f_2}$$

$$= 30 + \frac{(25-18).(10)}{(2)(25)-18-10}$$

$$= 30 + \frac{70}{22} = 33.18$$

Class	Frequency (f_i)	Class Mark (x_i)	$u_i=(x_i-35)/10$	f_iu_i	$f_iu_i^2$
10-20	12	15	-2	-24	48
20-30	18	25	-1	-18	18
30-40	25	35	0	0	0
40-50	10	45	1	10	10
50-60	7	55	2	7	14
Total	N=72			-25	90

Table 3.1

$$\bar{u} = \frac{\Sigma fiui}{N} = \frac{-25}{72} = -0.347$$

$$\bar{x} = 35+10\,\bar{u} = 35\text{-}3.47 = 31.53$$

$$\sigma_u^2 = \frac{\Sigma f_i u_i^2}{N} - (\bar{u})^2 = \frac{90}{72} - (-0.347)^2$$

$= 1.25 - 0.1204 = 1.1296$

$\sigma_x^2 = (10^2)\,\sigma_u^2 = (100)(1.1296) = 112.96$

$\sigma_x = \sqrt{112.96}$

$= 10.6283$

$$Sk_p = \frac{(\text{Mean-Mode})}{\text{S.D.}}$$

$$= \frac{31.53 - 33.18}{10.6283}$$

$$= \frac{-1.65}{10.6283}$$

$= -0.1552$; $Sk_p < 0$, implies negative skewness

3.2.3 Bowley's coefficient of skewness:

Prof. Bowley developed this relative measure of skewness. The formula for which is given by:

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

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

Let's decipher this further; For a symmetric distribution, the value that Sk_b will assume is 0 as Q_1 , Q_2 and Q_3 are equidistant (because Q_3-Q_2=Q_2-Q_1 for this case).

For a positive skewed distribution Q_1 is nearer to Q_2 as compared to Q_3 because the values are accumulated towards the lower part (left hand side) of the distribution.

Hence it follows that $(Q_3\text{-}Q_2)>(Q_2\text{-}Q_1)$. This implies that he numerator is positive and as the denominator is inherently positive, SK_b is positive. Conversely, for a negatively skewed distribution, Q_3 is nearer to Q_2 than Q_1 , hence Q_3-Q_2< Q_2-Q_1, therefore the value of Sk_b is negative (because the numerator assumes negative value).

The relative positioning of Q_1, Q_2 and Q_3 is as shown in the figure below for positive and negative skewed distribution:

Positive skew:

$(Q_3-Q_2) > (Q_2-Q_1)$

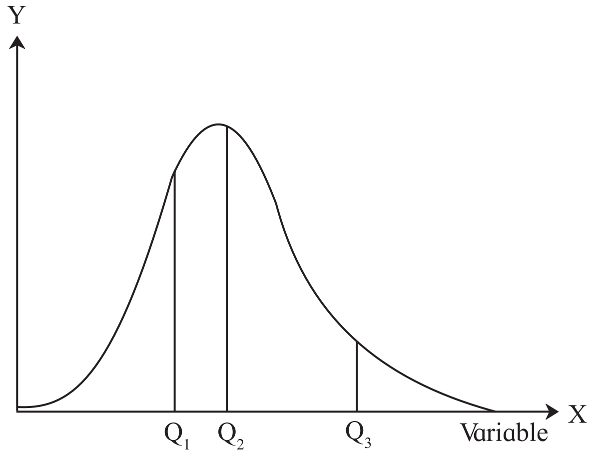

Fig 3.6 Positively Skewed Distribution

Negative Skew:

$(Q_3-Q_2) < (Q_2-Q_1)$

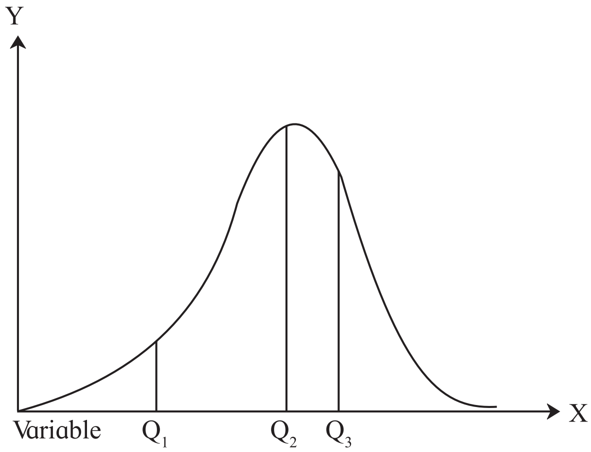

Fig 3.7 Negatively Skewed Distribution

Remarks:

1) $Sk_b > 0$ is an indicator of positively skewed distribution.

Similarly, $Sk_b < 0$ is an indicator of negatively skewed distribution.

$Sk_b = 0$ indicated symmetric distribution.

2) Sk_b lies between -1 and 1.

3) Sk_b can be computed for distributions with open ended classes at extremity as well. As these open ended classes do not impact quartiles on which the SKb is based. To clarify this let's look at the following example.

SOLVED EXAMPLES

Ex.1 If $Q_1 = 80$, $Q_2 = 100$, $Q_3 = 120$, find Bowley's coefficient of skewness.

Solution :

$$Sk_b = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$= \frac{120 + 80 - 2(100)}{40} = \frac{200 - 200}{40} = \frac{0}{40}$$

$$= 0$$

The Bowley's coefficient of skewness is 0, hence the data is symmetric.

Ex.2 Calculate Bowley's coefficient of skewness for the following distribution of weekly wages of workers.

Wages	Below 300	300-400	400-500	500-600	600-700	Above 700
No. of Workers	5	8	18	35	27	7

Class	Frequency	Less Than Type cumulative frequency
Below 300	5	5
300-400	8	13
400-500	18	31
500-600	35	66
600-700	27	93
Above 700	7	100
Total	N=100	

Table 3.2

Q_1 = value of $\left(\frac{N}{4}\right)^{th}$ observation.

= value of 25^{th} observation.

25^{th} observation belongs to the class 400-500.

Its frequency (f) is equal to 18 and cumulative frequency of the previous class is equal to 13.

$$Q_1 = L + \frac{h.(N/4 - C.F.)}{f}$$

$$= 400 + \frac{(100)(25-13)}{18}$$

$$= 466.67$$

$$Q_2 = \text{value of } \left(\frac{N}{2}\right)^{th} \text{ observation}$$

= value of 50^{th}observation.

50^{th} observation is from class 500 – 600

f=35 and C.F. = 31

$$Q_2 = L + \frac{h.(N/2 - C.F.)}{f}$$

$$= 500 + \frac{(100)(50-31)}{35}$$

$$= 554.28$$

$$Q_3 = \text{value of } \left(\frac{3N}{4}\right)^{th} \text{ observations}$$

= value of 75^{th} observation

75^{th} observation belongs to class 600 to 700

f = 27, C.F. = 66

$$Q_3 = L + \frac{h.(3N/4 - C.F.)}{f}$$

$$= \frac{(100)(75-66)}{27}$$

$$= 633.33$$

Bowley's coefficient of skewness is given by

$$Sk_b = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$= \frac{633.33 + 66.67 - 2(554.28)}{633.33 - 466.67}$$

$$= \frac{1100 - 1108.56}{166.66}$$

$$= -0.05$$

Ex.3 For a skewed distribution

Mean = 100, median = 98.5 and SD = 9. Find the mode and the Pearsonian coefficient of skewness (Sk_p) of the distribution.

Mean – mode = 3(mean –median)

100-mode = 3(100-98.5)

Mode = 100 – 4.5

= 95.5

$$Sk_p = \frac{\text{(Mean-Mode)}}{\text{S.D.}}$$

$$= \frac{100\text{-}95.5}{9}$$

$$= \frac{4.5}{9}$$

$$= 0.5$$

$Sk_p > 0$, the distribution is positively skewed.

