Chapter : Skewness

Introduction

Measures of central tendency give us representative value of the distribution. These measure do not tell us about variability in distribution. Measures of dispersion measure variability in the distribution but these measures have one serious limitation. These measures don't tell us about bias or asymmetry in the distribution. In other words measures of central tendency also called as measures of first order and measure of dispersion also called as measure of second order do not give us the complete picture of the distribution. These two types of measure don't tell us about the bias in distribution. Two distribution may have same average and same variability but may have different bias. Measures of skewness are used to find and measures this bias.

Meaning

Let us study how different authors have defined skewness:

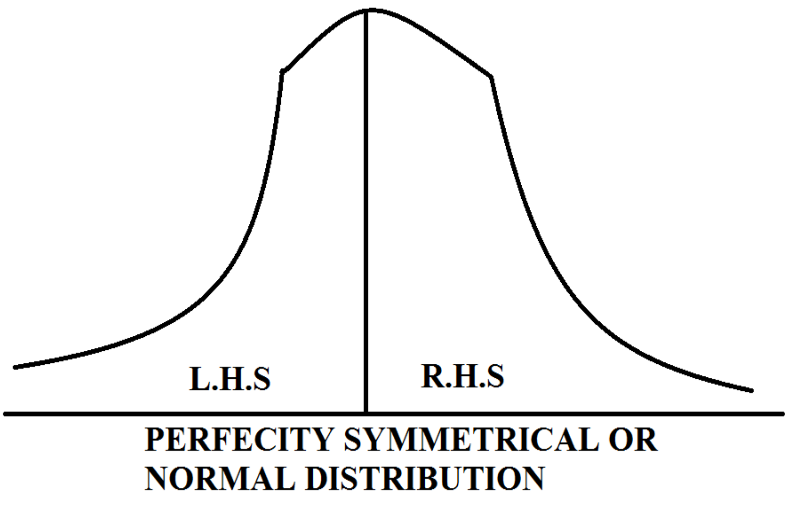
"When a series is not symmetrical it is said to asymmetrical or skewed."

"Sknewness is lack of symmetry"

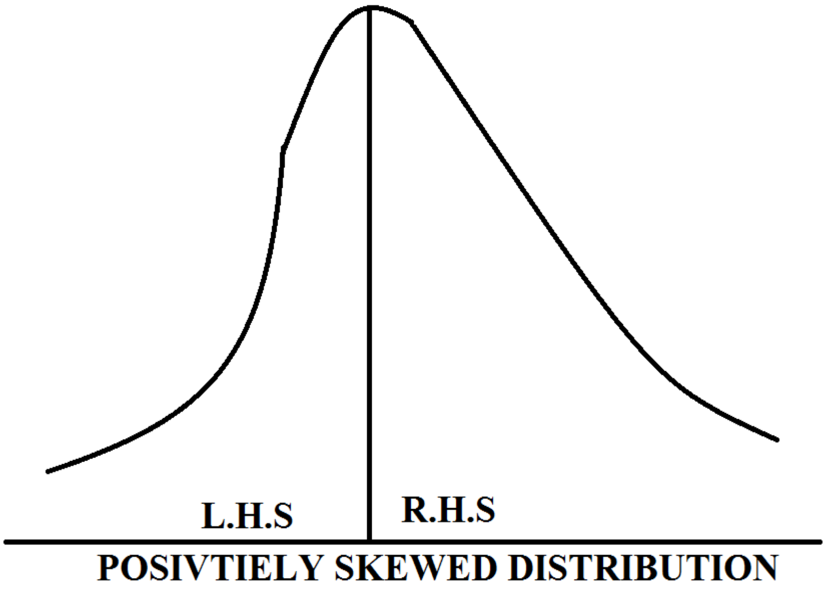
"Skewness refers to the asymmetry in the shape of a frequency distribution."

Thus we find that by skewness we mean the presence of bias or asymmetry in the distribution.

