CHAPTER 04

Measures of Dispersion, Moments, Skewness and Kurtosis

Chapter Contents

You should read this chapter if you need to learn about:

zeestats@gmail.com

- Sometimes when two or more different data sets are to be compared using measure of central tendency or averages, we get the same result.

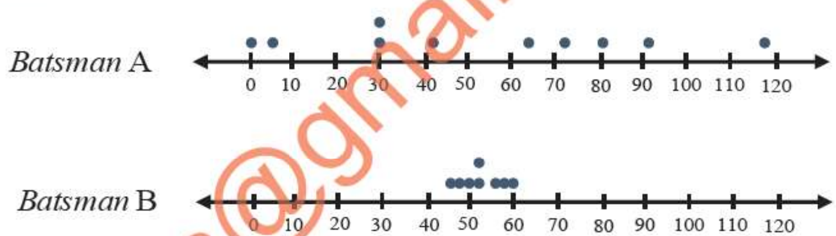
Consider the runs scored by two batsmen in their last ten matches as follows:

Batsman A: 30, 91, 0, 64, 42, 80, 30, 5, 117, 71
Batsman B: 53, 46, 48, 50, 53, 53, 58, 60, 57, 52

Clearly, mean of the runs scored by both the batsmen A and B is same i.e. 53

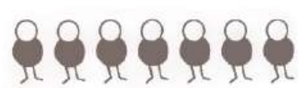
Can we say that the performance of two players is same? Clearly No, because the variability in the scores of batsman A is from 0 to 117, whereas, the variability of the runs scored by batsman B is from 46 to 60.

Let us now plot the above scores as dots on a number line. We find the following diagrams:



We can see that the dots corresponding to batsman B are close to each other and is clustering around the measure of central tendency (mean), while those corresponding to batsman A are scattered or more spread out. Thus, the measures of central tendency are not sufficient to give complete information about a given data. In such a situation the comparison becomes very difficult. We therefore, need some additional information for comparison, concerning with, how the data is dispersed about (more spread out) the average. This can be done by measuring the dispersion. Like 'measures of central tendency' we want to have a single number to describe variability. This single number is called a **'measure of dispersion'**.

Dispersion



No Dispersion

"The variability (spread) that exists between the values of a data is called dispersion".

OR

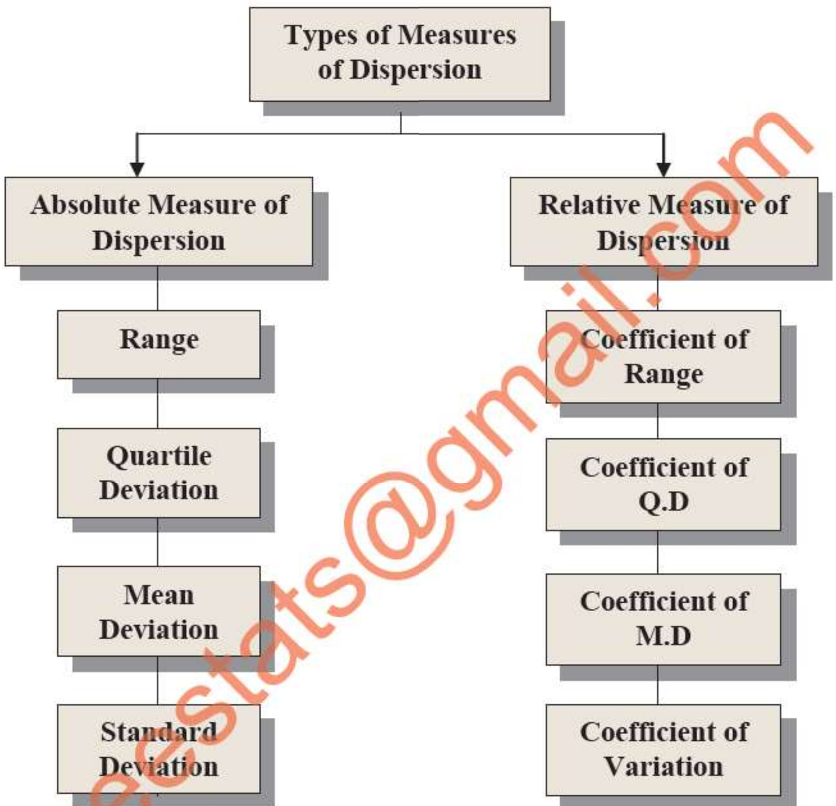
"The extent to which the observations are spread around an average is called dispersion or the scatter".

Dispersion

zeestats@gmail.com

Measures of Dispersion

As we know that, there are quite a few ways of measuring the central tendency of a data set i.e. A.M, G.M, H.M, Mode and Median. Similarly, we have different ways of measuring and comparing the dispersion of the distribution(s). There are two important types of measures of dispersion.



zeestats@gmail.com

Absolute Measure of Dispersion

"An absolute measure of dispersion measures the variability in terms of the same units of the data" e.g. if the units of the data are Rs, meters, kg, etc. The units of the measures of dispersion will also be Rs, meters, kg, etc.

The common absolute measures of dispersion are:

- Range
- Quartile Deviation or Semi Inter-Quartile Range
- Average Deviation or Mean Deviation
- Standard Deviation

Relative Measure of Dispersion

"A relative measure of dispersion compares the variability of two or more data that are independent of the units of measurements"

In other word *"A relative measure of dispersion, expresses the absolute measure of dispersion relative to the relevant average and multiplied by 100 many times"* *i.e.*

$$Relative\,Dispersion = \frac{Absolute\ Dispersion}{Average}$$

$$Relative\,Dispersion = \frac{Absolute\ Dispersion}{Average} \times 100$$

This is a pure number and independent of the units in which the data has been expressed. It is used for the purpose to compare the dispersion of a data with the dispersion of another data.

The common relative measures of dispersion are:

- Coefficient of Dispersion or Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation
- Coefficient of Standard Deviation or Coefficient of Variation (C.V)

VS.

The major difference b/w Absolute and Relative Measures of Dispersion is that the Absolute measure of dispersion measures only the variability of the data, further it has the unit of measurement; on the other hand Relative measure of dispersion is used to compare the variation of two or more distributions, further it is unit less.

Range

- **Ungrouped Data and for Discrete Grouped Data**

"The difference between the largest and the smallest value in a set of data is called range" *i.e.*

$$R = X_m - X_0$$

Where R is the range, X_m is the largest value and X_0 is the smallest value.

zeestats@gmail.com

- **Continuous Grouped Data**

"In continuous grouped data the difference between the upper class boundary of the highest class and lower class boundary of the lowest class is called range".

Coefficient of Range or Coefficient of Dispersion

The coefficient of range or coefficient of dispersion is a relative measure of dispersion and is given by:

$$Coefficient\ of\ Range = \frac{X_m - X_0}{X_m + X_0}$$

Historical Note

In 1892, Pearson introduced statistical concept of "range"

zeestats@gmail.com

EXAMPLE 4.01

Find Range and the Coefficient of Range from the following data:

51, 50, 40, 90, 75, 60, 44, 30, 23, 20 **(ungrouped data)**

Solution

Here $X_m = 90; X_0 = 20$

$$R = X_m - X_0 = 90 - 20 = 70$$

$$Coefficient\ of\ Range = \frac{X_m - X_0}{X_m + X_0} = \frac{90 - 20}{90 + 20} = 0.64$$

EXAMPLE 4.02

Find Range and the Coefficient of Range from the following data: **(Discrete Grouped data)**

Marks (X)	13	14	15	16	17
No. of Students (f)	2	5	13	7	3

Solution

Here $X_m = 17; X_0 = 13$

$$R = X_m - X_0 = 17 - 13 = 4$$

$$Coefficient\ of\ Range = \frac{X_m - X_0}{X_m + X_0} = \frac{17 - 13}{17 + 13} = 0.13$$

EXAMPLE 4.03

Find Range and the Coefficient of Range from the following data: **(Continuous Grouped data)**

Weight	11- 20	21- 30	31- 40	41-50	51-60
f	1	2	3	2	1

Solution

Weight	f	Class Boundaries
11- 20	1	10.5- 20.5
21- 30	2	20.5- 30.5
31- 40	3	30.5- 40.5
41-50	2	40.5-50.5
51-60	1	50.5-60.5
Total	9	--

Here $X_m = 60.5;\ X_0 = 10.5$

$$R = X_m - X_0 = 60.5 - 10.5 = 50$$

$$Coefficient\ of\ Range = \frac{X_m - X_0}{X_m + X_0}$$

$$= \frac{60.5 - 10.5}{60.5 + 10.5} = 0.70$$

Merits and Demerits of Range

Merits

- It is the simplest measure of dispersion.
- It gives a quick picture of the variability.

