Dispersion
(Measures of Variability)

Introduction and Definition :

Measures of Central tendency are called averages of first order, but these are not sensitive to the variability among the data. Two distributions may have same **Mean**, **Median** and **Mode** but the variability among the data in two distributions may be quiet different.

For example consider two groups ‘A’ and ‘B’ as

Group A	Group B
65	42
66	54
67	58
68	62
71	67
73	77
74	77
77	85
77	93
77	100

Computing Averages we get,

Group A	Group B
Mean = 71.5 Median = 72.0 Mode = 77	Mean = 71.5 Median = 72.0 Mode = 77

It is clear that Group A and Group B have same values of Mean. Median and Mode, but careful perusal of data in both the groups show that the values in Group B are much more widely scattered than the values in group A

Sometimes the two series may have similar formation but their measurement of Measures of Central Tendency may be different.

A	B	C
10	30	100
11	31	101
12	32	102
13	33	103
14	34	104

The series have entirely different average but the same formation. Clearly measurement of central tendency do not indicate how the individual values in the distribution differ from each other or from the central value.

When extent of variation of individual values in relation to other values or in relation to the central value is large, the Measures of Central Tendency fail to represent the series fully.

The Measures of Dispersion (or variability) coupled with the Measures of Central Tendency gives a fairly good idea (not the full idea) about the nature of the distribution.

To have a complete idea about the nature of data **Moments** and **Kurtosis** must also be measured.

Dispersion is the spread or scatter of values from the Measure of Central Tendency.
A Measure of Dispersion is designed to state the extent to which individual observations (or items) vary from their average. Here we shall account only the amount of variation but not the direction.

D.C. Brooks define dispersion as "Dispersion or spread is the degree of scatter or variation of variable about the central value".

Measures of Dispersion are called Averages of **Second order** because they are based on the deviations of the different values from the mean or other measures of central tendency which are called averages of **First order**.

Objectives of Dispersion:

1. To know the average variation of different values from the average of a series.
2. To know about the composition of a series or the dispersion of the values on either sides of the central tendency.
3. To know the range of values.
4. To compare the disparity between two or more series expressed in different units in order to find out the degree of variation.
5. To know whether the Central Tendency truly represent the series or not. If the dispersion is more the central tendency do not represent the series.

Importance of Dispersion :

1. Conclusion drawn from the central tendency carries no meaning without knowing variation of various items of the series from the average.
2. Inequalities in the distribution of wealth and income can be measured in dispersion.
3. Dispersion is used to compare and measure concentration of economic power and monopoly in the country.
4. Dispersion is used in output control and price control.

Characteristics of a good Measure of Dispersion :

1. It should be simple to understand and easy to calculate.
2. It should be rightly defined.
3. It should be based on the all items of the series.
4. It should not be unduly affected by the extreme items of the series.
5. It should be least affected by the sample fluctuations.
6. It should be amenable to the further algebraic treatment.

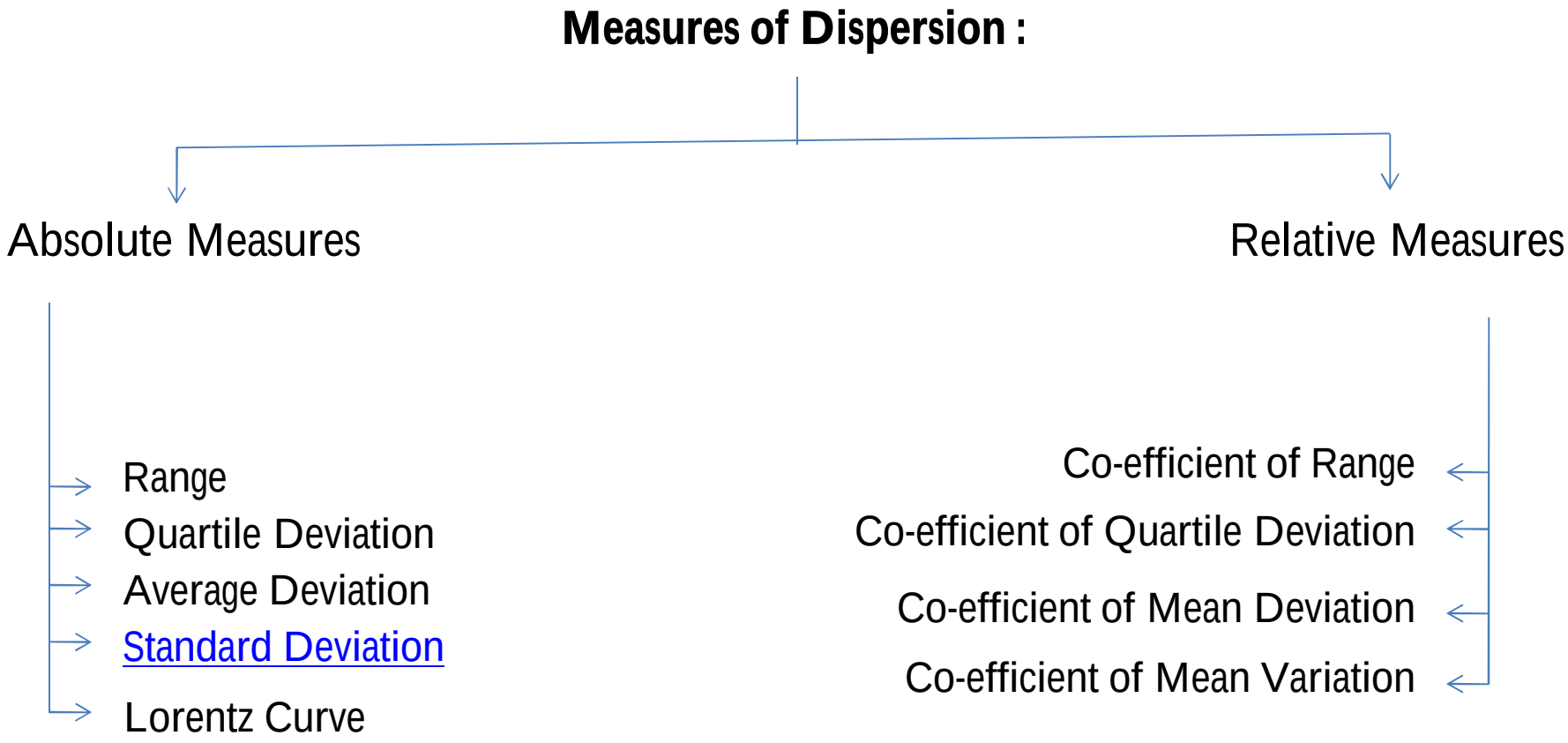
Merits :