Ex.4 Mode is greater than the mean by 7 and the variance is 100. Compute the coefficient of skewness. Is the data positively or negatively skewed?

Solution:

Mode – Mean = 7

Mean – mode= –7

Variance = 100 ----as given.

S.D. = 10

$$SK_p = \frac{\text{(Mean-Mode)}}{\text{S.D.}}$$

$$= \frac{-7}{10}$$

$$= -0.7.$$

$Sk_p < 0$, the distribution is negatively skewed.

Ex.5 If Arithmetic mean = 214

Mode = 210

Variance = 196

Karl Pearson's coefficient of skewness and nature of distribution is given by:

$$Sk_p = \frac{(\text{Mean-Mode})}{\text{S.D.}}$$

$$= \frac{214 - 210}{\sqrt{196}}$$

$$= \frac{-4}{14}$$

$$= -0.2857 < 0$$

$\therefore$ The distribution is negatively skewed.

Ex.6 For a frequency distribution, the lower quartile is 35 and median is 40. If the distribution is symmetric, find the upper quartile.

Lower quartile = Q_1 = 35

Median =40

Sk_b = 0 ----------the distribution is symmetric.

$$Sk_b = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

$$0 = Q_3 + Q_1 - 2Q_2$$

$$Q_3 = 2Q_2 - Q_1$$

$$= 2(40) - 35$$

$$= 80 - 35$$

$$= 45$$

The upper quartile Q_3 = 45.

5) Q_1 = 15

Q_2 = 21

Q_3 = 29

Calculate Bowley's coefficient of skewness

$$Sk_b = \frac{29 + 15 - 42}{29 - 15}$$

$$= \frac{2}{14}$$

$$= 0.143 > 0, \text{ Hence the}$$

the distribution is positively skewed

ADDITIONAL SOLVED EXAMPLES

Ex.1 Obtain the relation between Mean and Mode so that Distribution is negatively skewed.

Solution : Since we want relation between Mean and Mode therefore we use the formula of Karl Person's coefficient of skewness.

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

$\therefore$ (Sk_p) (S.D.) = Mean − Mode

S.D. is always positive and distribution is negatively skew.

$\therefore$ Mean − Mode < 0

$\therefore$ Mean < Mode

Which is the required relation.

Ex.2 Obtain the relation between upper quartile, lower quartile and median quartile so that Distribution is symmetric.

Solution : We wan't the relation between quartile so we use Bowley's coefficient of skewness

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

Since distribution is symmetric $\therefore$ $Sk_b = 0$

$\therefore$ $Q_3 - Q_2 = Q_2 - Q_1$

$2Q_2 = Q_1 + Q_3$

$\therefore$ Q_1, Q_2, Q_3 are in A.P.

Ex.3 Q_1, Q_2, Q_3 are in A.P. then discuss the nature of skewness

Solution : Simplest A.P. is 1, 2, 3

$Q_1 = 1$, $Q_2 = 2$, $Q_3 = 3$

Here we have to use Bowley's coefficient of skewness

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

$$= \frac{(3-2)-(2-1)}{(3-2)+(2-1)}$$

$$= \frac{1-1}{1+1} = \frac{0}{2}$$

$$= 0$$

$\therefore$ The Distribution is symmetric.

Ex.4 Discus the nature of distribution of Mean >Mode.

Solution : Here we use Karl Pearson's coefficient of skewness because we have relation between mean and mode

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

As S.D. is always positive sign of Sk_p depends on sign of Mean-Mode. We have Mean > Mode

$\therefore$ Mean − Mode > 0

$\therefore$ $Sk_p > 0$

$\therefore$ Distribution is positively skewed.

EXERCISE 3.1

1) For a distribution, mean =100 mode = 127 and SD = 60. Find the Pearson coefficient of skewness Sk_p.

2) The mean and variance of the distribution is 60 and 100 respectively. Find the mode and the median of the distribution if $Sk_p = -0.3$

3) For a data set, sum of upper and lower quartiles is 100, difference between upper and lower quartiles is 40 and median is 30. Find the coefficient of skewness.