Demerits

- It does not based on each and every value of the data.
- It cannot be computed in case of open-end distributions
- It is affected by extreme values.
- It is not capable of further algebraic treatment.
- It is affected by fluctuations of sampling.
- It is unsatisfactory for statistical inference.

zeestats@gmail.com

Test Yourself

Find the Range and Coefficient of Range from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

Quartile Deviation or Semi-inter-quartile Range

"Half of the difference between the upper quartile and lower quartile is called the semi-inter quartile range or quartile deviation" i.e.

$$Quartile\ deviation = \frac{Q_3 - Q_1}{2}$$

REMEMBER

The difference between the upper quartile and lower quartile is called inter quartile range i.e.

Inter quartile range = $Q_3 - Q_1$

Coefficient of Quartile Deviation

The coefficient of quartile deviation is a relative measure of dispersion and is given by:

$$Coefficient\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

EXAMPLE 4.04

Find Q.D and the Coefficient of Q.D from the following data:

50,51,52,53,54,55,56,57,58,59,60; (n = 11)

zeestats@gmail.com

Solution

$$Q_1 = \text{size of}\left(\frac{1(n+1)}{4}\right)\text{th observation}$$

$$Q_1 = \text{size of}\left(\frac{1(11+1)}{4}\right)\text{th observation}$$

$$= \text{size of 3th observation} = 52$$

$$Q_3 = \text{size of}\left(\frac{3(n+1)}{4}\right)\text{th observation}$$

$$Q_3 = \text{size of}\left(\frac{3(11+1)}{4}\right)\text{th observation}$$

$$= \text{size of 9th observation} = 58$$

Here $Q_1 = 52;\ Q_3 = 58$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{58 - 52}{2} = 3$$

$$\text{Coefficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{58 - 52}{58 + 52} = 0.0273$$

zeestats@gmail.com

EXAMPLE 4.05

Find Q.D and the Coefficient of Q.D from the following data

20, 21, 22, 23, 24, 25, 26, 27; **(ungrouped data)** (n = 8)

Solution

$$Q_1 = \text{size of}\left(\frac{1(n+1)}{4}\right)\text{th observation}$$

$$Q_1 = \text{size of}\left(\frac{1(8+1)}{4}\right)\text{th observation}$$

$$= \text{size of 2.25th observation}$$

$$= \text{size of}\left[\text{2nd} + 0.25(\text{3rd - 2nd})\right]\text{observation}$$

$$= 21 + 0.25(22 - 21) = 21.25$$

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

$$Q_3 = size\ of \left(\frac{3(8+1)}{4}\right) th\ observation$$

$$= size\ of\ 6.75th\ observation$$

$$= size\ of \left[6th + 0.75(7th - 6th)\right] observation$$

$$= 25 + 0.75(26 - 25) = 25.75$$

Here $Q_1 = 21.25;\ Q_3 = 25.75$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{25.75 - 21.25}{2} = 2.25$$

$$Coefficient\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{25.75 - 21.25}{25.75 + 21.25} = 0.0957$$

zeestats@gmail.com

EXAMPLE 4.06

Find Q.D and the Coefficient of Q.D from the following data: **(Discrete grouped data)**

X	20	21	22	23	24	25
f	1	3	5	2	2	2

Solution

X	f	Cumulative Frequency
20	1	1
21 ←	3	4
22	5	9
23	2	11
24 ←	2	13
25	2	15
Total	15	--

$$Q_1 = size\ of \left(\frac{1(n+1)}{4}\right) th\ observation$$

$$\Rightarrow Q_1 = size\ of \left(\frac{1(15+1)}{4}\right) th\ observation$$

$$= size\ of\ 4th\ observation = 21$$

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

$$\Rightarrow Q_3 = size\ of \left(\frac{3(15+1)}{4}\right) th\ observation$$

$$= size\ of\ 12th\ observation = 24$$

Here $Q_1 = 21;\ Q_3 = 24$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{24 - 21}{2} = 1.5$$

$$Coefficient\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{24 - 21}{24 + 21} = 0.0667$$

EXAMPLE 4.07

Find Q.D and the Coefficient of Q.D from the following data: **(Continuous grouped data)**

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
f	8	87	190	304	211	85	20

Solution

Marks	No. of Students	C.F	Class boundaries
30-39	8	8	29.5-39.5
40-49	87	95	39.5-49.5
50-59	190	285 →	49.5-59.5
60-69	304	589	59.5-69.5
70-79	211	800 →	69.5-79.5
80-89	85	885	79.5-89.5
90-99	20	905	89.5-99.5
Total	905	--	--

zeestats@gmail.com

Step 1:

$$Q_1 = Size\ of \left(\frac{1 \times n}{4}\right) th\ observation$$
$$= Size\ of \left(\frac{905}{4}\right) th\ observation$$
$$= Size\ of\ 226.25th\ observation$$

And since 226.25th observation lies in the class (49.5-59.5); hence this is the lower quartile class.

Here $l = 49.5, f = 190, C = 95, h = 10$

Step 2:

$$Q_1 = l + \frac{h}{f}\left(\frac{1 \times n}{4} - C\right)$$
$$= 49.5 + \frac{10}{190}(226.25 - 95)$$
$$= 56.40$$

Step 1:

$$Q_3 = Size\ of \left(\frac{3 \times n}{4}\right) th\ observation$$
$$= Size\ of \left(\frac{3 \times 905}{4}\right) th\ observation$$
$$= Size\ of\ 678.75th\ observation$$

And since 678.75th observation lies in the class (59.5-69.5); hence this is the median quartile class.

Here $l = 69.5, f = 211, C = 589, h = 10$

Step 2:

$$Q_3 = l + \frac{h}{f}\left(\frac{3 \times 905}{4} - C\right)$$
$$= 69.5 + \frac{10}{211}(678.75 - 589)$$
$$= 73.75$$

Here $Q_1 = 56.40; Q_3 = 73.75$

$$Q.D = \frac{Q_3 - Q_1}{2} = \frac{73.75 - 56.40}{2} = 8.6750$$

$$Coefficient\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{73.75 - 56.40}{73.75 + 56.40} = 0.1333$$

Merits and Demerits of Quartile Deviation

Merits

- It is simple to understand and easy to calculate.
- It is a good measure for open-end distributions.