1. They indicate the dispersal character of the statistical series.
2. They speak the dependability or reliability of the average value of a series.
3. The enable the statistician in comparing between two or more statistical series with regard to the character of their uniformity or consistency or equitability.
4. They enable the one in controlling the variability of a phenomenon under his purview .
5. They facilitate in making further statistical analysis of the series through devices like co-efficient of Skewness, co-efficient of Kurtosis, co-efficient of correlation, variance analysis etc.
6. They supplement Measures of Central Tendency in finding out more and more information related to the nature of a series.

Demerits :

1. They are liable to misinterpretations and wrong generalization by a statistician of a biased character.
2. They are liable to yield inappropriate results as there are different methods of calculating the dispersion.
3. Except one or two, most of the dispersion involve complicated process of computing.
4. They by themselves can not give any idea about the symmetrical or skewed character of a series.
5. Like measures of central tendency, most of the measures of dispersion do not give a convincing idea about a series to a layman.

Different Measures of Dispersion :



An absolute Measure of Dispersion is expressed in terms of the units of the measurement of the variable. The relative measure of dispersion generally known as co-efficient of dispersion is expressed as a pure number independent of the units of measurement of the variable. The main disadvantage of the absolute measure of dispersion is that it can not be used to compare the variability of two expressions measured with different units. Comparison of distribution with respect to their variability from the central value is done by relative measure of dispersion.

Range:

It is defined as difference between extreme value in the distribution, i.e.,

Range = Largest Value in the Distribution - Smallest Value in the Distribution

In case of continuous frequency distribution range is calculated by any one of the following two methods.

By subtracting the lower limit of the lowest class from the upper limit of the highest class.

OR

By subtracting the mid-value of the lower from mid value of the highest class.

Important:

1. In calculation of Range only the values of the variable are taken in to account and the frequencies are completely ignored.
2. Open ended classes have no Range since they have no highest and lowest value.
3. Some times, variability of two series is measured by Range only though it is a rough measure of variability.

Co-efficient of Range :

$$\text{Co-efficient of Range} = \frac{\text{Max. Value} - \text{Min. Value}}{\text{Max. Value} + \text{Min. Value}}$$

$$= \frac{\text{Absolute Range}}{\text{Sum of the extreme values}}$$

1. Find out the range and its coefficient from the following series.

110, 117, 129, 300, 357, 100, 500, 630, 750

Solution:

Range = Maximum value – Minimum value

Range = 750 – 100 = 650

$$C_r = \frac{Max - Min}{Max + Min}$$

$$C_r = \frac{750 - 100}{750 + 100} = \frac{650}{850} = 0.764$$

2. Find out the range and its coefficient from the following data.

x	*10*	*11*	*12*	*13*	*14*	*15*
f	*8*	*10*	*16*	*20*	*4*	*2*

Solution:

x	*f*
10	*8*
11	*10*
12	*16*
13	*20*
14	*4*
15	*2*

Range = Maximum value – Minimum value

Range = 15 – 10 = 5

$$C_r = \frac{Max - Min}{Max + Min}$$

$$C_r = \frac{15 - 10}{15 + 10} = \frac{5}{25} = 0.2$$

3. Find out the range and its coefficient from the following series.

X	10 – 60	60 – 120	120 – 180	180 – 240	240 – 300
f	3	5	6	3	2

Solution:

x	*f*
10 – 60	3
60 – 120	5
120 – 180	6
180 – 240	3
240 – 300	2

Range = Maximum value – Minimum value

Range = 300 – 10 = 290

$$C_r = \frac{Max - Min}{Max + Min}$$

$$C_r = \frac{300 - 10}{300 + 10} = \frac{290}{310} = 0.93$$

Inter-quartile Range :

$$IR = Q_3 - Q_1$$

Percentile Range :

$$P.R. = P_{90} - P_{10}$$

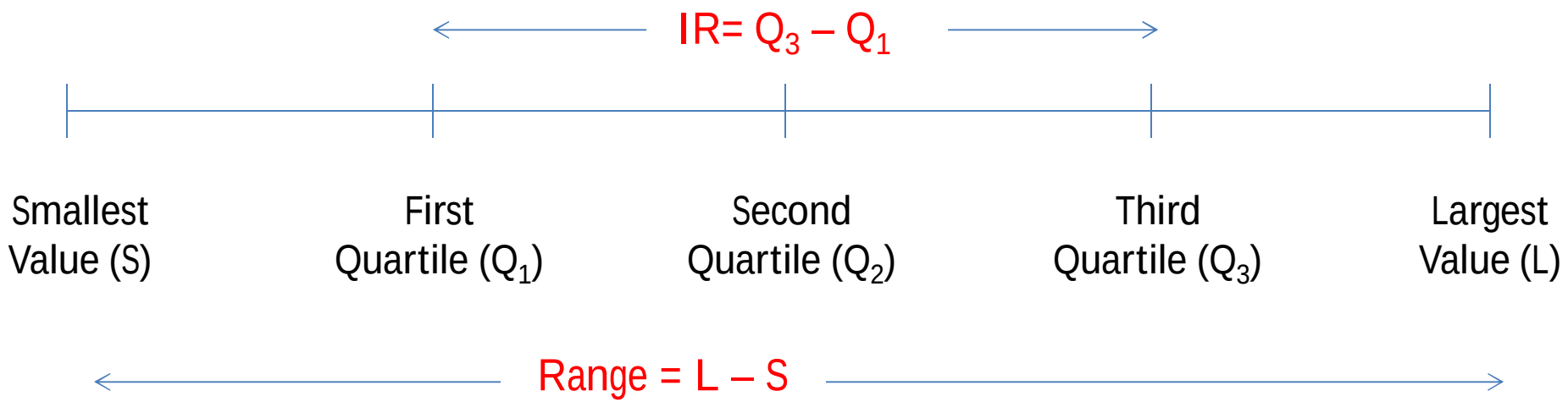
Quartile Deviation :

= Half the difference between the upper and lower quartile

= $(Q_3 - Q_1)/2$ = Semi-inter quartile Range.