(A)

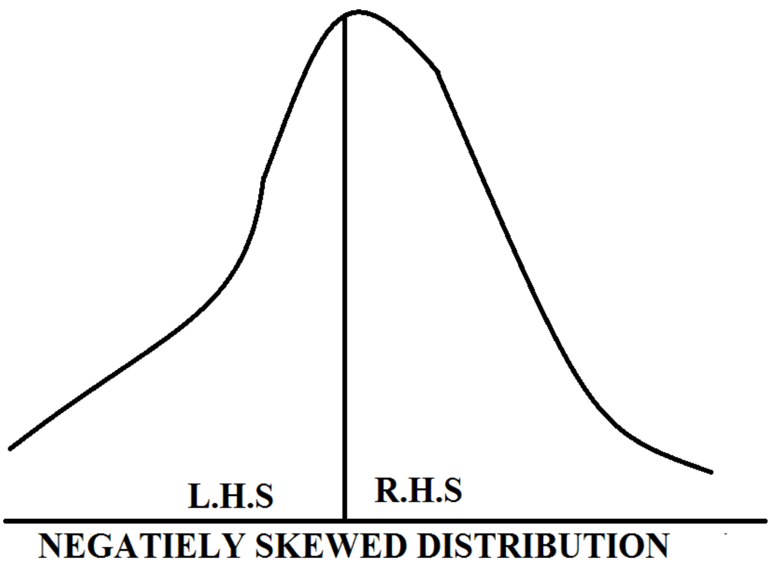


(B)



\

(C)



The meaning of skewness can be made clear with the help of flowing diagrams:

In distribution A, left hand side and right hand side of the curve are perfectly similar. There is no bias in the distribution. Distribution is perfectly symmetrical. Skewness is not present in the distribution. Thus type of distribution is called as normal distribution or unbiased distribution or perfectly symmetrical distribution.

In distribution B, the two sides are not perfectly similar. The distribution is biased towards right hand side. Positive skewness is present. The distribution will be called as positively skewed distribution.

In distribution C, also two sides of the distribution are not perfectly similar. The distribution is biased towards left hand side. Negative skewness is present in the distribution. The distribution is called as negatively skewed distribution.

OBJECTIVES OF SKEWNESS

Following are the objectives of skewness

1. It tells us whether the distribution is normal or not.
2. It gives us an idea about the nature and degree of concentration of observations

3. The empirical relation of mean, median and mode are based on a moderately skewed distribution
4. It tells us the direction and extent of asymmetry in a series and permit us to the compare two or more series with regard to these.

Testing Skewness:-Skewness in any distribution is tested with the help of any one of the following methods:

1. **By Graphic Method:-** In this method a given frequency distribution is plotted and a frequency curve is constructed. If this curve is similar to curve A (Discussed above), it means no skewness in present.
2. **By Comparing Mean, Median and Mode:** In this method firstly mean, median and mode is calculated. Then they are compared. Three cases arise:

(a) If Mean = Median = Mode, it implies there is no skewness in the distribution.

(b) If Mean > Median > Mode, it implies there is positive skewness in distribution .

(c) If Mean < Median < Mode, it implies there is negative skewness in the distribution.

Different Methods of Measuring Skewness:- Three methods are available for measuring skewness. These methods have their absolute version and relative version. For comparison we use relative versions.

1.**Karl Pearson's Method**

Absolute version:

$$\text{Skewness} = \text{Mean} - \text{Mode}$$

$$\text{Coefficient of Skewness} = \frac{\text{Mean} - \text{Mode}}{\text{Standard Deviation}}$$

In some case mode is ill define. We get two more than two modes. In such cases we use following relation for calculating mode.

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

Substituting this value

$$Skewness = \text{Mean} - [3\ \text{Median} - 2\text{Mean}]$$

$$\text{Skewness} = \text{MEan} - 3\text{Median} + 2\text{Mean}$$

$$\text{Skewness} = 3\text{Mean} - 3\text{Median}$$

$$\text{Skewness} = 3[\text{Mean} - \text{Media}]$$

$$\text{Cofficient of skewness} = \frac{3[\text{Mean} - \text{Median}]}{\text{Standard Deviation}}$$

Step. 1. Calculate mean of the distribution.

2.Calculate mode of the distribution.

3.Calculate standard deviation.

4.Put these values in the formula.

Karl pearsons' method of measuring skewness is considered the best among all the methods, since it is based upon mean and standard deviation.

II. Bowley's Method

Second method of measuring skewness is given by Prof. Bowley. It is biased upon quartiles. Therefore, it is also called as the quartile method of measuring skewness.