Demerits

- It does not based on each and every value of the data.
- It is not capable of further algebraic treatment.
- It is affected by fluctuations of sampling.
- It is unsatisfactory for statistical inference.

zeestats@gmail.com

Test Yourself

Find the Q.D and Coefficient of Q.D from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15, 20, 19, 21

2) 30, 33, 23, 22, 34, 40, 41, 28, 35, 39

3)

X	20	25	30	35	40
f	2	4	9	3	1

4)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

Mean Absolute Deviation or Mean Deviation (Average Deviation)

"The arithmetic mean of the absolute deviations from an average (mean, median, etc.) is called mean deviation or average deviation"

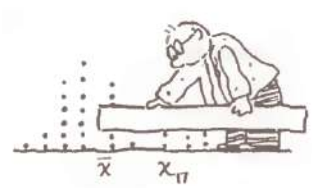
	Ungrouped Data	Grouped Data
M.D from Mean	$M.D = \frac{\sum \lvert x_i - \bar{x} \rvert}{n}$	$M.D = \frac{\sum f \lvert x_i - \bar{x} \rvert}{n}$
M.D from Median	$M.D = \frac{\sum \lvert x_i - Med \rvert}{n}$	$M.D = \frac{\sum f \lvert x_i - Med \rvert}{n}$

Coefficient of Mean Deviation

The coefficient of mean deviation is a relative measure of dispersion and is given by:

$$Coefficient\ of\ M.D\ (from\ mean) = \frac{M.D(from\ mean)}{Mean}$$

$$Coefficient\ of\ M.D\ (from\ median) = \frac{M.D(from\ median)}{Median}$$



EXAMPLE 4.08

Find M.D and the Coefficient of M.D from mean.

Using the data: 50,51,52,53,54,55,56,57,58,59,60; **(ungrouped data)**

Solution

Here $\bar{x} = \frac{\sum x_i}{n} = \frac{605}{11} = 55$

$$M.D = \frac{\sum \lvert x_i - \bar{x} \rvert}{n} = \frac{30}{11} = 2.7273$$

$$Coefficient\ of\ M.D = \frac{M.D}{\bar{X}} = \frac{2.7273}{55} = 0.0496$$

X	$Xi - \bar{X}$	$\lvert Xi - \bar{X} \rvert$
50	-5	5
51	-4	4
52	-3	3
53	-2	2
54	-1	1
55	0	0
56	1	1
57	2	2
58	3	3
59	4	4
60	5	5
605	--	30

zeestats@gmail.com

EXAMPLE 4.09

Find M.D and the Coefficient of M.D from median.

Using the data: 50,51,52,53,54,55,56,57,58,59,61; **(ungrouped data)**

Solution

$$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$$
$$= size\ of \left(\frac{11+1}{2}\right) th\ observation$$
$$= size\ of\ 6th\ observation = 55$$

X	Xi - Med	|Xi - Med|
50	-5	5
51	-4	4
52	-3	3
53	-2	2
54	-1	1
55	0	0
56	1	1
57	2	2
58	3	3
59	4	4
61	5	6
--	--	31

$$M.D = \frac{\sum |x_i - median|}{n} = \frac{31}{11} = 2.8182$$
$$Coefficient\ of\ M.D = \frac{M.D}{Median}$$
$$= \frac{2.8182}{55}$$
$$= 0.0512$$

zeestats@gmail.com

EXAMPLE 4.10

Find M.D and the Coefficient of M.D from mean. **(Discrete grouped data)**

X	20	21	22	23	24	25
f	1	3	5	2	2	2

Solution

Here $\bar{x} = \frac{\sum fx_i}{n} = \frac{337}{15} = 22.47$

$$M.D = \frac{\sum f|x_i - \bar{x}|}{n} = \frac{18.41}{15} = 1.23$$

$$Coefficient\ of\ M.D = \frac{M.D}{\bar{X}}$$
$$= \frac{1.23}{22.47}$$
$$= 0.05$$

X	f	fX	$\lvert Xi - \bar{X}\rvert$	$f\lvert Xi - \bar{X}\rvert$
20	1	20	2.47	2.47
21	3	63	1.47	4.41
22	5	110	0.47	2.35
23	2	46	0.53	1.06
24	2	48	1.53	3.06
25	2	50	2.53	5.06
Total	15	337	--	18.41

EXAMPLE 4.11

Find M.D and the Coefficient of M.D from median. **(Discrete grouped data)**

X	20	21	22	23	24	25
f	1	3	5	2	2	2

Solution

$$Median = size\ of \left(\frac{n+1}{2}\right) th\ observation$$

$$= size\ of \left(\frac{15+1}{2}\right) th\ observation$$

$$= size\ of\ 8th\ observation = 22$$

X	f	C.F	$f\lvert Xi - Med\rvert$
20	1	1	2
21	3	4	3
22 ←	5	9	0
23	2	11	2
24	2	13	4
25	2	15	6
Total	15	--	17

$$M.D = \frac{\sum f\lvert x_i - median\rvert}{n} = \frac{17}{15} = 1.13$$

$$Coefficient\ of\ M.D = \frac{M.D}{Median} = \frac{1.13}{22}$$

$$= 0.05$$

EXAMPLE 4.12

Find M.D and the Coefficient of M.D: from mean **(Continuous grouped data)**

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Here $\bar{x} = \frac{\sum fx_i}{n} = \frac{58902.5}{905} = 65.09$

$$M.D = \frac{\sum f\lvert x_i - \bar{x}\rvert}{n} = \frac{8449.88}{905} = 9.34$$

$$Coefficient\ of\ M.D = \frac{M.D}{\bar{X}}$$

$$= \frac{9.34}{65.09}$$

$$= 0.14$$

Marks	X	f	fX	$f\lvert Xi - \bar{X}\rvert$
30-39	34.5	8	276	244.69
40-49	44.5	87	3871.5	1790.95
50-59	54.5	190	10355	2011.27
60-69	64.5	304	19608	178.03
70-79	74.5	211	15719.5	1986.43
80-89	84.5	85	7182.5	1650.22
90-99	94.5	20	1890	588.29
Total	--	905	58902.5	8449.88

zeestats@gmail.com

EXAMPLE 4.13

Find M.D and the Coefficient of M.D from median. **(Continuous grouped data)**

Marks	30-39	40-49	50-59	60-69	70-79	80-89	90-99
No. of Students	8	87	190	304	211	85	20

Solution

Step 1:

$$Median = Size\ of\left(\frac{n}{2}\right)th\ observation$$

$$= Size\ of\left(\frac{905}{2}\right)th\ observation$$

$$= Size\ of\ 452.5th\ observation$$

And since 452.5th observation lies in the class (59.5-69.5); hence this is the median class.

Here $l = 59.5, f = 304, C = 285, h = 10$

Step 2:

$$Median = l + \frac{h}{f}\left(\frac{n}{2} - C\right)$$

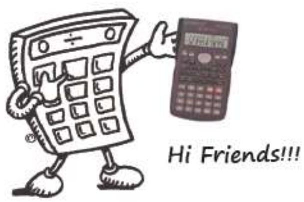
$$= 59.5 + \frac{10}{304}(452.5 - 285)$$

$$= 65$$

Marks	*X*	*No. of Students (f)*	*C.F*	*Class boundaries*	$f\lvert Xi - Median\rvert$
30-39	34.5	8	8	29.5-39.5	244
40-49	44.5	87	95	39.5-49.5	1783.5
50-59	54.5	190	285	49.5-59.5	1995
60-69	64.5	304	589 →	59.5-69.5	152
70-79	74.5	211	800	69.5-79.5	2004.5
80-89	84.5	85	885	79.5-89.5	1657.5
90-99	94.5	20	905	89.5-99.5	590
Total	--	905	--	--	8426.5

$$M.D = \frac{\sum f\lvert x_i - median\rvert}{n} = \frac{8426.5}{905} = 9.31$$

$$Coefficient\ of\ M.D = \frac{M.D}{Median} = \frac{9.31}{65} = 0.14$$



zeestats@gmail.com

Merits and Demerits of Mean Deviation

Merits

- It is simple to understand.
- It is based on each and every value of the data.

Demerits

- It is not a good measure for open-end distributions.
- It is not capable of further algebraic treatment.
- It is difficult to handle it mathematically; because there is an element of artificiality i.e. the deviations are not taken with their proper signs.
- It is unsatisfactory for statistical inference.

Test Yourself

a) Find the M.D from **Mean** and Coefficient of M.D from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15, 20, 19, 21

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

b) Find the M.D from **Median** and Coefficient of M.D from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15, 20, 19, 21

2) 30, 33, 23, 22, 34, 40, 41, 28, 35, 39

3)

X	20	25	30	35	40
f	2	4	9	3	1

4)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

zeestats@gmail.com

Standard Deviation

"The positive square root of variance is called as Standard deviation".