Co-efficient of Quartile Deviation :

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



4. Find out the Inter quartile range, semi – interquartile range and coefficient of Q D from the following data.

x	10	11	12	13	14	15

Solution:

$Q_3 = 68.5$ $Q_1 = 22.5$ IR = 46 SIQR = 46/2 = 23

$$\text{Coefficient of QD} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0.505$$

5. Find out the Inter quartile range, semi – interquartile range and coefficient of Q D from the following data.

X	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40
f	4	5	6	10	11	9	4	1

Solution:

$Q_3 = 25.83$ $Q_1 = 22.5$ IR = 12.92 SIQR = 12.92/2 = 6.46

$$\text{Coefficient of QD} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = 0.333$$

For Symmetric Distribution :

$$
\begin{aligned}
Q_3 - Median &= Median - Q_1 \\
2Median &= Q_3 + Q_1 \\
&= Q_3 + Q_1 + Q_1 - Q_1 \\
&= (Q_3 - Q_1) + 2Q_1 \\
&= \left(\frac{Q_3 - Q_1}{2}\right) + Q_1 \\
Median &= Q_1 + Q.D. \ldots\ldots\ldots\ldots\ldots\ldots (1)
\end{aligned}
$$

$$
\begin{aligned}
2Median &= Q_3 + Q_1 + Q_3 - Q_3 \\
&= Q_3 - \left(\frac{Q_3 - Q_1}{2}\right) \\
Median &= Q_3 - Q.D. \ldots\ldots\ldots\ldots\ldots\ldots (2)
\end{aligned}
$$

For Asymmetric Distribution :

Median $\neq Q_1$ + Q.D.
Median $\neq Q_3$ – Q.D.

Merits of Quartile Deviation :

1. Simple to understand and easy to compute.
2. Not affected by extreme values.
3. Computed even if distribution has unequal intervals.
4. Computed in case of open ended intervals.

Demerits of Quartile Deviation :

1. It is not based on the all observations of the series because it does not take frequencies below the lower quartile and above the upper quartile into consideration. .
2. Not amendable to algebraic treatment.
3. Affected by sample fluctuations.
4. It is a distance on the scale and is not a measure from average. Therefore, it fails to show variations around an average.

Use :

The quartile deviation, as a measure of dispersion, is mainly employed in open ended distributions. In many situations, we encounter such distributions because of the need to keep certain information confidential.

Mean Deviations :

Mean deviation (also called Average Deviation) is defined as the arithmetic mean of the absolute deviations of all the values from their Mean or Median or Mode.

$$Mean \quad Deviation \quad = \sum_{i=1}^{n} \frac{\left| x_i - Median \right|}{n} = \sum_{i=1}^{n} \frac{\left| x_i - Mean \right|}{n}$$

for frequency distributi on

$$Mean \quad Deviation \quad = \sum_{i=1}^{n} \frac{f\left| x_i - Median \right|}{N} = \sum_{i=1}^{n} \frac{f\left| x_i - Mean \right|}{N} \qquad M.D. = \frac{\left(\sum f\left| dx \right|\right)}{\sum f}$$

Where N = Σf.

Steps to Calculate Mean deviation in Individual Values (or Observations):

In case of individual observations, the following steps are involved in the calculation of Mean Deviation :

1. Calculate the Mean or Median of a given series.
2. Write down the deviations (dx_i)of each item (x_i) either from the Mean or the Median without considering the sign.
3. Sum up the deviations disregarding the signs, This is = $\left(\sum_{i=1}^{n} |dx_i| \right)$.
4. Divide the total of the deviations by the number of observations and the resulting value is the Mean Deviation.

Steps to Calculate Mean deviation in – Discrete Series :

In case of discrete series, the following steps are involved in the calculation of Mean Deviation :

1. Calculate the Mean or Median of a given series.
2. Write down the deviations (dx_i)of each item (x_i) either from the Mean or the Median without considering the sign.
3. Multiple the deviations by frequencies ($f|dx_i|$).
4. Find sum of the products so obtained. This is $\sum f|dx|$
5. Divide the sum of products by the total frequency and the resulting value is the mean deviation. Expressed as a formula form :

Coefficient of Mean Deviation :

It is a relative measure of dispersion and is computed by the following formula :

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Mean}}$$,When mean is used as a reference point

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Median}}$$,When median is used as a reference point

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Average Used}}$$

5. A rainwear manufacturing company wants to launch some new products in a new state. The rainfall in the state (in cm) for the past 10 years is given in table below. Find out the MD and coefficient of MD

Table : Rainfall for 10 years (1995 - 2004)

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
Rainfall (cm)	110	120	130	135	140	150	160	170	180	190

Solution: Mean = 148.5

Year	*Rainfall (x)*	*(x_i - Mean)*	I(x - Mean)I
1995	110	- 38.5	38.5
1996	120	- 28.5	28.5
1997	130	- 18.5	18.5
1998	135	- 13.5	13.5
1999	140	- 8.5	8.5
2000	150	1.5	1.5
2001	160	11.5	11.5
2002	170	21.5	21.5
2003	180	31.5	31.5
2004	190	41.5	41.5
Total			$\sum_{i=0}^{n} (Ix_i - \bar{x})\, I = 215$

$$Mean\ \ Deviation = \sum_{i=1}^{n} \frac{|x_i - Median|}{n} = \sum_{i=1}^{n} \frac{|x_i - Mean|}{n}$$

$$= \frac{215}{10} = 21.5$$

$$\text{Coefficient of M D} = \frac{\text{Mean Deviation}}{\text{Mean}}$$

$$\text{Coefficient of M D} = \frac{21.5}{148.5} = 0.14$$

6. The weekly earnings of 187 employees of a company is given in table below. Find MD and coefficient of MD.

Table : Weekly earnings and number of employees

Weekly earnings INR	100	120	140	160	180	200	210
No of employees	5	8	12	16	22	44	80