Absolute Version:

$$Skewness = Q_3 + Q_1 - 2\ Median$$

Relative Version:

$$Coefficient\ of\ skewness = \frac{Q_3 + Q_1 - 2Median}{Q_3 - Q_1}$$

Bowley’s method is not always dependable. However, if extreme items are present in the distribution or open end class intervals are there or problem is qualitative in nature, Bowley’s method gives better result than pearson’s coefficient since the later is not suitable for these conditions.

Steps 1. Calculate Q_1 $i.e$ first quartile

2.Calculate$Q_3 i.e$. third quartile

3. Calculate median

4. Substitute these values in the formula.

III. Kelly’s Method

This method is the slight modification of Bowleys's method. Like bolwley's it is also based upon positional measures of central tendency. But in place of quartiles percentiles are used in this method.

$$Skewness = P_{90} + P_{10} - 2Median$$

$or\ D_9 + D_1 - 2Median$p

Relative Version

$$\text{Coefficient of Skewness} = \frac{P_{90} + P_{10} - 2Median}{P_{90} - P_{10}}$$

$$\text{or } \frac{\mathrm{D_9 + D_1 - 2Median}}{\mathrm{D_9 - D_1}}$$

This Formula is not very popular. It is suitable under those conditions where Bowley's method is suitable.

Step. 1 Calculate P_{20} or Ninetheenth percentile

3.Calculate P_{10} or Tenth percentile

4.Substitute these in then formula.

Example:3Calculate Karl pearson's coefficient of shewness from the give below:

X 10 20 30 40 50 60 70

F 1 5 12 22 17 9 4

Solution:

X	F	A = 40 dX = X – A	C = 10 $d'X$	$fd'X$	$d'X^2$	$-fd'X^2$
10	1	-30	12	-3	9	9
20	5	-20	-2	-10	4	20
30	12	-10	-1	-12	1	12
40	22	0	0	0	0	0
50	17	10	1	17	1	17
60	9	20	2	18	4	36
70	4	30	3	12	9	36
	N =			22		130

	26					

$$Mean = A + \frac{\sum fd'X}{N}, C = 40 + \frac{22}{70} \times = 430.1$$

$$Mode = 40(By\ inspection)$$

$$\sigma = \sqrt{\frac{\sum fd'X^2}{N} - \left(\frac{\sum fd'X}{N}\right)^2} \times C$$

$$\sigma = \sqrt{\frac{130}{70} - \left(\frac{22}{70}\right)^2} \times 10 = 13.26$$

Karl pearson's coefficient of

$$Skewness = \frac{Mean - Mode}{\sigma} = \frac{43.1 - 40}{13.26}$$

Example:- Find out skewness and coefficient of skewness by karl pearson's method:

Month	Price of Milk (Rs./litre)
January	2.5
February	3.0
March	3.0

April	3.5
May	4.0
June	4.5
July	5.0
August	5.0
September	4.5
October	4.0
November	3.0
December	3.0

Solution

S. No	X	$X - \bar{X}$	$(X - \bar{X})^2$
1	2.5	-1.25	1.5625
2	3.0	-0.75	0.5625
3	3.0	-0.25	0.5625
4	3.5	0.25	0.0625
5	4.0	0.75	0.0625
6	4.5	1.25	1.5625

7	5.0	0.75	1.5625
8	5.0	1.25	0.5625
9	4.5	0.75	0.5625
10	4.0	0.25	0.0625
11	3.0	0.75	05625
12	3.0	0.75	0.5625
	$\sum X = 45.0$		$\sum (\bar{X})^2 = 8.2500$

$$\bar{X} = \frac{\sum X}{n} = \frac{45}{12} = Rs. 3.75$$

Mode = 3 Rs. (It occurs maximum times)

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{n}} = \sqrt{\frac{8.2500}{12}}$$

$$\sigma = \sqrt{0.6875} = 0.83$$

$$Skeweness = \bar{X} - Mode$$

$$Skeweness = 3.75 - 3.00 = Rs. 0.75$$

$$Coefficient\ of\ skewness = \frac{3.75 - 3.00}{0.83} = 0.90$$

Example: Find out coefficient of skewness by Bowley's Method.