OR

"The positive square root of the arithmetic mean of the squared deviations from the mean is called the standard deviation"

REMEMBER

- The arithmetic mean of the squared deviations of the values measured from the mean is called variance.
- For a set of data: Range > S.D > M.D > Q.D
- The measures of dispersion are always positive.

	Ungrouped Data	Grouped Data
S.D for Population	$\sigma = \sqrt{\frac{\sum (xi-\mu)^2}{N}}$	$\sigma = \sqrt{\frac{\sum f(xi-\mu)^2}{N}}$
S.D for Sample	$S = \sqrt{\frac{\sum (xi-\bar{x})^2}{n}}$	$S = \sqrt{\frac{\sum f(xi-\bar{x})^2}{n}}$

Methods of Calculating Variance and Standard Deviation

Methods	Ungrouped Data	
	Variance	Standard Deviation
Direct Method	$S^2 = \frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2$	$S = \sqrt{\frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2}$
Short cut Method	$S^2 = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2$	$S = \sqrt{\frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2}$
Step-deviation Method	$S^2 = h^2\left[\frac{\sum ui^2}{n} - \left(\frac{\sum ui}{n}\right)^2\right]$	$S = h\sqrt{\frac{\sum ui^2}{n} - \left(\frac{\sum ui}{n}\right)^2}$
Methods	**Grouped Data**	
	Variance	**Standard Deviation**
Direct Method	$S^2 = \frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2$	$S = \sqrt{\frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2}$
Short cut Method	$S^2 = \frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2$	$S = \sqrt{\frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2}$
Step-deviation Method	$S^2 = h^2\left[\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2\right]$	$S = h\sqrt{\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2}$

zeestats@gmail.com

Coefficient of Standard Deviation OR Coefficient of Variation

The coefficient of standard deviation is a relative measure of dispersion and is given by:

$$Coefficient\ of\ S.D = \frac{Standard\ Deviation}{Mean}$$

The coefficient of standard deviation is also called the coefficient of variation, denoted by **C.V** and is given by:

$$C.V = \frac{Standard\ Deviation}{Mean} \times 100$$

Coefficient of Variation was introduced by **Karl Pearson**. It is used to compare the variation or to compare the performance of two sets of data. A large value of C.V indicates that there is greater variability and vice versa. Similarly, the smaller the C.V the more consistent is the performance and vice versa.

Historical Note

In 1893, Pearson introduced statistical concept of "S.D". He also introduced the coefficient of variation for the comparison between two different groups.

zeestats@gmail.com

EXAMPLE 4.14

Find Variance and Standard deviation from the following data: **(ungrouped data)**

2, 4, 6, 8, 10

Solution

Direct Method

X	X^2
2	4
4	16
6	36
8	64
10	100
30	220

$$Variance = S^2 = \frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2 = \frac{220}{5} - \left(\frac{30}{5}\right)^2 = 8$$

$$Standard\ Deviation = S = \sqrt{\frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2} = \sqrt{\frac{220}{5} - \left(\frac{30}{5}\right)^2} = 2.8$$

Short-cut Method Let $A = 4$

X	$D = X_i - A$	D^2
2	-2	4
4	0	0
6	2	4
8	4	16
10	6	36
--	10	60

$$S^2 = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2 = \frac{60}{5} - \left(\frac{10}{5}\right)^2 = 8$$

$$S = \sqrt{\frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2} = \sqrt{\frac{60}{5} - \left(\frac{10}{5}\right)^2} = 2.8$$

Step-deviation Method Here $h = 2$ and let $A = 8$

X	$u = \frac{X_i - A}{h}$	u^2
2	-3	9
4	-2	4
6	-1	1
8	0	0
10	1	1
30	-5	15

$$S^2 = h^2\left[\frac{\sum ui^2}{n} - \left(\frac{\sum ui}{n}\right)^2\right] = 2^2\left[\frac{15}{5} - \left(\frac{-5}{5}\right)^2\right] = 8$$

$$S = h\sqrt{\frac{\sum ui^2}{n} - \left(\frac{\sum ui}{n}\right)^2} = 2\sqrt{\frac{15}{5} - \left(\frac{-5}{5}\right)^2} = 2.8$$

zeestats@gmail.com

EXAMPLE 4.15

Find Variance and Standard deviation from the following data: **(Discrete grouped data)**

X	10	15	20	25	30
f	1	2	3	2	1

Solution

Direct Method

X	f	fX	fX^2
10	1	10	100
15	2	30	450
20	3	60	1200
25	2	50	1250
30	1	30	900
Total	9	180	3900

$$S^2 = \frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2 = \frac{3900}{9} - \left(\frac{180}{9}\right)^2 = 33.3$$

$$S = \sqrt{\frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2} = \sqrt{\frac{3900}{9} - \left(\frac{180}{9}\right)^2} = 5.7$$

Short-cut Method Here $A = 20$

X	f	$D = X_i - A$	fD	fD^2
10	1	-10	-10	100
15	2	-5	-10	50
20	3	0	0	0
25	2	5	10	50
30	1	10	10	100
Total	9	--	0	300

$$S^2 = \frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2 = \frac{300}{9} - \left(\frac{0}{9}\right)^2 = 33.3$$

$$S = \sqrt{\frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2} = \sqrt{\frac{300}{9} - \left(\frac{0}{9}\right)^2} = 5.7$$

Step-deviation Method Here $A = 20$, $h = 5$

X	f	$u = \frac{X_i - A}{h}$	fu	fu^2
10	1	-2	-2	4
15	2	-1	-2	2
20	3	0	0	0
25	2	1	2	2
30	1	2	2	4
Total	9	--	0	12

$$S^2 = h^2\left[\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2\right] = 5^2\left[\frac{12}{9} - \left(\frac{0}{9}\right)^2\right] = 33.3$$

$$S = h\sqrt{\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2} = 5\sqrt{\frac{12}{9} - \left(\frac{0}{9}\right)^2} = 5.7$$

zeestats@gmail.com

EXAMPLE 4.16

Find Variance and Standard deviation from the following data: **(Continuous grouped data)**

Weight(kg)	11- 20	21- 30	31- 40	41-50	51-60
f	1	2	3	2	1

Solution

Direct Method

Weight(kg)	f	X	fX	fX^2
11- 20	1	15.5	15.5	240.25
21- 30	2	25.5	51.0	1300.5
31- 40	3	35.5	106.5	3780.75
41-50	2	45.5	91.0	4140.5
51-60	1	55.5	55.5	3080.25
Total	9	--	319.5	12542.25

$$S^2 = \frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2$$

$$= \frac{12542.25}{9} - \left(\frac{319.5}{9}\right)^2 = 133.3kg^2$$

$$S = \sqrt{\frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2}$$

$$= \sqrt{\frac{12542.25}{9} - \left(\frac{319.5}{9}\right)^2} = 11.5kg$$

Short-cut Method Here $A = 35.5$

Weight	f	X	$D = X_i - A$	fD	fD^2
11- 20	1	15.5	-20	-20	400
21- 30	2	25.5	-10	-20	200
31- 40	3	35.5	0	0	0
41-50	2	45.5	10	20	200
51-60	1	55.5	20	20	400
Total	9	--	--	0	1200

$$S^2 = \frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2 = \frac{1200}{9} - \left(\frac{0}{9}\right)^2 = 133.3kg^2$$

$$S = \sqrt{\frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2} = \sqrt{\frac{1200}{9} - \left(\frac{0}{9}\right)^2} = 11.5kg$$

- To compute the Variance or S.D, round-off it one more decimal place than the original data values.
- The unit of the S.D is the same as that for the raw data, so it is preferable to use the S.D instead of the Variance.

