SKEWNESS

The word skewness is opposite of symmetry and its presence tell us that a particular distribution is not symmetrical or in other words it is skewed. The word skewness can be understood by the following definitions given by different statistician & Economist :-

1) As per Croxten and cowden, "When a series is not symmetrical it is skewed"

2) According to Garrett, " A distribution is said to be skewed when the mean and Median ~~and mode are~~ fall at different points in the distribution and the balance or centre of gravity is shifted to one side or the other.

3) In the words of Simpson and Kafka, " Measures of skewness tell us the direction and the extent of skewness. In symmetrically distribution the arithmatic mean, median and mode are identicle. The more the mean moves away from mode, the larger the asymmetry or skewness.

From the above definition it is clear that skewness is the lack of symmetry. Measures of skewness indicates the difference between the manner in which items are distributed in a particular distribution compared with symmetrical or normal distribution. In a symmetrical distribution frequencies go on an increasing upto a point and then begin to decrease in the same fashion. There are various posible patterns of symetrical distribution and normal distribution which is bell shaped in one of these. Some of the posible patterns of symetrical distribution are as follows :-

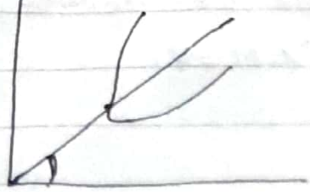


Fig - I

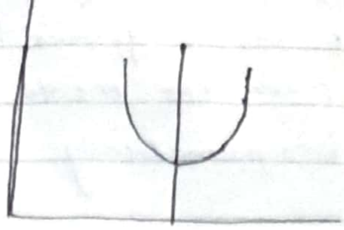


Fig - II

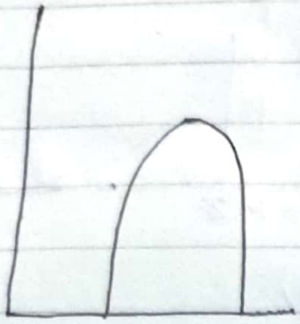


Fig - III

Fig. IV

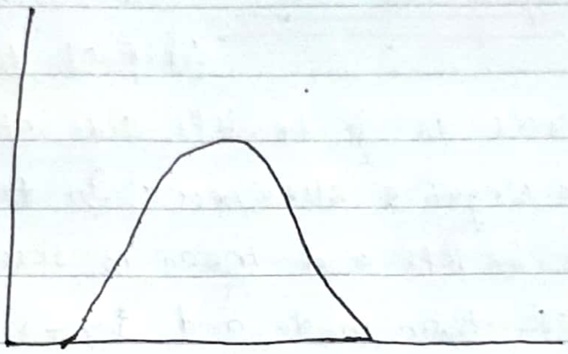


Fig. V (symmetrical bell shaped distribution)

In a symmetrical distribution, Mean = Median = Mode and they lies at the centre of the distribution. When symmetry is disturbed these values are pulled apart. The skewness may be broadly of two types:—

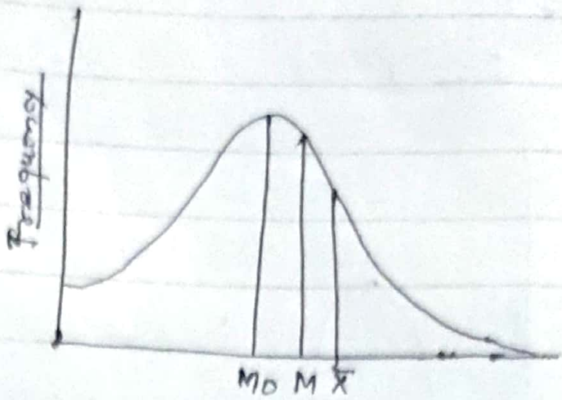
(1) Positive skewness :-

(2) Negative skewness :—

(1) Positive skewness :— A distribution in which more than half of the area under the curve is to be right side of the mode it is said to be positive skewed distribution. In this type of skewness the right side

mean is greater than median and the median is greater than the mode, and $Q_3 - M > M - Q_1$

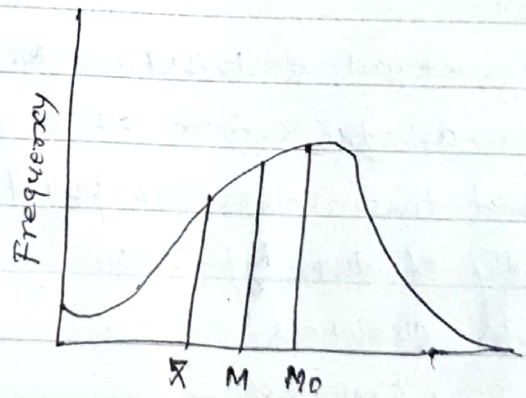
Diagramatically,



$(\bar{X} > M > Mo)$

(2) Negative Skewness :- A distribution in which more than half of the area under the distribution curves is to be left side of the mode, it is said to be Negative skewness. In this case, the elongated tail is to be left and mean is less than the median which is less than mode and $Q_3 - M < M - Q_1$

Diagramatically,



Measures of skewness :- Various measures of skewness are being developed. Each new one is an important on the earlier one. A few important measures of skewness are :-

Scanned with CamScanner

(1) First measures of Skewness :— They are based on the assumption that is in a ~~skewness~~ skewed distribution the value of mean, median and mode do not coincide and the difference between any two of these values indicates the extent of skewness. As per these method skewness can be measured by anyone of these :—

(a) Mean − Mode (b) Mean − Median (c) Median − Mode

Drawbacks :— This method suffers from the following drawbacks :—

(1) The result obtained cannot be compared with others because they are expressed in the same units as the items of distribution.

