8
MEASURES OF DISPERSION

In a series, all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion.

Averages are central values. They enable comparison of two or more sets of data. They are not sufficient to depict the true nature of the sets. For example, consider the following marks of two students.

Student I	Student II
68	85
75	90
65	80
67	25
70	65

Both have got a total of 345 and an average of 69 each. The fact is that the second student has failed in one paper. When the averages alone are considered, the two students are equal.

Less variation is a desirable characteristic. First student has less variation. That is, he is almost equally good in all the subjects. To quote Simpson and Kafka, "*An average does not tell the full story. It is hardly fully representative of a mass, unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to guage how representative the average is*".

Consider the following example.

Group I

Student C		Student D		Student E	
Mark	Deviation from Mean	Mark	Deviation from Mean	Mark	Deviation from Mean
X	$X-\overline{X}$	X	$X-\overline{X}$	X	$X-\overline{X}$
60	0	50	–10	20	–40
60	0	55	– 5	40	–20
60	0	60	0	60	0
60	0	65	5	80	20
60	0	70	10	100	40
Total: 300		300		300	
Mean($\overline{X}$):60		60		60	

Group II

Student F		Student G		Student H	
Mark	Deviation from Mean	Mark	Deviation from Mean	Mark	Deviation from Mean
X	$X-\overline{X}$	X	$X-\overline{X}$	X	$X-\overline{X}$
0	–10	60	–10	80	–10
5	– 5	65	– 5	85	– 5
10	0	70	0	90	0
15	5	75	5	95	5
20	10	80	10	100	10
Total : 50		350		450	
Mean($\overline{X}$): 10		70		90	

The three students in Group I have equal mean marks. The variations in their marks are not equal. There is no variation in the marks of C. There is smaller variation in the marks of D compared with that in the marks of E. When the mean mark is considered as a representative of all the marks of a student, C's mean mark is a very good representative. In Statistics, there is no chance for all the values being equal. There is small variation in the marks of D. His mean mark is a good representative of his marks. His marks are spread close to the mean. In other words, they scatter not much away from the mean. Mean is reliable in this case. When the mean is said to be 60, people expect the individual marks to be near 60. They are so. This property is not found in the marks of E. The individual marks are far away from the mean. The expectation of the people that his individual marks will be near 60 fails.

The reliability of mean as a good representative fails due to large variation in the individual values. Thus, smaller is the variation in the individual values, more is the reliability of a measure of central tendency.

The three students in Group II do not have equal mean marks. The variations in their marks seem to be equal.

There will be other students whose mean marks and variations in marks differ. Certain measures of variation determine the amount of variation in a series; certain measures of variation are useful for comparison of two or more series.

According to A.L. Bowley, *"Dispersion is the measure of the variation of the items"*. In the words of Murray R.Spiegel, *"The degree to which numerical data tend to spread about an average value is called* the ***variation or dispersion of the data"***.

Importance or Significance of Measures of Dispersion

Dispersion is measured for the following purposes:

1. ***The reliability of a measure of central tendency is known.*** Less dispersion means less variation in the values. The values are close to each other. They are near a measure of central tendency. The measure of central tendency in such a case is more representative and more reliable. A measure of dispersion indicates to what extent a measure of central tendency represents the data and is reliable.

2. ***Measures of Dispersion provide a basis for the control of variability.*** They tell what is the quantum of variation in any set of data. On the basis of that it can be decided whether the variation may be allowed to exist as it is or is to be reduced. In Statistical Quality Control (S.Q.C.), for example, whether the variation is due to assignable causes or chance causes is found out and the next step is planned.

3. ***They help to compare two or more sets of data with regard to their variability.*** Relative measures of dispersion are used for this purpose.

4. ***They enhance the utility and scope of statistical techniques.*** This aspect is seen in the study of correlation, regression, tests of significance, S.Q.C., etc.

Criteria or requisites or characteristics or desirable properties of a measure of dispersion

An ideal measure of dispersion is expected to possess the following properties. The properties are given briefly. They are similar to those of measures of central tendency.

1. It should be rigidly defined.
2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy to calculate.
6. It should have sampling stability.