S. No	1	2	3	4	5	6	7	8	9	10
X	15	22	25	30	31	35	40	42	50	54

Solution

S. No	X
1	15
2	22
3	25
4	30
5	31
6	35
7	40
8	42

9	50
10	54

$$Q_1 = size\ of\ \frac{N}{4} = \frac{100}{4} = 25th\ item$$

$$Q_1 = 2nd\ item + 0.75(3rd\ item - 2nd\ item)$$

$$Q_1 = 22 + 0.75(25 - 22) = 24.25$$

$$Q_3 = size\ of\ \frac{3(n+1)}{4} th\ item$$

$$Q_3 = size\ of\ \frac{3(10+1)}{4} th\ item$$

$$Q_3 = size\ of\ \frac{33}{4} = 8.25\ th\ item$$

$$Q_3 = size\ item + 0.25(9th\ item - 8th\ item)$$

$$Q_3 = 42 + 0.25(50 - 42)$$

$$Q_3 = 44$$

$$Median = size\ of\ \frac{10+1}{2} = 5.5th\ item$$

$$Median = 5th\ item + 0.5(6th\ item - 5th\ item)$$

$$Median = 31 + 0.5(35 - 31)$$

$$Median = 33$$

$$Bowley'scoefficent\ of\ skewness = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{44 + 24.25 - 66}{44 - 24.25} = \frac{2.25}{19.75} = 0.11$$

Example: From the following data find out Bowley's coefficient of skewness:

Size: 5– 7 8 – 10 11 – 13 14 – 16 17 – 19

Number of persons : 14 24 38 20 4

Solution: Class intervals are in inclusive form. For finding median and quartiles these will have converted into exclusive form.

X	f	c.f.
4.5 – 7.5	14	14

7.5 – 10.5 10.5 – 13.5 13.5 – 16.5 16.5 – 19.5	24 38 20 4	38 76 96 100
	N = 100	

$$Q_1 = \text{size of } \frac{\text{N}}{4} = \frac{100}{4}$$

$$= 25\text{th item. it lies in interval 7.5 to 10.5}$$

$$Q_1 = L + \frac{\frac{N}{4} - c.f.}{F} \times i$$

$$Q_1 = 7.5 + \frac{25 - 14}{24}.3 = 8.87$$

$$Q_3 = \text{size of } \frac{3\text{N}}{4} = \frac{3(100)}{4}$$

$$= 75 \text{ th item. it lies in interval 10.5 to 13.5}$$

$$Q_3 = L + \frac{\frac{3N}{4} - c.f.}{f} \times i$$

$$Q_2 = 10.5 + \frac{75-38}{38} \times 3 = 13.42$$

$$\text{Median} = \text{size of } \frac{\text{N}}{2} = \frac{100}{2}$$

$$= 50\text{th item. it lies in interval } 10.5 \text{to } 13.5$$

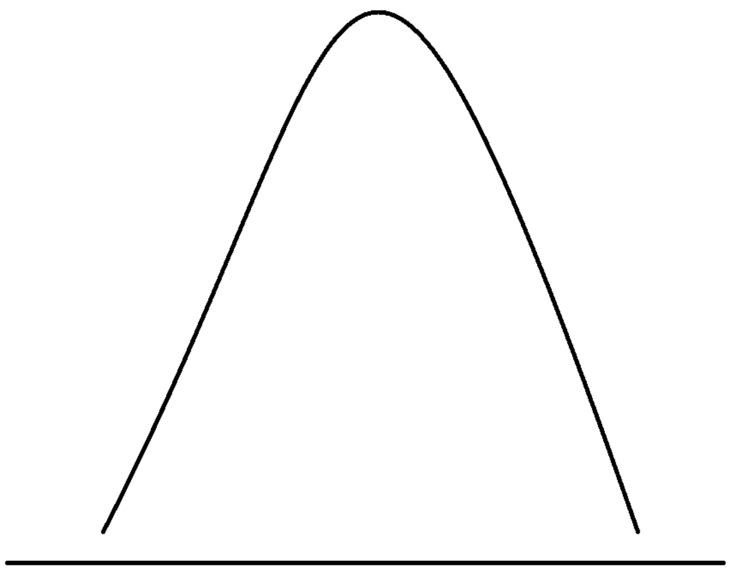
$$\text{Median} = \text{L} + \frac{\frac{\text{N}}{2} - \text{c.f.}}{\text{f}}.\text{i} \quad \text{Median}$$