Solution: Mean = 188.55 INR

x	f	fx	$x_i - Mean$	$\|x_i - Mean\|$	$f\|x_i - Mean\|$
100	5	500	- 88.55	88.55	442.75
120	8	960	- 68.55	68.55	548.4
140	12	1680	- 48.55	48.55	582.6
160	16	2560	- 28.55	28.55	456.8
180	22	3960	- 8.55	8.55	188.1
200	44	8800	11.45	11.45	503.8
210	80	16,800	21.45	21.45	1716
Total	187	35260		275.65	4438.45

$$Mean\ \ Deviation = \sum_{i=1}^{n} \frac{f\left|x_i - Median\right|}{N} = \sum_{i=1}^{n} \frac{f\left|x_i - Mean\right|}{N} = \frac{4438\ .45}{187} = 23\ .73$$

$$\text{Coefficient of Mean Deviation} = \frac{\text{Mean Deviation}}{\text{Mean}} = \frac{23.73}{188.5} = 0.1258$$

Ex :

Find the Mean Deviation around Median of the following series :

Marks (x)	5	10	15	20	25
Students	6	7	8	11	8

Solution :

Marks (x)	f	*c.f.*	$\lvert d = x - 15 \rvert$	$f \lvert d \rvert$
5	6	6	10	60
10	7	13	5	35
15	8	21	0	0
20	11	32	5	55
25	8	40	10	80
Total	N = 40			$\sum f \lvert d \rvert = 230$

Median = Value for [(N+1)/2]th term = 20.5 th term = 15

Mean Deviation $= \frac{\sum f|d|}{\sum f} = \frac{230}{40} = 5.75 marks$

Ex :
Find the Mean Deviation around Mean of the following data :

Class Interval	Frequency
0 – 10	8
10 – 20	12
20 – 30	10
30 – 40	8
40 – 50	3
50 – 60	2
60 – 70	7

Solution :
For calculating the Mean deviation :

Mid Value (x)	f	fx	$x - 29$	$\lvert x - \bar{x} \rvert$	$f \lvert x - \bar{x} \rvert$
5	8	40	-24	24	192
15	12	180	- 14	14	168
25	10	250	- 4	4	40
35	8	280	6	6	48
45	3	135	16	16	48
55	2	110	26	26	52
65	7	455	36	36	252
	$N = 50$	$\sum fx = 1450$			$\sum f_i \lvert x_i - \bar{x} \rvert = 800$

$$Mean = \frac{\sum fx}{\sum f} = \frac{1450}{50} = 29$$

$$M.D. = \frac{1}{\sum f} \sum f_i \left| x_i - \bar{x} \right| = \frac{1}{50} \times 800 = 16$$

Standard Deviation :

Standard Deviation is the most important, the most reliable and the most widely used measure of dispersion. The term 'standard' is assigned to this measure of variation probability because of the following reasons.

(i) It is the most commonly used and is the most flexible in terms of variety of applications of all measures of variation.

(ii) The area under any symmetrical curve rather normal curve remains the same with in a fixed number of standard deviations from the Mean on either side of it, e.g., in any normal curve area with in Mean $\pm$ standard deviation is always 68.27% of the total area and the area is 95.45% of the total area with in mean $\pm$ 2 standard deviation.

(iii) The sum of squares of the deviations about the Mean is the least as compared to the sum of the squares of the deviations about the Median or Mode, therefore, root Mean square deviation about the Mean is the least.

It is most important of all the measures of dispersion because it is used in many other statistical operations, e.g., sampling techniques, correlation and regression analysis, finding co-efficient of variation, skewness, kurtosis, etc. standard deviation is also called 'Mean Error' or 'Mean Square Error' or 'Root-Mean Square Deviation' Unlike the Mean Deviation, which may be calculated around any average, the standard deviation is always computed around the Mean.

It is the square-root of the Arithmetic Mean of the squared deviations of all values from their Mean.

Standard Deviation :

For discrete Distribution :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}}$$

For Frequency Distribution :

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i\left(X_i - \overline{X}\right)^2}{\sum f}}$$

Short-Cut Method For Finding Standard Deviation :

$$\sigma = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

Standard deviation is calculated for Continuous Series by calculating the Mid-Points.

Standard Deviation from Step-Deviation Method :

$$d' = \frac{X - A}{h} = \frac{d}{h}$$

$$S.D. = h\sqrt{\frac{\sum d'^2}{n} - \left(\frac{\sum d'}{n}\right)^2}$$

And, if the frequencies are given, then

$$S.D. = h\sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2}$$

Combined Standard Deviation of Two or More Groups :

Let A and B be two groups with n_1 and n_2 the respective number of values and σ_1 and σ_2 the respective standard deviations, then their combined S.D. σ_{12} is given by

$$\sigma_{12} = \sqrt{\frac{n_1\left(\sigma_1^2 + d_1^2\right) + n_2\left(\sigma_2^2 + d_2^2\right)}{n_1 + n_2}}$$

This can also be given by the following formula :

$$\sigma_{12} = \sqrt{\frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1 \times n_2}{(n_1 + n_2)^2}(\bar{x}_1 - \bar{x}_2)^2}$$

$$n_1 = n_2 \Rightarrow$$

$$\sigma_{12} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2} + \frac{1}{4}(\bar{x}_1 - \bar{x}_2)^2}$$

$$\bar{x}_1 = \bar{x}_2 \Rightarrow$$

$$\sigma_{12} = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}$$

Co-efficient of Variation :

$$\text{Karl Pearson's Co-efficient of Variation} = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100\ \%$$

Relations Between Measures of Dispersion :

1. Q.D. = (2/3) S.D.
2. M.D. = (4/5) S.D.
3. A.M. $\pm$ Q.D. would cover 50 % of the items.
4. A.M. $\pm$ S.D. would cover 68.27 % of the item.
5. A.M. $\pm$ M.D. would cover 57.51 % of the items.