Absolute and Relative Measures

There are two kinds of measures of dispersion, viz., absolute measures of dispersion and relative measures of dispersion. Absolute measures indicate the amount of variation in a set of values. They are quoted in terms of the units of observations. For example, when rainfall on different days are available in cm., any absolute measure of dispersion gives the variation in rainfall in cm. On the contrary, when rainfall on different days are quoted in mm., an absolute measure of dispersion gives the variation in rainfall in mm. Relative measures are used to compare the variation in two or more sets. They are free from the units of measurements of the observations. They are pure numbers. For example, when rainfall on different days are given in cm., a relative measure such as coefficient of variation does not give the variation in cm. Consequently rainfalls in two places, say, one in cm and the other in inch, can be compared using coefficient of variation. Further, the set which has less variation is said to be less variable or more stable or more consistent or more uniform or more homogeneous, etc. The various absolute and relative measures of dispersion are listed below.

Absolute Measures	Relative Measures
1. Range	1. Coefficient of Range
2. Quartile Deviation (Q.D.) or Semi Inter Quartile Range	2. Coefficient of Quartile Deviation
3. (i) Mean Deviation (M.D.) (about Mean)	3 (i) Coefficient of Mean Deviation (about Mean)
(ii) Mean Deviation (M.D.) about Median	(ii) Coefficient of Mean Deviation about Median
(iii) Mean Deviation (M.D.) about Mode	(iii) Coefficient of Mean Deviation (about Mode)
4. Standard Deviation (S.D.)	4. Coefficient of Variation
5. Variance	

Lorenz curve exhibits the inequality in the distribution of values.

RANGE

Definition : *Range is the difference between the greatest (largest) and the smallest of the values.*

In symbols, Range = L–S

L - Largest Value

S - Smallest Value

In individual observations and discrete series, L and S are easily identified. In continuous series, the following two methods are followed.

Method 1 :

L - Upper boundary of the highest class

S - Lower boundary of the lowest class

Method 2 :

L - Mid value of the highest class

S - Mid value of the lowest class

$$\text{Coefficient of Range} = \frac{L-S}{L+S}$$

Example 1 : Find the value of range and its coefficient for the following data :

8 10 5 9 12 11

(B.Com. Bharathiar, A 02)

Solution : L = 12 S = 5

$\therefore$ Range = L – S

= 12 – 5

= 7

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{12-5}{12+5} = \frac{7}{17} = 0.4118$$

Example 2 : Calculate range and its coefficient from the following distribution:

Size :	60-62	63-65	66-68	69-71	72-74
Number :	5	18	42	27	8

(B.Com. Bharathiar, A 01)

Solution :

Method 1 : After rewriting the class intervals continuously, the lower boundary of the lowest class, S = 59.5 and the upper boundary of the highest class, L = 74.5.

$$\therefore \text{Range} = L - S = 74.5 - 59.5 = 15$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{74.5-59.5}{74.5+59.5} = \frac{15}{134} = 0.1119$$

Method 2 :

Mid value of the lowest class, S = 61

Mid value of the highest class, L = 73

$$\therefore \text{Range} = L - S = 73 - 61 = 12$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{73-61}{73+61} = \frac{12}{134} = 0.0896.$$

Uses of Range

Range is not a popular measure of dispersion. It is used in the following few situations.

1. Range is used in finding the control limits of Mean chart and Range chart in S.Q.C.

2. While quoting the prices of shares, bonds, gold, etc. on daily basis or yearly basis, the minimum and the maximum prices are mentioned.

3. The minimum and the maximum temperature likely to prevail on each day are forecasted.

Merits:

1. It is simple to understand and easy to calculate.
2. It can be calculated in no time.

Demerits :

1. Its definition does not seem to suit continuous series. Two methods are there for its calculation from such series.
2. It is based on the two extreme items. It does not consider the other items.
3. It is unduly affected by the extreme items.
4. It can not be manipulated algebraically. The range of combined set cannot be found from the range of the individual sets. It can be found only when the individual values are known; or atleast the values of the extreme items are known.
5. It does not have sampling stability.
6. It cannot be calculated from open-end class intervals.
7. It is a very rarely used measure. Its scope is limited to the three situations indicated earlier.

QUARTILE DEVIATION (Q.D.)