Step-deviation Method Here $A = 35.5$, $h = 10$

Weight	f	X	$u = \frac{X_i - A}{h}$	fu	fu^2
11- 20	1	15.5	-2	-2	4
21- 30	2	25.5	-1	-2	2
31- 40	3	35.5	0	0	0
41-50	2	45.5	1	2	2
51-60	1	55.5	2	2	4
Total	9	--	--	0	12

$$S^2 = h^2\left[\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2\right]$$

$$= 10^2\left[\frac{12}{9} - \left(\frac{0}{9}\right)^2\right] = 133.3kg^2$$

$$S = h\sqrt{\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2}$$

$$= 10\sqrt{\frac{12}{9} - \left(\frac{0}{9}\right)^2} = 11.5kg$$

zeestats@gmail.com

Test Yourself

Find the Variance and S.D from the following data:

1) 1, 3, 5, 7, 9, 11, 13, 15, 20, 19, 21

2)

X	20	25	30	35	40
f	2	4	9	3	1

3)

Weight	21- 30	31- 40	41- 50	51-60	61-70
f	1	3	5	4	2

It will be incorrect if we get a negative answer in calculating measures of dispersion.

EXAMPLE 4.17

The number of runs scored by two cricketers A and B during a test series of 5 test matches is shown below for each of the 10 innings. Using coefficient of variation, find who will be more consistent player?

A	5	26	97	76	112	89	6	108	24	16
B	51	47	36	60	58	39	44	42	71	50

Solution

Cricketer A:	Cricketer B:
$C.V(x) = \frac{S.D(x)}{\overline{x}} \times 100$ $\overline{x} = \frac{\sum xi}{n}$ $S.D(x) = \sqrt{\frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2}$	$C.V(y) = \frac{S.D(y)}{\overline{y}} \times 100$ $\overline{y} = \frac{\sum yi}{n}$ $S.D(y) = \sqrt{\frac{\sum yi^2}{n} - \left(\frac{\sum yi}{n}\right)^2}$

x	x^2	y	y^2
5	25	51	2601
26	676	47	2209
97	9409	36	1296
76	5776	60	3600
112	12544	58	3364
89	7921	39	1521
6	36	44	1936
108	11664	42	1764
24	576	71	5041
16	256	50	2500
559	48883	498	25832

Cricketer A:	Cricketer B:
$\overline{x} = \frac{559}{10} = 55.9$ $S.D(x) = \sqrt{\frac{48883}{10} - \left(\frac{559}{10}\right)^2} = 41.993$ $C.V(x) = \frac{41.993}{55.9} \times 100 = 75.12\%$	$\overline{y} = \frac{498}{10} = 49.8$ $S.D(y) = \sqrt{\frac{25832}{10} - \left(\frac{498}{10}\right)^2} = 10.156$ $C.V(y) = \frac{10.156}{49.8} \times 100 = 20.39\%$

Since the C.V for player B is smaller than C.V for player A, therefore player B is more consistent.

zeestats@gmail.com

EXAMPLE 4.18

Goals scored by two teams A and B in a football season were as follows:

No. of goals scored in a match (xi)	Number of Matches (frequencies)	
	A	B
0	27	17
1	9	9
2	8	6
3	5	5
4	4	3

Using coefficient of variation, find which team may be considered more consistent?

Solution

Team A:	Team B:
$C.V = \frac{S.D}{\overline{x}} \times 100$ $\overline{x} = \frac{\sum f_A xi}{n_A}$ $S.D = \sqrt{\frac{\sum f_A xi^2}{n_A} - \left(\frac{\sum f_A xi}{n_A}\right)^2}$	$C.V = \frac{S.D}{\overline{x}} \times 100$ $\overline{x} = \frac{\sum f_B xi}{n_B}$ $S.D = \sqrt{\frac{\sum f_B xi^2}{n_B} - \left(\frac{\sum f_B xi}{n_B}\right)^2}$

x_i	f_A	f_B	$f_A x_i$	$f_A x^2$	$f_B x_i$	$f_B x^2$
0	27	17	0	0	0	0
1	9	9	9	9	9	9
2	8	6	16	32	12	24
3	5	5	15	45	15	45
4	4	3	16	64	12	48
Total	53	40	56	150	48	126

Team A:	Team B:
$\overline{x} = \frac{56}{53} = 1.06$ $S.D = \sqrt{\frac{150}{53} - \left(\frac{56}{53}\right)^2} = 1.308$ $C.V = \frac{1.308}{1.06} \times 100 = 123.4\%$	$\overline{x} = \frac{48}{40} = 1.20$ $S.D = \sqrt{\frac{126}{40} - \left(\frac{48}{40}\right)^2} = 1.308$ $C.V = \frac{1.308}{1.20} \times 100 = 109.0\%$

Since the C.V for Team B is smaller than C.V for Team A, therefore team B is more consistent.

zeestats@gmail.com

Test Yourself

1) The number of runs scored by two cricketers A and B during a test series of 5 test matches is shown below for each of the 10 innings. Using coefficient of variation, find who will be more consistent player?

A	15	34	27	55	0	0	6	4	123	34
B	5	67	36	55	89	33	37	89	88	111

2) Goals scored by two teams A and B in a football season were as follows:

No. of goals scored in a match (xi)	Number of Matches (frequencies)	
	A	B
0	16	20
1	8	7
2	4	8
3	6	2
4	3	1

Using coefficient of variation, find which team may be considered more consistent?

zeestats@gmail.com

Merits and Demerits of Standard Deviation

Merits

- It is simple to understand.
- It is clearly defined by a mathematical formula.
- It is based on each and every value of the data.
- It is capable of further algebraic treatment.
- It is less affected by the fluctuations of sampling.
- It provide basis for statistical inference.

Demerits

- Its calculation is not very simple.
- It is affected by the extreme values.
- It is not a good measure for open-end distributions.

Properties of Variance and Standard Deviation

- **Variance and S.D of a constant is zero i.e.**

  $Var(c) = 0$ **and** $S.D(c) = 0$ **where "c" is any constant.**

- **The variance and S.D are unaffected by the change of origin i.e. when a constant is added to or subtracted from each value of a variable, the variance and S.D remain unchanged i.e.**

  $Var(X \pm c) = Var(X)$ **and** $S.D(X \pm c) = S.D(X)$

- **Variance and S.D are affected by the change of scale i.e. when each observation of a variable is multiplied or divided by a constant, then variance and S.D are affected by these changes i.e.**

  $Var(c\,X) = c^2 Var(X)$ **and** $S.D(c\,X) = |c|\,S.D(X)$

  $Var\left(\frac{X}{c}\right) = \left(\frac{1}{c^2}\right) Var(X)$ **and** $S.D\left(\frac{X}{c}\right) = \left|\frac{1}{c}\right| S.D(X)$

- **The variance and S.D of the sum or difference of two independent variables is equal to the sum of their respective variances and S.D's respectively i.e.**

  $Var(X \pm Y) = Var(X) + Var(Y)$ **and** $S.D(X \pm Y) = S.D(X) + S.D(Y)$

zeestats@gmail.com

Moments

"The arithmetic mean of the r^{th} power of deviations taken either from mean, zero or from any arbitrary origin (provisional mean) are called moments".

- When the deviations are computed from the arithmetic mean, then such moments are called moments about mean (mean moments) or sometimes called **central moments**, denoted by m_r and given as follows:

Ungrouped Data	Grouped Data
$m_r = \frac{\sum (x_i - \bar{x})^r}{n}$	$m_r = \frac{\sum f(x_i - \bar{x})^r}{n}$
Where r = 1,2,3,4…	

- When the deviations of the values are computed from origin or zero, then such moments are called the **moments about origin**, denoted by m'_r and are given by:

Ungrouped Data	Grouped Data
$m'_r = \frac{\sum x_i^r}{n}$	$m'_r = \frac{\sum f x_i^r}{n}$
Where r = 1,2,3,4…	

REMEMBER

Moments about provisional mean and moments about zero are called raw moments (denoted by m'_r)

- When the deviations of the values are computed from any arbitrary value say "A" (provisional mean), then such moments are called **moments about provisional mean**, denoted by m'_r.