(2) Similar looking curves may be show large differences in the result.

Relative Measure of distribution :— They are obtained by dividing the absolute measures by anyone of measures of dispersion. They are also called co-efficient of skewness. Various relative measures of skewness are :—

(a) Karl Pearsons co-efficient of Skewness :—

This method uses standard deviation as a divisor. It is expressed by the following formula :—

$$\boxed{S_p = \frac{\bar{X} - M_o}{\sigma} \text{ or } \frac{\text{Mean} - \text{Mode}}{\text{standard deviation}}} \quad \text{— I}$$

According to Pearson if mode is ill defined the following formula should be used

$$\boxed{\frac{3(\bar{X} - M_o)}{\sigma} \text{ or } \frac{3(\text{Mean} - \text{mode})}{\text{standard deviation}}} \quad \text{— II}$$

Scanned with CamScanner

If the value comes out to be 0 (zero), the distribution is symmetrical. In case of skewness the value lies between ±(1) for formula – I, and ±3 for formula (ii)

(b) Bowley's co-efficient of skewness :– It based upon the value of Quartiles and Median. When $Q_3 - M = M - Q_1$, the distribution is said to symmetrical.

Bowleys relative measures of skewness (S_B)

$$\boxed{S_B = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}}$$

The theoretical ~~limits~~ limits of the above formula are ±1. If the distribution is symmetrical, the value comes out to be zero.

Example :– Calculate Karl Pearson's co-efficient of skewness from the following data and comment on its bias.

Value	100	200	300	400	500	600	700
Frequency	1	5	12	22	17	9	4

Soln.

x	f	$(dx)\frac{(x-A)}{100}$	dx^2	fdx	fdx^2
100	1	−3	9	−3	9
200	5	−2	4	−10	20
300	12	−1	1	−12	12
400	22	00	00	00	00
500	17	+1	1	+17	17
600	9	+2	4	+18	36
700	4	+3	9	+12	36
	N=70			$\Sigma fdx = 22$	Σfdx^2 130

Let A = 400

i = 100

$$\bar{X} = A + \frac{\Sigma f dx}{N} \times i$$

$$= 400 + \frac{22}{70} \times 100$$

$$= 400 + 0.3143 \times 100$$

$$400 + 31.43$$

$$\boxed{\bar{X} = 431.43}$$

Highest frequency is 22 $\therefore$ Mode $= 400$

$$\sigma = \sqrt{\left(\frac{\Sigma f dx^2}{N} - \left(\frac{\Sigma f dx}{N}\right)^2\right)} \times i$$

$$= \sqrt{\frac{130}{70} - \left(\frac{22}{70}\right)^2} \times 100$$

$$\sigma = 132.6$$

$$S_p = \frac{\bar{X} - M_o}{\sigma} = \frac{431.43 - 400}{132.6} = +0.24$$

Karl Pearson's co-efficient Sp comes out to be +0.24 this means the given distribution is positively skewed.

Example Find out co-efficient of skewness from the following table:-

Marks obtained:-	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f	5	10	25	40	26	11	3

Scanned with CamScanner

Soln.

Marks Obtained	Frequency	MV	dx	fdx	fdx^2
0 - 10	5	5	-30	-150	4500
10 - 20	10	15	-20	-200	4000
20 - 30	25	25	-10	-250	2500
30 - 40	40	35	0	0	0
40 - 50	26	45	+10	+260	2600
50 - 60	11	55	+20	220	4400
60 - 70	3	65	+30	+90	2700
	N = 120			Σfdx -30	Σfdx^2 = 20700

Let A = 35

$$\bar{X} = A + \frac{\Sigma fdx}{N}$$

$$= 35 + \frac{-30}{120}$$

$$= 35 - \frac{30}{120}$$

$$= 35 - 0.25 = \boxed{34.75}$$

Highest frequency is 40 ∴ Mode class is 30-40

$$Mo = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 30 + \frac{40 - 25}{2 \times 40 - 25 - 26} \times 10$$

$$= 30 + \frac{15}{80 - 51} \times 10$$

$$= 30 + \frac{150}{29}$$

$$30 + 5.17 = \boxed{35.17}$$

$$\sigma = \frac{1}{N}\sqrt{\Sigma f dx^2 \times N - (\Sigma f dx)^2}$$

$$= \frac{1}{120}\sqrt{20700 \times 120 - (-30)^2}$$

$$= \frac{1}{120}\sqrt{248400 - 900}$$

$$= \frac{1}{120}\sqrt{2483100}$$

$$= \frac{1}{120} \times 1575.79 = \boxed{13.13}$$

$$S_P = \frac{\bar{X} - M_o}{\sigma}$$

$$= \frac{34.75 - 35.17}{13.13} = \frac{0.42}{13.13} = \boxed{-0.032}$$

$$\boxed{S_P = -0.032}$$

Co-efficient of skewness by Bowley :-

example : - Calculate Quartile co-efficient of Skewness of the following observation :-