Definition : *Quartile Deviation is half of the difference between the first and the third quartiles.* Hence it is called **Semi Inter Quartile Range.**

In symbols, Q.D. = $\frac{Q_3 - Q_1}{2}$ Q.D. is the abbreviation. Among the quartiles Q_1, Q_2 and Q_3, the range is $Q_3 - Q_1$. Hence, $Q_3 - Q_1$ is called inter quartile range and $\frac{Q_3 - Q_1}{2}$, semi inter quartile range.

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

As mentioned in the previous chapter, 25% of the values are below or equal to Q_1 ; 25% above or equal to Q_3. $Q_3 - Q_1$ is the distance between Q_1 and Q_3. Central 50% of the items lie between Q_1 and Q_3. It is customery to consider $\frac{Q_3 - Q_1}{2}$ as an absolute measure of dispersion.

Definitions and calculations of Q_1 and Q_3 for all types of data were considered in the previous chapter.

Example 3 : What do you mean by Quartile Deviation? Find the Quartile Deviation for the following:-

391, 384, 591, 407, 672, 522, 777, 733, 1490, 2488.

(B.Com. Madras, S 74)

Solution: The given values in ascending order :

384, 391, 407, 522, 591, 672, 733, 777, 1490, 2488.

Position of Q_1 is $\frac{N+1}{4} = \frac{10+1}{4} = 2.75$.

$\therefore Q_1 = 2^{nd}$ value + 0.75 (3^{rd} value – 2^{nd} value)

$= 391 + 0.75\,(407-391)$

$= 391 + 0.75 \times 16$

$= 391 + 12$

$= 403$

Position of Q_3 is $3\left(\frac{N+1}{4}\right) = 3 \times 2.75 = 8.25$

$\therefore Q_3 = 8^{th}$ value + 0.25 (9^{th} value– 8^{th} value)

$= 777 + 0.25\,(1490 - 777)$

$= 777 + 0.25 \times 713$

$= 777 + 178.25$

$= 955.25$

$$\therefore Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \frac{955.25 - 403.00}{2}$$

$$= \frac{552.25}{2}$$

$$= 276.13$$

Example 4: Weekly wages of a labourer are given below. Calculate Q.D. and coefficient of Q.D.

Weekly Wage (Rs.)	100	200	400	500	600	Total
No. of Weeks	5	8	21	12	6	52

Solution:

Weekly Wage (Rs)	No.of Weeks	Cum. Freq.
100	5	5
200	8	13
400	21	34
500	12	46
600	6	52
Total	N=52	--

Position of Q_1 is $\frac{N+1}{4} = \frac{52+1}{4}$

$= 13.25.$

$\therefore Q_1 = 13^{th}$ value + 0.25 (14^{th} value – 13^{th} value)

$= 200 + 0.25\,(400-200)$

$= 200 + 0.25 \times 200$

$= 200 + 50$

$= 250$

Position of Q_3 is $3\left(\frac{N+1}{4}\right) = 3 \times 13.25 = 39.75$

$\therefore Q_3 = 39^{th}$ value + 0.75 (40^{th} value – 39^{th} value)

$= 500 + 0.75\,(500 - 500)$

$= 500 + 0.75 \times 0$

$= 500 + 0$

$= 500$

$\therefore$ Q.D. $= \frac{Q_3 - Q_1}{2} = \frac{500-250}{2} = \frac{250}{2} = 125$

$\therefore$ Coefficient of Q.D.$= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{500-250}{500+250} = \frac{250}{750} = 0.3333$

Example 5: For the data given here, give the quartile deviation.