Ungrouped Data	Grouped Data
$m'_r = \frac{\sum D^r}{n}$	$m'_r = \frac{\sum f D^r}{n}$
Where r = 1,2,3,4… D = xi –A	

zeestats@gmail.com

EXAMPLE 4.19

Calculate the first four moments about the mean from the following data.

2, 4, 6, 8, 10

Solution

Here $\bar{x} = \frac{\sum x_i}{n} = \frac{30}{5} = 6$

xi	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
2	-4	16	-64	256
4	-2	4	-8	16
6	0	0	0	0
8	2	4	8	16
10	4	16	64	256
30	0	40	0	544

$m_1 = \frac{\sum (x_i - \bar{x})}{n} = 0$, $m_2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{40}{5} = 8$

$m_3 = \frac{\sum (x_i - \bar{x})^3}{n} = 0$, $m_4 = \frac{\sum (x_i - \bar{x})^4}{n} = \frac{544}{5} = 108.8$

EXAMPLE 4.20

Calculate the first four moments about the zero from the following data.

2, 4, 6, 8, 10

Solution

x_i	x_i^2	x_i^3	x_i^4
2	4	8	16
4	16	64	256
6	36	216	1296
8	64	512	4096
10	100	1000	10000
30	220	1800	15664

$$m'_1 = \frac{\sum x_i}{n} = \frac{30}{5} = 6 \quad , \quad m'_2 = \frac{\sum x_i^2}{n} = \frac{220}{5} = 44$$

$$m'_3 = \frac{\sum x_i^3}{n} = \frac{1800}{5} = 360 \quad , \quad m'_4 = \frac{\sum x_i^4}{n} = \frac{15664}{5} = 3132.8$$

EXAMPLE 4.21

Calculate the first four moments about the P.M from the following data.

2, 4, 6, 8, 10

Solution

Here D = xi – A and (let A = 4)

X	$D = X_i - A$	D^2	D^3	D^4
2	-2	4	-8	16
4	0	0	0	0
6	2	4	8	16
8	4	16	64	256
10	6	36	216	1296
30	10	60	280	1584

$$m'_1 = \frac{\Sigma D}{n} = \frac{10}{5} = 2 \quad , \quad m'_2 = \frac{\Sigma D^2}{n} = \frac{60}{5} = 12$$

$$m'_3 = \frac{\Sigma D^3}{n} = \frac{280}{5} = 56 \quad , \quad m'_4 = \frac{\Sigma D^4}{n} = \frac{1584}{5} = 316.8$$

zeestats@gmail.com

EXAMPLE 4.22

Calculate the first four moments about the mean from the following data:

xi	2	3	4	5	6
f	1	3	7	3	1

Solution

Here $\bar{x} = \frac{\sum fx_i}{n} = \frac{60}{15} = 4$

xi	f	fx	$(xi-\bar{x})$	$f(xi-\bar{x})$	$f(xi-\bar{x})^2$	$f(xi-\bar{x})^3$	$f(xi-\bar{x})^4$
2	1	2	-2	-2	4	-8	16
3	3	9	-1	-3	3	-3	3
4	7	28	0	0	0	0	0
5	3	15	1	3	3	3	3
6	1	6	2	2	4	8	16
tal	15	60	--	0	14	0	38

$$m_1 = \frac{\sum f(x_i-\bar{x})}{n} = \frac{0}{15} = 0 \quad , \quad m_2 = \frac{\sum f(x_i-\bar{x})^2}{n} = \frac{14}{15} = 0.933$$

$$m_3 = \frac{\sum f(x_i-\bar{x})^3}{n} = \frac{0}{15} = 0 \quad , \quad m_4 = \frac{\sum f(x_i-\bar{x})^4}{n} = \frac{38}{15} = 2.533$$

EXAMPLE 4.23

Calculate the first four moments about zero from the following data:

xi	2	3	4	5	6
f	1	3	7	3	1

Solution

xi	f	fx	fx^2	fx^3	fx^4
2	1	2	4	8	16
3	3	9	27	81	243
4	7	28	112	448	1792
5	3	15	75	375	1875
6	1	6	36	216	1296
Total	15	60	254	1128	5222

$$m'_1 = \frac{\sum fx_i}{n} = \frac{60}{15} = 4 \quad , \quad m'_2 = \frac{\sum fx_i^2}{n} = \frac{254}{15} = 16.93$$

$$m'_3 = \frac{\sum fx_i^3}{n} = \frac{1128}{15} = 75.2 \quad , \quad m'_4 = \frac{\sum fx_i^4}{n} = \frac{5222}{15} = 348.13$$

zeestats@gmail.com

EXAMPLE 4.24

Calculate the first four moments about P.M from the following data:

xi	2	3	4	5	6
f	1	3	7	3	1

Solution

Here D = xi – A and (let A = 3)

xi	*f*	*D = xi – A*	*fD*	fD^2	fD^3	fD^4
2	1	-1	-1	1	-1	1
3	3	0	0	0	0	0
4	7	1	7	7	7	7
5	3	2	6	12	24	48
6	1	3	3	9	27	81
Total	15	--	15	29	57	137

$$m'_1 = \frac{\sum fD}{n} = \frac{15}{15} = 1 \quad , \quad m'_2 = \frac{\sum fD^2}{n} = \frac{29}{15} = 1.933$$

$$m'_3 = \frac{\sum fD^3}{n} = \frac{57}{15} = 3.8 \quad , \quad m'_4 = \frac{\sum fD^4}{n} = \frac{137}{15} = 9.133$$

Test Yourself

Find Moments about **Mean**, about **Zero** and about **P.M** from the following data:

1) 11, 13, 15, 17, 19, 21, 23, 25, 30, 29, 31

2)

xi	20	30	40	50	60
f	1	5	9	4	1

TECHNICAL STUFF

All the raw moments can then be converted into central moments or mean moments or moments about mean, by using the following relations:

- $m_1 = 0$
- $m_2 = m'_2 - (m'_1)^2$
- $m_3 = m'_3 - 3m'_1 m'_2 + 2(m'_1)^3$
- $m_4 = m'_4 - 4m'_1 m'_3 + 6(m'_1)^2 m'_2 - 3(m'_1)^4$

zeestats@gmail.com

EXAMPLE 4.25

The first four moments about origin X = 0 are 4, 16.93, 75.2 and 348.13 respectively. Find moments about mean?

Solution Given that

$m'_1 = 4$, $m'_2 = 16.93$
$m'_3 = 75.2$, $m'_4 = 348.13$

Now we use:

$m_1 = 0$
$m_2 = m'_2 - (m'_1)^2$
$\Rightarrow m_2 = 16.93 - (4)^2 = 0.93$
$m_3 = m'_3 - 3m'_1m'_2 + 2(m'_1)^3$
$\Rightarrow m_3 = 75.2 - 3(4)(16.93) + 2(4)^3 = 0.04$
$m_4 = m'_4 - 4m'_1m'_3 + 6(m'_1)^2m'_2 - 3(m'_1)^4$
$\Rightarrow m_4 = 348.13 - 4(4)(75.2) + 6(4)^2(16.93) - 3(4)^4 = 2.21$

zeestats@gmail.com

EXAMPLE 4.26

The first four moments about X = 12 are 2.40, 43.0, 337.50 and 5500 respectively. Find moments about mean?

Solution Given that

$m'_1 = 2.40$, $m'_2 = 43.0$
$m'_3 = 337.50$, $m'_4 = 5500$

For population data we use " μ " instead of "m" in all the formulae of Moments.