$$= 10.5 + \frac{50-38}{38}.3 = 11.447$$

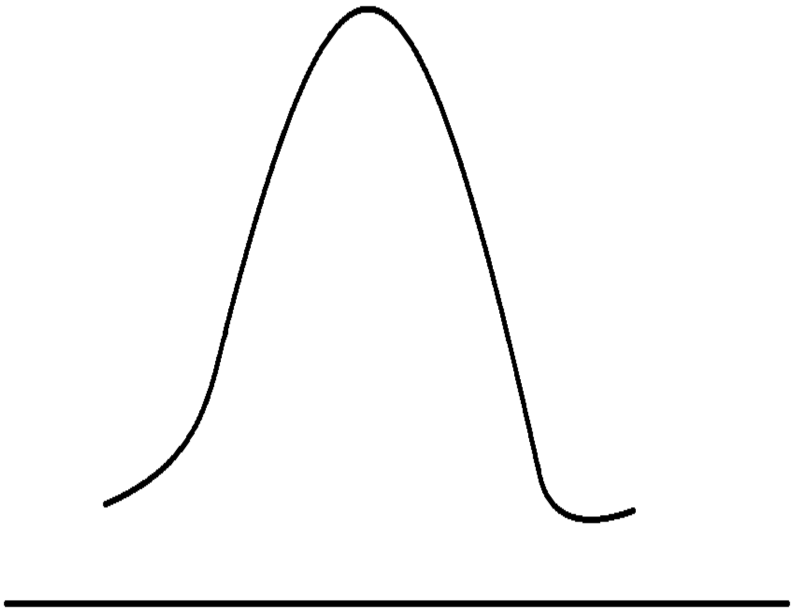
$$\text{Bowley}'\text{scoefficient of skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{13.42 + 8.87 - 2 \times 11.477}{13.42 - 8.87} = -0.13$$

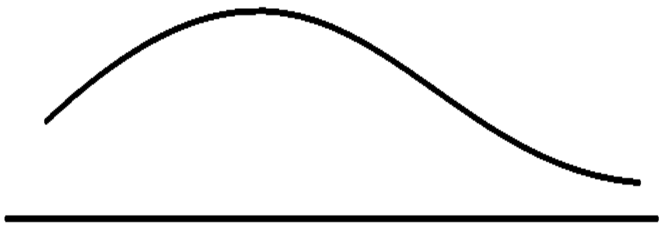
KURIOSIS



(A)



(C)



(C)

We have studied three measures to describe the characteristics of any distribution. These three measures are-measure of central tendency measures of dispersion and measures of skewness. In reality these three measures fail to describe any distribution completely. Two distributions may have same average, dispersion and skewness and yet may differ from one another. Infact in order to identify a distribution completely we require one more measure which has been called as ‘Kurtosis’ by Prof. Karl Pearson.

Meaning:- Kurtosis is a greek word. It means business. In simple word in statistics by kurtosis we mean the degree of flatness or peakedness in the region about mode of a distribution. Let us see how famous authors have defined kurotsis.

According to croxton and cowden, “A measure of kurtosis indicates the degree to which a curve of a frequency distribution is peaked on flat topped.”

According to C. H. Myres, “Kurotsis refers to the degree of peakedness of the humb of the distribution.”

Prof. Karl Pearson described three pattern of peakedness which are shown as below:

Lepto Kurtic Mesco Kurtic Platy Kurtic

(A)

Curve of type A which is highly peaked is called as Leptokurtic. It is said to lack kurtosis or to have negative kurtosis. It is like the leaping of Kangaroos.

Curve of type B which is neither too flat nor too peaked is called as mesokurtic. It is medium peaked. It is like a normal curve and shape of its humb is accepted as a standard one.

Curve of type C which is highly flat is called as platykurtic. It is said to procsses kurtosis in excess or has positive kurtosis.

Different between skewness and kurtosis: As Clear from above, kurtosis is different from skewness. Whereas skewness deals with the right or left tails of the distribution, kurtosis enables us to have an idea about the shape and nature of middle part of the distribution. In order words, kurtosis is concerned with the flatness or peakedness of the distribution.

Measurement of Kurtosis

To measure kurtosis Prof. Karl pearson gave two coefficients based upon moments. These two coefficients are Beta two (β_2) and Gamma two (γ_2). These two coefficients are defined as below:

$$\beta_2 = \frac{\mu_4}{{\mu^2}_2} = \frac{\mu_4}{\sigma^4}$$

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu^2{}_2} - 3 = \frac{\mu_4 - 3\sigma^4}{\sigma^4}$$

(i) If $\beta_2 > 3,\ i.e., \gamma_2 > 0$ distribution is leptokurtic.