18, 16, 11, 24, 14, 18, 25, 22, 28, 24, 29, 17, 10, 20, 12.

Soln - 10, 11, 12, 14, 16, 17, 18, 18, 20, 22, 24, 24, 25, 28, 29

$$Q_1 = \frac{N+1}{4} = \frac{15+1}{4} = \frac{16}{4} = 4\text{th item ie } 14 \therefore Q_1 = 14$$

$$Q_3 = \frac{3(N+1)}{4} = \frac{3(15+1)}{4} = \frac{3 \times 16}{4} = \frac{48}{4} = 12 \quad 12^{th} \text{ item ie. } 24$$

$$Q_3 = 24$$

$M = \frac{N+1}{2} = \frac{15+1}{2} = \frac{\cancel{16}^{8}}{\cancel{2}} = 8^{th}$ item i.e. 18

$M = 18$

$$S_B = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}$$

$$= \frac{(24-18) - (18-14)}{(24-18) + (18-14)}$$

$$= \frac{(6-4)}{(6+4)} = \frac{2}{10} = 0.20$$

$\boxed{S_B = 0.20}$ Ans.

Example :- Using Bowley's method calculate co-fficient of skewness from the following data

Marks :-	3	5	7	10	12	15
Frequency	3	6	10	6	3	2

Soln

Marks	f	cf
3	3	3
5	6	9
7	10	19
10	6	25
12	3	28
15	2	30
	N=30	

$Q_1 = \frac{N+1}{4} = \frac{30+1}{4} = \frac{31}{4} = 7.75^{th}$.

7.75 lies on cf - 9 ∴

$\therefore Q_1 = 5$

$Q_3 = \frac{3(N+1)}{4} = \frac{3(30+1)}{4} = \frac{3 \times 31}{4}$

$= \frac{93}{4} = 23.25^{th}$

23.25^{th} item lies on cf 25

$\therefore Q_3 = 10$

$M = \frac{N+1}{2} = \frac{30+1}{2} = \frac{31}{2} = 16.5$

16.5^{th} item lies on cf 19

$\therefore M = 7$

$$S_B = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)}$$

$$S_B = \frac{(10-7) - (7-5)}{(10-7) + (7-5)}$$

$$= \frac{3-2}{3+2} = \frac{1}{5} = 0.20$$

$\boxed{S_B = 0.20}$ Ans.

Example :- From the following data calculate the Quartile measure of Skewness for the following series :-

Class Interval	Frequency f	c f
0-10	4	4
10-20	5	9
20-30	7	16
30-40	12	28
40-50	27	55
50-60	22	77
60-70	14	91
70-80	12	103
80-90	10	113
90-100	7	120
	N=120	

$M = \frac{N}{2} = \frac{120}{2} = 60$th item.

60th item lies on c.f = 77 i.e. class Interval 50-60

$$M = L_1 + \frac{i}{f}(m - c)$$

$$= 50 + \frac{10}{22}(60-55)$$

$$= 50 + \frac{10}{22}(5)$$

$$= 50 + \frac{50}{22}$$

$$M = 50 + 2.27 = \boxed{52.27}$$

$Q_1 = \left\{\frac{N}{4}\right\} = \frac{120}{4} = 30$th item.

30th item lies in c.f. 55, class Interval - 40-50

$$Q_1 = L_1 + \frac{i}{f}(Q_1 - c)$$

$$= 40 + \frac{10}{27}(30-28)$$

$$= 40 + \frac{10}{27}(2)$$

$$= 40 + \frac{20}{27}$$

$$= 40 + 0.741$$

$$\boxed{Q_1 = 40.75}$$

$$Q_3 = \frac{3(N)}{4} = \frac{3 \times 120}{4} = \frac{\cancel{360}^{90}}{\cancel{4}}$$

90th item which is lies in c.f 91, class interval 60-70

$$Q_3 = L_1 + \frac{i}{f}(Q_3 - c)$$

$$= 60 + \frac{10}{14}(90 - 77)$$

$$= 60 + \frac{10}{14}(13)$$

$$= 60 + \frac{130}{14}$$

$$= 60 + 9.3$$

$$\boxed{Q_3 = 69.3}$$

$$S_g = \frac{Q_3 + Q_1 - 2M}{Q_1 + Q_3}$$

$$= \frac{69.3 + 40.75 - 2 \times 52.27}{40.75 + 69.3}$$