Now we know that

$m_1 = 0$
$m_2 = m'_2 - (m'_1)^2$
$\Rightarrow m_2 = 43.0 - (2.4)^2 = 37.24$
$m_3 = m'_3 - 3m'_1m'_2 + 2(m'_1)^3$

μ (meu)

$\Rightarrow m_3 = 33.7 - 3(43)(2.40) + 2(2.40)^3 = 55.548$
$m_4 = m'_4 - 4m'_1m'_3 + 6(m'_1)^2m'_2 - 3(m'_1)^4$
$\Rightarrow m_4 = 5500 - 4(2.40)(337.5) + 6(2.40)^2(43) - 3(2.40)^4 = 3646.5472$

Test Yourself

1) The first four moments about origin X = 0 are 8, 83.71, 1019.43 and 123100 respectively. Find moments about mean?

2) The first four moments about X = 25 are -1.9, 20.5, -96.3 and 906.1 respectively. Find moments about mean?

Symmetrical Distribution

- A distribution in which the values of mean, median and mode are equal is called symmetrical distribution i.e.

Mean = Median = Mode

- A distribution is which the two quartiles are equidistant from the median is called a symmetrical distribution i.e.

$$Q_3 - \text{Median} = \text{Median} - Q_1$$

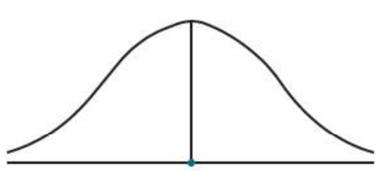
or $$Q_3 + Q_1 - 2\,\text{Median} = 0$$

- A distribution is said to be symmetrical if:

$$b_1 = 0$$

- A distribution in which the two tails are equal in length from the central value then it is called symmetrical distribution. The symmetrical distribution is always in the form of a bell.

Moment-Ratios	1st Moments Ratio	2nd Moments Ratio
Sample	$b_1 = \frac{m_3^2}{m_2^3}$	$b_2 = \frac{m_4}{m_2^2}$
Population	$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$	$\beta_2 = \frac{\mu_4}{\mu_2^2}$
Moment-Ratios are independent of the origin and units of measurements i.e. they are dimensionless quantities.		



zeestats@gmail.com

Skewness

We know that for symmetrical distribution the values of mean, median and mode are equal and that the two tails of the distribution are equal in length from the central value etc.

"*Skewness is the degree of asymmetry*"

OR

"*Skewness is the lack (absence) of symmetry around central value (average)*"

The presence skewness tells us that a particular distribution is not symmetrical or in other words it is skewed. In skewed distribution the curve is turned more to one side than the other.

Positive Skewness

- Skewness is said to be positive, if mean is greater than the median and median is greater than mode i.e.

  Mean > Median > Mode

- Skewness is said to be positive, if:

  $Q_3 + Q_1 - 2\text{ Median} > 0$

- In terms of moments, skewness is said to be positive if:

  $$\alpha_3 > 0$$

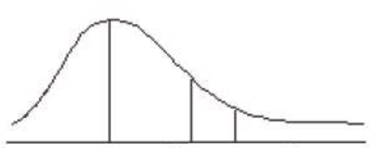
- Skewness is said to be positive, if the right tail of a distribution is longer than its left tail.

TIP

To measure the skewness we will use:

$$\alpha_3 = \sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}}$$

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



Mean > Median > Mode

zeestats@gmail.com

Negative Skewness

- Skewness is said to be negative, if mean is smaller than the median and median is smaller than mode i.e.

$$\text{Mean} < \text{Median} < \text{Mode}$$

- Skewness is said to be negative, if:

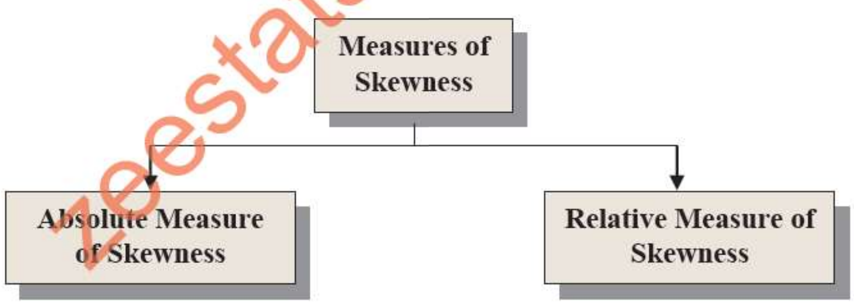
$$Q_3 + Q_1 - 2\ \text{Median} < 0$$

- In terms of moments, skewness is said to be negative if:

$$\alpha_3 < 0$$

- Skewness is said to be negative, if the left tail of a distribution is longer than its right tail.

Mean < Median < Mode



Absolute measures of skewness	Relative measures of skewness
• $A.S = Mean - Mode$ • $A.S = Mean - Median$ • $A.S = Q_3 + Q_1 - 2\ Median$	• Karl Pearson's measures of Skewness • Bowley's measures of Skewness • Coefficient of Skewness based on Moments

zeestats@gmail.com

EXAMPLE 4.27

Calculate absolute skewness if Mean = 13.25 and Median = 12.96.

Solution Given that

Mean = 13.25
Median = 12.96

To calculate absolute skewness we use the formula:

Absolute skewness = Mean – Median = 13.25 – 12.96 = 0.29

Hence the distribution is **positively skewed**.

EXAMPLE 4.28

Calculate absolute skewness if Mean = 12.61and Mode = 13.25.

Solution Given that

Mean = 12.61
Mode = 13.25

To calculate absolute skewness we use the formula:

Absolute skewness = Mean – Mode = 12.61– 13.25 = - 0.64

Hence the distribution is **negatively skewed**.

EXAMPLE 4.29

Calculate absolute skewness if $Q_1 = 13.73$, $Q_3 = 38.29$ and Median = 26.01

Solution To calculate absolute skewness we use the formula:

Absolute skewness = $Q_3 + Q_1 - 2Median = 38.29 + 13.73 - 2(26.01) = 0$

Hence the distribution is **symmetrical**.

zeestats@gmail.com

Test Yourself

1) Calculate absolute skewness if Mean = 33.25 and Median = 32.96.
2) Calculate absolute skewness if Mean = 42.61 and Mode = 43.25.
3) Calculate absolute skewness if $Q_1 = 124.87$, $Q_3 = 146.53$ and Median = 135.7

Karl Pearson's measures of Skewness

It is defined as:

$$S_k = \frac{Mean - Mode}{Standard\ Deviation}$$

"It is to be noted that, this measure is suggested by Karl Pearson (1857-1936) and is known as Pearsonian coefficient."

Since in many cases mode is ill-defined, therefore we replace (Mean – Mode) by its equivalent from the empirical relation i.e. 3 (Mean – Median) and hence:

$$S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$$

This coefficient usually varies between –3 and +3.

Historical Note

Karl Pearson introduced coefficient of skewness based on Mean, Median, Mode and S.D to measure the skewness of a distribution

zeestats@gmail.com

EXAMPLE 4.30

Calculate Pearson's coefficient of skewness if Mean = 13.25, Mode = 12.61 and S.D = 3.73

Solution Given that

Mean = 13.25, Mode = 12.61, S.D = 3.73

To calculate absolute skewness we use the formula:

$$S_k = \frac{Mean - Mode}{Standard\ Deviation} = \frac{13.25 - 12.61}{3.73} = 0.1716$$

Hence the distribution is **positively skewed**.

Bowley's measures of Skewness

It is defined as:

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

"It is also to be noted that, this measure is suggested by Bowley (1869-1957) and is known as Bowley's coefficient".

This coefficient usually varies between –1 and +1.