(ii) If $\beta_2 = 3, i.e. \gamma_2 = 0$ distribution is mesokurtic.

(iii) If $\beta_2 < 3,\ i.e., \gamma_2 < 0$ distribution is platykurtic.

Uses of kurtosis

(i) The study of kurtosis helps us to know the nature of the distribution.

(ii) The study of kurtosis helps us to choose among the various averages. If distribution is mesokurtic mean is the best measure of central tendency. If distribution is leptokurtic, median is more suitable. If distribution platykurtic study of quartiles is more useful.

Example:- The following information is given regarding holdings of 5 farmers.

Farmer: A B C D E

Land Holdings (Acres): 2 3 7 8 10

(i) Find out different central moments

(ii) Calculate coefficient of skewness

(iii) Calculate measures of kurtosis

Solution: It is individual observation series.

S. No	X(Acres)	$X-\bar{X}$	$(X-\bar{X})^2$	$(X-\bar{X})^3$	$(X-\bar{X})^4$
1	2	-4	16	-64	256
2	3	-3	9	-27	81
3	7	1	1	-1	16
4	8	2	4	8	256
5	10	4	16	64	
	$\sum X = 30$	0	46	-18	610

$$\bar{X} = \frac{\sum X}{n} = \frac{30}{5} = 6(Acres)$$

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

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

$$\mu_2 = \frac{\sum(X - \bar{X})^3}{n} = \frac{-18}{5} = -3.6$$

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

(iv)Coefficients of skewness:

$$\beta_1 = \frac{{\mu^2}_3}{{\mu^3}_2} = \frac{(-3.6)^2}{9.2^3} = \frac{12.96}{778.688} = 0.0166, \gamma_1 = \sqrt{\beta_1}$$

$$= \sqrt{0.0166} = 0.128$$

Since μ_3 is negative, distribution is vely skewed

(iii)Measures of kurtosis:

$$\beta_2 = \frac{\mu_4}{{\mu^2}_2} = \frac{122}{9.2^2} = 1.4\ \gamma_2 = \beta_2 - 3 = 1.4 - 3 = -16$$

Since β_2 is less than three and γ_2 is -16, distribution is play kurtic or flat at the top.

Example: Marks off 100 students are given below. Using moments find measures of skewness and measures of kurtosis.

Marks : $0 - 10$ $10 - 20$ $20 - 30$ $30 - 40$ $40 - 50$

No. of students: 10 20 40 20 10

X	f	X Mid value	fX	$X - \bar{X}$	$f(X - \bar{X})$	$f(X - \bar{X})^2$	$f(X - \bar{X})^3$	$f(X - \bar{X})^4$
$0 - 10$	10 20	5 15	50 300	-20 -10	-200 -200	4000 2000	-8000 -20000	16000 200000

10 − 20 20 − 30 30 − 40 40 − 50	40 35 45	25 35 45	100 0 700 450	0 10 20	0 200 200	0 2000 4000	0 20000 80000	0 200000 160000 0
	10 0		250 0		0	1200 0	0	360000 0

$$\bar{X} = \frac{\sum fX}{N} = \frac{2500}{100} = 25$$

$$\mu_1 = \frac{\sum(X - \bar{X})}{N} = \frac{0}{100} = 0(Always)$$

$$\mu_2 = \frac{\sum(X - \bar{X})^2}{N} = \frac{12000}{100} = 120 = Vatiance$$

$$\mu_3 = \frac{\sum(X-\bar{X})^3}{N} = \frac{0}{100} = 0$$

$$\mu_4 = \frac{\sum(X-\bar{X})^4}{N} = \frac{3600000}{100} = 36000$$

Measures of skewness :

$$\beta_1 = \frac{\mu^2{}_3}{\mu^2{}_2} = \frac{0}{1203} = 0, \gamma_1 = \sqrt{\beta_1} = 0$$

Distribution is without skewness

Measures of Kurtosis:

$$\beta_2 = \frac{\mu_4}{\mu_2} = \frac{36000}{120^2} = 2.5, \gamma_2 = \beta_2 - 3 = 2.5 - 3$$

$$= -0.5$$