4) For a data set with upper quartile equal to 55 and median equal to 42. If the distribution is symmetric, find the value of lower quartile.

5) Obtain coefficient of skewness by formula and comment on nature of the distribution.

Height in inches	No. of Females
Less than 60	10
60-64	20
64-68	40
68-72	10
72-76	2

6) Find Sk_p for the following set of observations.

   17,17,21,14, 15, 20, 19, 16, 13, 17 , 18

7) Calculate Sk_b for the following set of observations of yield of wheat in kg from 13 plots:

   4.6, 3.5, 4.8, 5.1, 4.7, 5.5, 4.7, 3.6, 3.5, 4.2, 3.5, 3.6, 5.2

8) For a frequency distribution $Q_3 - Q_2 = 90$

   And $Q_2 - Q_1 = 120$

   Find Sk_b

Let's Remember

Karl Pearsons Coefficient of skewness (Pearsonian Coefficient of skewness)

$$Sk_p = \frac{\text{Mean} - \text{Mode}}{\text{S.D.}}$$

If mode is intermediate then

$$\text{Mean} - \text{Mode} = 3\left(\frac{\text{Mean-Median}}{\text{S.D.}}\right)$$

$$\therefore \; Sk_p = 3\left(\frac{\text{Mean-Median}}{\text{S.D.}}\right)$$

- Features of Pearsonian Coefficient :

  (1) For Symmetric distribution $Sk_p = 0$

  (2) For Positively skewed distribution $Sk_p > 0$

  (3) For Negatively skewed distribution $Sk_p < 0$

  (4) Most Sk_p values lie between - 1 and 1. All Sk_p values lie between - 3 and 3

- Bowley's coefficient of skewness :

$$Sk_b = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

- Features of Bowley's coefficient :

  (1) If $Sk_b > 0$ then the distribution is positively skewed.

  (2) If $Sk_b < 0$ then the distribution is negatively skewed.

  (3) If $Sk_b = 0$ then the distribution is symmetric.

  (4) $-1 < Sk_b < 1$.

MISCELLANEOUS EXERCISE - 3

1. For a distribution, mean = 100, mode = 80 and S.D. = 20. Find Pearsonian coefficient of skewness Sk_p.

2. For a distribution, mean = 60, median = 75 and variance = 900. Find Pearsonian coefficient of skewness Sk_p.

3. For a distribution, $Q_1 = 25$, $Q_2 = 35$ and $Q_3 = 50$. Find Bowley's coefficient of skewness Sk_b.

4. For a distribution, $Q_3 - Q_2 = 40$, $Q_2 - Q_1 = 60$ Find Bowley's coefficient of skewness Sk_b.

5. For a distribution, Bowley's coefficient of skewness is 0.6. Find the upper and lower quartiles.

6. For a frequency distribution, the mean is 200, the coefficient of variation is 8% and Karl Pearsonian's coefficient of skewness is 0.3. Find the mode and median of the distribution.

1. Calculate Karl Pearsonian's coefficient of skewness Sk_p from the following data :

Marks above	0	10	20	30	40	50	60	70	80
No of students	120	115	108	98	85	60	18	5	0

2. Calculate Bowley's coefficient of skewness Sk_b from the following data :

Marks above	0	10	20	30	40	50	60	70	80
No of students	120	115	108	98	85	60	18	5	0

3. Find Sk_p for the following set of observations 18,27,10,25,31,13,28.

4. Find Sk_b for the following set of observations 18,27,10,25,31,13,28.

Activity 3.1

Collect marks in mathematics subject of 12 students in your class and arrange data in ascending order.

Complete the following table and write your comment?

$Q_1 =$ ☐, $Q_2 =$ ☐, $Q_3 =$ ☐ $Sk_b =$ ☐

Activity 3.2

Plot the points and draw free hand curve using given data

x	10	20	30	40	50	60	70	80	90	100
f	3	9	16	14	13	10	8	5	5	2

Write your comment about skewness and write relation between mean and mode.

❖ ❖ ❖