Historical Note

Bowley introduced coefficient of skewness based on quartiles to measure the skewness of a distribution

EXAMPLE 4.31

Calculate Bowley's coefficient of skewness if $Q_1 = 14.6$, $Q_3 = 25.2$ and Median = 18.8

Solution To calculate absolute skewness we use the formula:

$$S_k = \frac{Q_3 + Q_1 - 2Median}{Q_3 - Q_1} = \frac{25.2 + 14.6 - 2(18.8)}{25.2 - 14.6} = 0.21$$

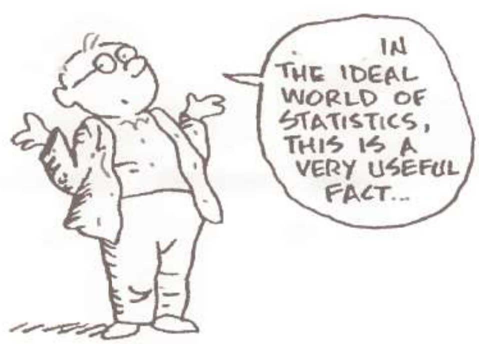
Hence the distribution is **positively skewed**.

zeestats@gmail.com

Coefficient of Skewness based on Moments

It is defined by:

$$\alpha_3 = \frac{m_3}{\sqrt{m_2^3}}$$



EXAMPLE 4.32

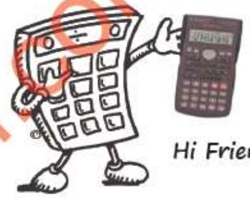
The first four moments about origin X = 0 are 4, 16.93, 75.2 and 348.13 respectively. Find moments about mean also find coefficient of skewness based on moments?

Solution Given that

$$m'_1 = 4 \quad , \quad m'_2 = 16.93$$
$$m'_3 = 75.2 \quad , \quad m'_4 = 348.13$$

Now we know that

$$m_1 = 0$$
$$m_2 = m'_2 - (m'_1)^2$$
$$\Rightarrow m_2 = 16.93 - (4)^2 = 0.93$$
$$m_3 = m'_3 - 3m'_1 m'_2 + 2(m'_1)^3$$
$$\Rightarrow m_3 = 75.2 - 3(4)(16.93) + 2(4)^3 = 0.04$$
$$m_4 = m'_4 - 4m'_1 m'_3 + 6(m'_1)^2 m'_2 - 3(m'_1)^4$$
$$\Rightarrow m_4 = 348.13 - 4(4)(75.2) + 6(4)^2(16.93) - 3(4)^4 = 2.21$$



Now $\alpha_3 = \dfrac{m_3}{\sqrt{m_2^3}} = \dfrac{0.04}{\sqrt{0.93^3}} = 0.0446$

Hence the distribution is **positively skewed**.

zeestats@gmail.com

Test Yourself

1) Calculate Pearson's coefficient of skewness if Mean = 50, Mode = 55 and S.D = 12.5
2) Calculate Bowley's coefficient of skewness if $Q_1 = 13.73$, $Q_3 = 38.29$ and Median = 26.01
3) The first four moments about origin X = 0 are 23.5, 297, 5299.6 and 110306.94 respectively. Find moments about mean also find coefficient of skewness based on moments?

A distribution is said to be normal if its $b_1=0$ and $b_2=3$ respectively. The curve of the normal distribution is Bell-shaped and symmetric.

- For a Bell-shaped symmetric distribution:
  - ✓ 68.27% area of the normal curve lies under the range $\mu \pm \delta$
  - ✓ 95.45% area of the normal curve lies under the range $\mu \pm 2\delta$
  - ✓ 99.73% area of the normal curve lies under the range $\mu \pm 3\delta$
  - ✓ $MeanDeviation = \frac{4}{5}(Standard\ Deviation)$
  - ✓ $QuartileDeviation = \frac{2}{3}(Standard\ Deviation)$
  - ✓ $QuartileDeviation = \frac{5}{6}(MeanDeviation)$

zeestats@gmail.com

Kurtosis

"The degree of peakedness or flatness of a frequency distribution relative to normal distribution is called Kurtosis". OR

"The characteristic by which we compare the "hump" of a distribution with normal distribution is called kurtosis".

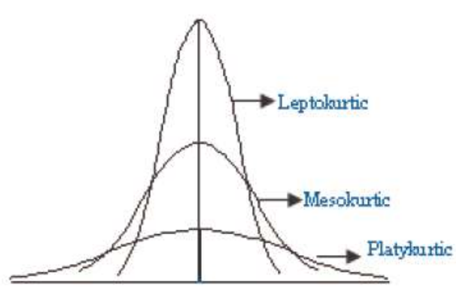
TIP

To measure the skewness we will use:

$$b_2 = \frac{m_4}{m_2^2}$$

Kurtosis indicates whether a particular distribution is flatter or more peaked than the normal curve. Kurtosis is measured by the b2

- If $b_2 > 3$, then the distribution is known as **leptokurtic**
- If $b_2 = 3$, then the distribution is known as **mesokurtic**
- If $b_2 < 3$, then the distribution is known as **platykurtic**



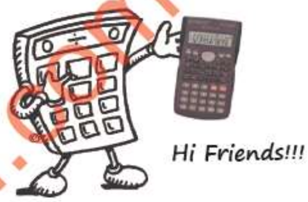
EXAMPLE 4.33

The first four moments about X = 170 are 5.2, 664, 10720 and 145600 respectively. Calculate b_2 and kurtosis of the distribution?

Solution Given that

$$m'_1 = 5.2 \quad , \quad m'_2 = 664$$
$$m'_3 = 10720 \quad , \quad m'_4 = 1456000$$

Now



$$m_1 = 0$$
$$m_2 = m'_2 - (m'_1)^2$$
$$\Rightarrow m_2 = 664 - (5.2)^2 = 637$$
$$m_3 = m'_3 - 3m'_1m'_2 + 2(m'_1)^3$$
$$\Rightarrow m_3 = 10720 - 3(5.2)(664) + 2(5.2)^3 = 642.82$$
$$m_4 = m'_4 - 4m'_1m'_3 + 6(m'_1)^2 m'_2 - 3(m'_1)^4$$
$$\Rightarrow m_4 = 1456000 - 4(5.2)(10720) + 6(5.2)^2(664) - 3(5.2)^4 = 1338558$$

$$b_2 = \frac{m_4}{m_2^2} = \frac{1338558}{637^2} = 3.3$$

Since b_2 is more than 3, therefore the distribution is **leptokurtic**.

zeestats@gmail.com

EXAMPLE 4.34

Find kurtosis using $m_4 = 2.533$ and $m_2 = 0.933$?

Solution Given that $m_4 = 2.533$ and $m_2 = 0.933$

Now $b_2 = \frac{m_4}{m_2^2} = \frac{2.533}{0.933^2} = 2.91$

Since b_2 is less than 3, therefore the distribution is **platykurtic**.

EXAMPLE 4.35

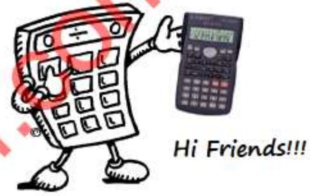
The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75 respectively. Test skewness and kurtosis?

Solution Given that

$$m_1 = 0 \quad , \quad m_2 = 2.5$$
$$m_3 = 0.7 \quad , \quad m_4 = 18.75$$

Skewness:

$$\alpha_3 = \frac{m_3}{\sqrt{m_2^3}} = \frac{0.7}{\sqrt{2.5^3}} = 0.18$$



Therefore the distribution is **positively skewed**.

Kurtosis:

$$b_2 = \frac{m_4}{m_2^2} = \frac{18.75}{2.5^2} = 3$$

Therefore the distribution is **mesokurtic**.

zeestats@gmail.com

Test Yourself

1) The first four moments about X = 34.5 are -11, 260, -5000 and 128000 respectively. Calculate b_2 and kurtosis of the distribution?
2) Find kurtosis using $m_4 = 3646.54$ and $m_2 = 37.24$?
3) The first four central moments of a distribution are 0, 37.24, 55.55 and 3646.54 respectively. Test skewness and kurtosis?