Calculate S.D. From the following data:

Class – interval	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
frequency	6	14	10	8	1	3	8

Solution:

Class Interval	*Mid value (x)*	*frequency (f)*	*x - 45 =d*	*d/10 = d'*	*fd'*	*fd'²*
0 - 10	5	6	-40	-4	-24	96
10 - 20	15	14	-30	-3	-42	126
20 - 30	25	10	-20	-2	-20	40
30 - 40	35	8	-10	-1	-8	8
40 - 50	45	1	0	0	0	0
50 - 60	55	3	10	1	3	3
60 - 70	65	8	20	2	16	32
Total		50	-70		-75	305

$$S.D. = 10\sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} = 10\sqrt{\frac{305}{50} - \left(\frac{-75}{50}\right)^2} = 19.6$$

Ex :

Calculate Semi-inter quartile Range and co-efficient of quartile deviation from the following data.

Age (in years)	15 – 25	25 – 35	35 – 45	45 – 55	55 – 65	65 – 75	75 – 85
No. of persons	3	61	132	153	140	51	3

Solution :

Age	Frequency	*c.f.*
15 – 25	3	3
25 – 35	61	64
35 – 45	132	196
45 – 55	153	349
55 – 65	140	489
65 – 75	51	540
75 – 85	3	543
	N = 543	

N/4 = 543/4

= 135.75,

3N/4 = 407.25.

Q_1 = 40.435

Q_3 = 59.16

$Q_3 - Q_1$ = 18.725

$Q_3 + Q_1$ = 99.595

Semi – inter quartile Range = $(Q_3 - Q_1)/2$

= 18.725/2 = 9.365

Co-efficient of quartile deviation = $(Q_3 - Q_1)/(Q_3 + Q_1)$

= 18.725/99.595

= 0.188

Ex :
Calculate the Mean deviation from the Mean for the following data :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70
No. of Students	6	5	8	15	7	6	3

Computational Table

Mid-Value (x)	Frequency (f)	fx	$\lvert x - \bar{x} \rvert$	$\sum f \lvert x - \bar{x} \rvert$
5	6	30	28.4	170.4
15	5	75	18.4	92.0
25	8	200	8.4	67.2
35	15	525	1.6	24.0
45	7	315	11.6	81.2
55	6	330	21.6	129.6
65	3	195	31.6	94.8
	N = 50	$\sum fx = 1670$		$\sum f \lvert x - \bar{x} \rvert = 659.2$

$$\bar{x} = \frac{\sum fx}{N} = \frac{1670}{50} = 33.4$$

$$M.D. = \frac{\sum f\left|x - \bar{x}\right|}{\sum f} = \frac{659.2}{50} = 13.184$$

Ex :

Determine the S.D. from following data :

Yield (in gm)	1216	1374	1167
	1232	1407	1453
	1202	1372	1278
	1141	1221	1329

Solution :

(x)	$x - 1200 = d$	d^2
1216	16	256
1232	32	1024
1202	02	4
1141	-59	3481
1374	174	30276
1407	207	42849
1372	172	29584
1221	21	441
1167	-33	1089
1453	253	64009
1278	78	6084
1329	129	16641
Total	$\sum d = 992$	$\sum d^2 = 195738$

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

$$= \sqrt{\frac{195738}{12} - \left(\frac{992}{12}\right)^2}$$

$$= \sqrt{16311.5 - 6833.7}$$

$$= 97.35$$

Ex :

Find out the S.D. from the following table giving the wages of 230 persons

Wages (Rs)	No. of Persons
140 – 160	12
160 – 180	18
180 – 200	35
200 – 220	42
220 – 240	50
240 – 260	45
260 – 280	20
280 – 300	8

Solution :

Wages (Rs)	Mid Value (x)	x - 230	$d/20 = d'$	f	fd'	d'^2	fd'^2
140 – 160	150	-80	-4	12	-48	16	192
160 – 180	170	-60	-3	18	-54	9	162
180 – 200	190	-40	-2	35	-70	4	140
200 – 220	210	-20	-1	42	-42	1	42
220 – 240	230	0	0	50	0	0	0
240 – 260	250	20	1	45	45	1	45
260 – 280	270	40	2	20	40	4	80
280 – 300	290	60	3	8	24	9	72
				$\Sigma f = 230$	$\Sigma fd' = -105$		$\Sigma fd'^2 = 733$

$$S.D. = \left[\sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f} \right)^2} \right] \times i$$

$$= 20 \sqrt{\frac{733}{230} - \left(\frac{-105}{230} \right)^2}$$

$$= 34.5$$

Case (1) :

Mr Ranveer wants to invest Rs. 10,000 in one of the two companies X and Y. Average return in a year from company X is Rs. 4000 with a Standard Deviation of Rs 25, while in company Y the average return in a year is Rs 5000 with a Standard Deviation of Rs. 40.

Which company would you recommend to Mr Ranveer for investment ? Justify your answer.

Solution :

Coefficient of variance of Company X $= \frac{25}{4000} \times 100 = 0.625$

Coefficient of variance of Company Y $= \frac{40}{5000} \times 100 = 0.8$

Since the co-efficient of variation of company X is less than of Y, Hence company X is more consistence and Mr Ranveer is suggested to invest in company X.

Ex :

Weekly salaries of 100 employees in a firm are given below.

Salaries per week	400 – 500	500 – 600	600 – 700	700 – 800
No. of Employees	2	15	21	30
Salaries per week	800 – 900	900 – 1000	1000 – 1100	
No. of Employees	20	9	3	

Calculate the percentage of employees with salaries, in Rs per week, in the following Range :
(Mean – 2 S.D.), (Mean + 2 S.D.).

Solution :

Salaries in Rs per week	Mid Value *(x)*	*f*	*x – 750 = d*	*d' = d/100*	*fd'*	*d'²*	*fd'²*	*cf*
400 – 500	450	2	- 300	- 3	- 6	9	18	2
500 – 600	550	15	- 200	- 2	- 30	4	60	17
600 – 700	650	21	- 100	- 1	- 21	1	21	38
700 – 800	750	30	0	0	0	0	0	68
800 – 900	850	20	100	1	20	1	20	88
900 – 1000	950	9	200	2	18	4	36	97
1000 – 1100	1,050	3	300	3	9	9	27	100
		$\sum f = 100$			$\sum fd' = -10$		$\sum fd'^2 = 182$	

$$Mean = 750 - \frac{10}{100} \times 100 = 740$$

$$S.D. = 100\sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd'^2}{n}\right)^2} = 100\sqrt{\frac{182}{100} - \left(\frac{-10}{100}\right)^2} = 10\sqrt{182 - 1} = 134.5$$

$$\Rightarrow 2 \times S.D. = 134.5 \times 2 = 269$$

$$Mean - 2 \times S.D. = 740 - 269 = 471$$

$$Mean + 2 \times S.D. = 740 + 269 = 1{,}009$$

If the value 471 corresponds to n_1th value and 1009 corresponds to n_2th value. then

$$471 = 400 + \frac{n_1 - 0}{2} \times 100 \Rightarrow n_1 = \frac{71 \times 2}{100} = 1.4$$

$$1{,}009 = 1{,}000 + \frac{n_2 - 97}{3} \times 100 \Rightarrow n_2 = \frac{9 \times 3}{100} + 97 = 97.27$$

Required Percentage is 97.27 – 1.4 = 96 % approx.

Complete the table showing the frequencies with which words of different number of letters occur in the extract reproduced below (omitting punctuation marks) treating as the variable the number of letters in each word and obtain the mean and standard of the deviation.

"Her eyes were blue: blue as autumn – blue as the blue we see between the retreating mouldings of hills and woody slopes on a sunny September morning: A misty and shady blue that had no beginning on surface, and was looked into rather than at".

Mean A = 4.35

S D = 2.23

No of latters x	*Frequency* f	$d = x - A$	fd	fd^2
1	2	- 5	- 10	50
2	8	-4	-32	128
3	9	-3	-27	81
4	10	-2	-20	40
5	5	-1	-5	5
6	4	0	0	0
7	3	1	3	3
8	1	2	2	4
9	3	3	9	27
10	1	4	4	16
Total	46		- 76	354

Calculate the mean and SD from the following data:

Value	90 - 99	80 - 89	70 - 79	60 - 69	50 - 59	40 - 49	30 – 39
Frequency	2	12	22	20	14	4	1

Mean A = 68.1

S D = 12.505

Initially there were 9 workers, all being paid a uniform wage. Later a 10^{th} worker is added whose wage rate is 20 INR less than for others. Compute

1. The effect on the mean wage.
2. SD of wages for the group of 10 workers.

Solution :

Let the constant wage of each of the 9 workers be 'c' INR. Then the wage of the 10^{th} worker is given to be (c – 20) INR.

The mean wage of 9 workers = c INR

The mean wage of 10 workers = (c – 2)INR

Hence the average wage is decreased by 2.

Variance of 9 workers = 0 (constant)

$$\sigma^2 = \frac{1}{10}\sum(x - \bar{x})^2 = \frac{1}{10}\left[(4 + 4 + 4 + \cdots \ldots + 4) + 324\right]$$

$$= \frac{360}{10} = 36$$

$$\Rightarrow \sigma = \sqrt{36} = 6$$

S No	x	x – Mean = x – (c - 2)	$(x - \text{Mean})^2$
1	C	2	4
2	C	2	4
3	C	2	4
4	C	2	4
5	C	2	4
6	C	2	4
7	C	2	4
8	C	2	4
9	C	2	4
10	C – 20	- 18	324

Lorenz Curve:

Graphically the Dispersion is studied by means of Lorenz Curve. Lorenz Curve is a cumulative percentage curve in which the percentage of items (or frequencies) are shown with the corresponding percentage of factors like income, wealth, profits, etc. The curve is also used to study the distribution of land, wages, income, etc, among the population of a country.

Lorentz curve is a graphic method of studying the dispersion in a distribution. It was first used by Max O Lorentz, an economic satisfaction for the measurement of economic inequalities such as in the distribution of income and wealth between different periods or different countries of the time. But today, Lorentz curve is also used to study disparities of the distribution of wages, profits, turnover, production, population etc.

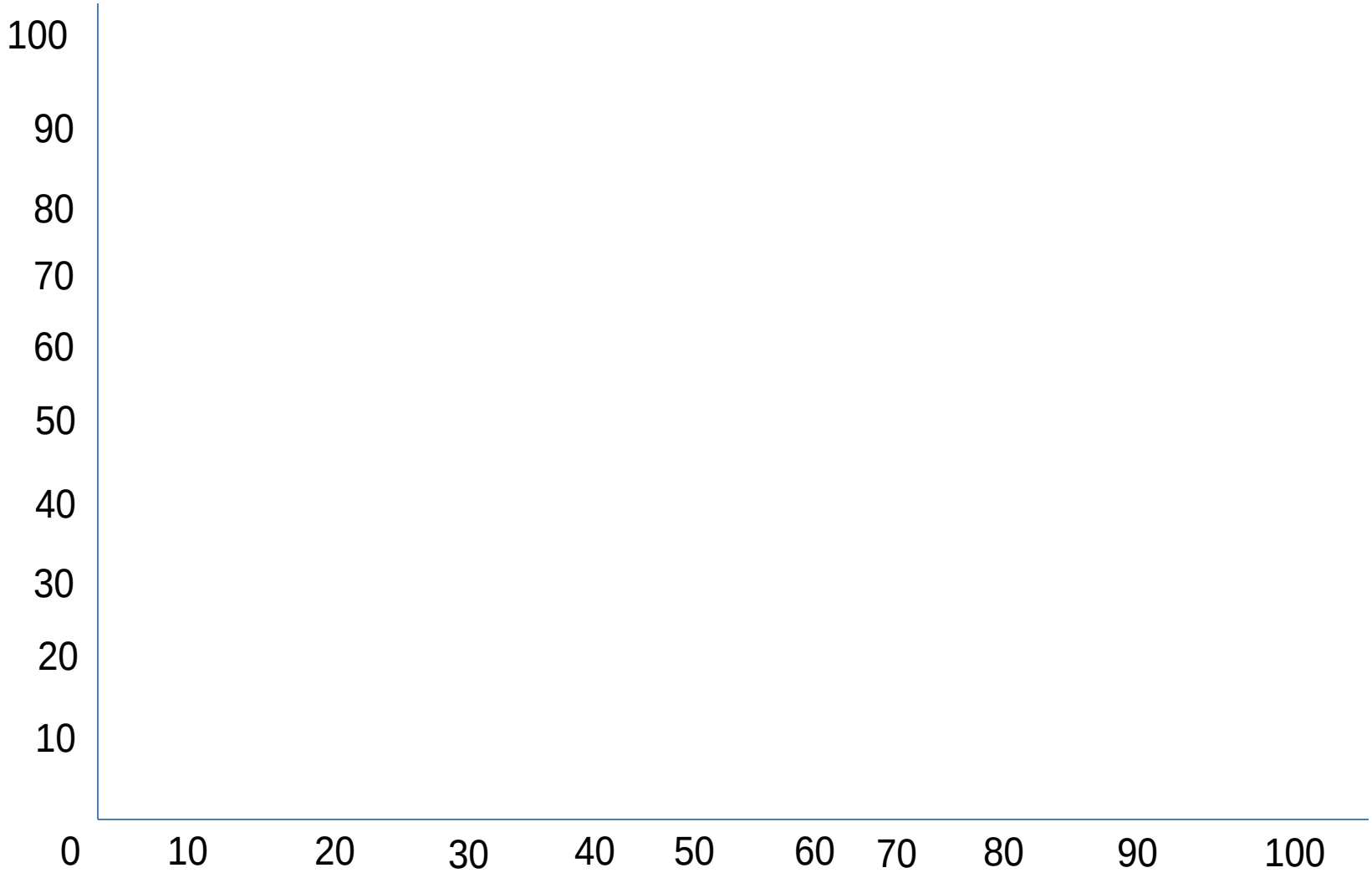
A very distinct feature of the Lorentz curve is in dealing with the cumulative values of the variable and the cumulative frequencies rather than absolute values and the given frequencies.

From the following table giving data regarding income of workers in a factory, draw a Lorentz curve to study inequality of income.

Income (INR)	*No of workers*
Below 500	6000
500 – 1000	4, 250
1, 000 – 2, 000	3, 600
2, 000 – 3, 000	1, 500
3, 000 – 4, 000	650

Solution :

Income	*Mid – value*	*Cumulative Income*	*% of cumulative income*	*No of workers (f)*	*Cf*	*% of cf*
0 – 500	250	250	2.94	6000	6000	37.5
500 – 1000	750	1000	11.76	4250	10250	64.1
1000 – 2000	1500	2500	29.41	3600	13850	86.6
2000 – 3000	2500	5000	58.82	1500	15350	95.9
3000 - 4000	3500	8500	100.00	650	16000	100.0
Total	**8500**			**16000**		


100
90
80
70
60
50
40
30
20
10
0
10
20
30
40
50
60
70
80
90
100

Moments, Skewness and Kurtosis

Group A	Group B
67	66
68	66
66	68
71	68
74	70
77	71
73	72
77	73
65	77
72	79

Mean A	71
Mean B	71
SD Group A	4.4
SD Group B	4.4

MOMENTS

For a frequency distribution having observations $x_1, x_2, \ldots, x_n$ with respective frequencies $f_1, f_2, \ldots\ldots, f_n$ the r^{th} moment about mean is defined as;

$$\mu_r = \frac{1}{N}\sum f_i\left(X - \overline{X}\right)^r$$

These moments are also called ***central moments.***

In particular;

If r = 0, we have

$$\mu_0 = \frac{1}{N}\sum f_i\left(X - \overline{X}\right)^0 = \frac{1}{N}\sum f_i \times 1 = \frac{\sum f_i}{N} = \frac{N}{N} = 1$$

If r = 1, we have

$$\mu_1 = 0$$

If r = 2, we have

$$\mu_2 = \sigma^2$$

Moments about an Arbitrary Value;

$$\mu'_r = \frac{1}{N}\sum f_i (X_i - A)^r$$

Where, A is assumed mean. These moments are also known as **Raw Moments.**

$$\mu'_1 = \overline{X} - A$$

Example;

Calculate first four moments about 30 for the following distribution of age group;

Class Interval	5 – 15	15 – 25	25 – 35	35 – 45	45 – 55
Frequency	8	12	15	9	6

Solution;

Class Intervals	***Frequency (f)***	***Mid Value (X)***	***X – 30***	***f(X – 30)***	***$f(X - 30)^2$***	***$f(X - 30)^3$***	***$f(X - 30)^4$***
5 – 15	8	10 y	- 20 y	- 160 y	3200 y	- 64000 y	1280000 y
15 – 25	12	20 y	- 10 y	- 120 y	1200 y	-12000 y	120000 y
25 – 35	15	30 y	0 y	0 y	0 y	0 y	0 y
35 – 45	9	40 y	10 y	90 y	900 y	9000 y	90000 y
45 – 55	6	50 y	20 y	120 y	2400 y	48000 y	960000 y
Total	50			- 70 y	7700 y	- 19000 y	2450000 y

$$\mu_1' = \frac{1}{50}\sum f_i(X_i - A)^1 = \frac{-70}{50} = -1.40 \quad y$$

$$\mu_2' = \frac{1}{50}\sum f_i(X_i - A)^2 = \frac{7700}{50} = 154 \quad y$$

$$\mu_3' = \frac{1}{50}\sum f_i(X_i - A)^3 = \frac{-19000}{50} = -380 \quad y$$

$$\mu_4' = \frac{1}{50}\sum f_i(X_i - A)^4 = \frac{2450000}{50} = 49000 \quad y$$

Coefficients Based on Moments;

So far it has been observed that moments also represented in terms of units of variable X. In order to compare moments of two or more distributions it is necessary to define them into coefficients.

Let us transform the variable X in to another variable;

$$z_i = \frac{X_i - \overline{X}}{\sigma} \quad \text{Where,} \quad \bar{X} = 0, \quad \sigma = 1$$

Then ***r^th^*** order moment of z about zero, denoted by α_r is given by;

$$\alpha_r = \frac{1}{N} \sum f_i z_i^r$$

From formula above;

$$\alpha_1 = 0$$

$$\alpha_2 = \frac{\mu_2}{\sigma^2} = 1$$

$$\alpha_3 = \frac{\mu_3}{\sigma^3}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{\mu_4}{\mu_2^2}$$

Karl Pearson Suggested two Beta Coefficients, β_1 and β_2

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

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \alpha_4$$

Skewness;

It refers to asymmetry of any distribution. The symmetry of a distribution means that from a given deviation from a central value, there are equal number of observations on either side of the it.

If the distribution is asymmetrical or skewed, its frequency curve would have a prolonged tail either towards its left or right hand side.

Hence the skewness of a distribution is defined as the **departure form symmetry.**

For a symmetrical distribution

Mean = Median = Mode

For Positively Skewed distribution

Mode < Median < Mean

(i.e. more values are on the right hand side of the distribution than left)

For Negatively Skewed distribution

Mean < Median < Mode

(i.e. more values are on the left hand side of the distribution than Right)

Measure of Skewness;

1. Based on Measure of Mean, Median and Mode.
2. Based on Quartiles and Percentiles.
3. Based on Moments.

Based on Measure of Mean, Median and Mode.

$$S_k = \frac{\overline{X} - M_o}{\sigma}$$

or

$$S_k = \frac{3\left(\overline{X} - M_d\right)}{\sigma}$$

If $S_k > 0$, the distribution is '+'vely skewed

If $S_k < 0$, the distribution is '–' vely skewed

If $S_k = 0$, distribution is symmetrical

Based on Quartiles and Percentiles;

$$S_Q = \frac{Q_3 + Q_1 - 2M_d}{Q_3 - Q_1}$$

Known as **Bowly's Coefficient of Skewness**

$$S_P = \frac{P_{90} + P_{10} - 2P_{50}}{P_{90} - P_{10}}$$

Value of S_k lies between + 1 and - 1

Based on Moments;

$$S_M = \alpha_3 = \frac{\mu_3}{\sigma^3} = \pm\sqrt{\beta_1} = \gamma_1$$

Since β is always '0' therefore sign of S_k is given by sign of μ_3

Example;

Calculate the Karl Pearson's coefficient of Skewness from the following data;

Size	1	2	3	4	5	6	7
Frequency	10	18	30	25	12	3	2

Solution;

For given distribution;

Mean = 3.28

S.D. = 1.35

Mode = 3.00

$$S_k = \frac{\overline{X} - M_o}{\sigma} = \frac{3.28 - 3.0}{1.35} = 0.207$$

The distribution is moderately positively skewed

Example;

Calculate the Bowly's coefficient of Skewness from the following data;

Size	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40
Frequency	7	10	20	13	17	10	14	9

$$S_Q = 0.06$$

Distribution is approximately symmetrical

Measures of Kurtosis;

A measure of coefficient of Kurtosis is given by Karl Pearson; That says if

β_2 = 3 Curve is Mesokurtic or Normal

β_2 > 3 Curve is Leptookurtic

β_2 < 3 Curve is Pletykurtic

Example;

Calculate Measure of Kurtosis;

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Frequency	10	20	40	20	10

Solution;

Here,

β_2 = 2.5, Therefore the distribution is platykurtic.

Group A	X - 71
67	-4
68	-3
66	-5
71	0
74	3
77	6
73	2
77	6
65	-6
72	1

Short Cut Method of finding Mean Deviation :

When mean (or median) is not a whole number but a fraction, the following short-cut formula can be used :

Mean Deviation from Mean

$$= \frac{[\text{Sum of the values} > \text{Mean}] - [\text{Sum of the values} < \text{Mean}]}{\text{Total Number of Values}}$$

Mean Deviation from Median

$$= \frac{\text{[Sum of the values } > \text{ Median] - [Sum of the values } < \text{ Median]}}{\text{Total Number of Values}}$$

Short Cut Method of finding Mean Deviation (for Frequency Distribution):

$$\text{Mean Deviation from Mean} = \frac{\sum f|dx| + (f_a - f_b) \times c}{\sum f} = \frac{\sum f|(x - A)| + (f_a - f_b) \times c}{\sum f}$$

Where,

$\sum f|dx| =$ Sum of the products of the absolute deviations [Considering all deviations positive or negative as positive] and the respective frequencies when the deviations are taken from Assumed Mean

$f_a =$ Sum of the frequencies above the Mean = Sum of the values of the frequency less than the Mean.

f_b = Sum of the frequencies below the Mean = Sum of the values of the frequency more than the Mean.

C = Difference between the real Mean and arbitrary Mean.

Similarly, We can find the Mean Deviation from Median

$$\text{Mean Deviation from Median} = \frac{\sum f|dx| + (f_a - f_b) \times c}{\sum f}$$

$\sum f|dx|$ = Sum of the products of the absolute deviations [Considering all deviations positive or negative as positive] and the respective frequencies when the deviations are taken from Assumed Median

f_b = Sum of the frequencies below the Mean = Sum of the values of the frequency more than the Median.

C = Difference between the real Median and arbitrary Median.

Note : The assumed Mean or Median must be selected from the class in which real Mean or Median lies.

Mean deviation may also be calculated from following method :

M.D. from Mean $= \dfrac{\sum Xf_a - \sum Xf_b - \left(\sum f_a - \sum f_b\right)\overline{X}}{N}$

$\sum Xf_a =$ Sum of the products of Mid-point (X) and frequencies above the Mean.

$\sum Xf_b =$ Sum of the products of Mid-point (X) and frequencies below the Mean.

$\sum f_a =$ Sum of frequencies of Mid-points above the Mean.

$\sum f_b =$ Sum of frequencies of Mid-points below the Mean.

$N =$ Total number of observations.

M.D. from Median $= \dfrac{\sum Xf_a - \sum Xf_b - \left(\sum f_a - \sum f_b\right)(Median)}{N}$

Steps to Calculate Mean deviation in – Continuous Series :

As regards the calculation of mean deviation in a continuous series, the procedure to be adopted same as in the case of discrete series but with a minor difference. Here classes are replaced by mid-values and frequencies are multiplied by deviation of the mid-values from the mean.