X	351-500	501-650	651-800	801-950	951-1100
f	48	189	88	47	28

(B.Com. Bharathidasan, A 01)

Solution :

X	f	True class intervals	Cumulative Frequency
351- 500	48	350.5 – 500.5	48
501- 650	189	500.5 – 650.5	237 ←
651- 800	88	650.5 – 800.5	325 ←
801- 950	47	800.5 – 950.5	372
951-1100	28	950.5 –1100.5	400
Total	N = 400	------	-----

$$\frac{N}{4} = \frac{400}{4} = 100; \quad Q_1 \text{ class is } 500.5 - 650.5$$

$$\therefore L_1 = 500.5;\ f_1 = 189;\ i_1 = 650.5-500.5=150;\ cf_1 = 48$$

$$\therefore Q_1 = L_1 + \left[\frac{i_1\left(\frac{N}{4} - cf_1\right)}{f_1}\right]$$

$$= 500.5 + \left[\frac{150(100-48)}{189}\right]$$

$$= 500.5 + \left[\frac{150 \times 52}{189}\right]$$

$$= 500.5 + 41.27$$

$$= 541.77$$

$$\frac{3N}{4} = 3 \times 100 = 300; \quad Q_3 \text{ class is } 650.5\text{–}800.5$$

$$\therefore L_3 = 650.5;\ f_3 = 88;\ i_3 = 800.5-650.5=150;\ cf_3 = 237$$

$$\therefore Q_3 = L_3 + \left[\frac{i_3\left(\frac{3N}{4} - cf_3\right)}{f_3}\right]$$

$$= 650.5 + \left[\frac{150(300-237)}{88}\right]$$

$$= 650.5 + \left[\frac{150 \times 63}{88}\right]$$

$$= 650.5 + 107.39$$

$$= 757.89$$

$$\therefore Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \frac{757.89 - 541.77}{2}$$

$$= \frac{216.12}{2}$$

$$= 108.06$$

Merits

1. It is simple to understand and easy to calculate.
2. It is not affected by extreme items.
3. It can be calculated for data with open end classes also.

Demerits

1. It is not based on all the items. It is based on two positional values Q_1 and Q_3 and ignores the extreme 50% of the items.
2. It cannot be manipulated algebraically.
3. It is affected by sampling fluctuations.
4. Like range, it does not measure the deviation about any measure of central tendency.

MEAN DEVIATION OR AVERAGE DEVIATION

Definition: *Mean deviation is the arithmetic mean of the absolute deviations of the values about their arithmetic mean or median or mode.*

M.D. is the abbreviation for Mean Deviation. There are three kinds of mean deviations, viz.,

(i) mean deviation or mean deviation about mean
(ii) mean deviation about median
(iii) mean deviation about mode.

Mean deviation about median is the least. It could be easily verified in individual observations and discrete series where the actual values are considered.

The relative measures are the following:

(i) Coefficient of Mean Deviation (about Mean)

$$= \frac{\text{Mean Deviation about Mean}}{\text{Mean}}$$

(ii) Coefficient of Mean Deviation about Median

$$= \frac{\text{Mean Deviation about Median}}{\text{Median}}$$

(iii) Coefficient of Mean Deviation about Mode

$$= \frac{\text{Mean Deviation about Mode}}{\text{Mode}}$$

Individual Observations

$$\text{Mean Deviation (about Mean)} = \frac{\Sigma\left|X - \overline{X}\right|}{N}$$

The mean, $\overline{X} = \frac{\Sigma X}{N}$ is calculated first. From each X,

$\overline{X}$ is subtracted. Every $(X-\overline{X})$ is treated positive and is written as $|X-\overline{X}|$ (read, modulus). Total of $|X-\overline{X}|$ is found and is divided by N to get the mean deviation about mean.

$X-\overline{X}$ is the deviation of X from $\overline{X}$. If $(X-\overline{X})$ is negative and negative sign is considered, the total deviation, $\Sigma(X-\overline{X})$ is zero. It is the first mathematical property of arithmetic mean. Consequently average deviation $\Sigma(X-\overline{X})/N$ is zero. It will be of no use to consider the average deviation in such a form. Hence, the above formula. Similarly,

Mean Deviation about Median $= \dfrac{\Sigma|X-M|}{N}$ and

Mean Deviation about Mode $= \dfrac{\Sigma|X-Z|}{N}$

Median or Mode, whichever is required, is calculated first. Then, as in M.D. about mean, other calculations follow.

Example 6 : Calculate mean deviation for the numbers 1 2 3 4 5 (B.Com.Bharathidasan, N 01)

Solution :

Numbers	
X	$\lvert X-\overline{X}\rvert$
1	2
2	1
3	0
4	1
5	2
$\Sigma X=15$	$\Sigma\lvert X-\overline{X}\rvert=6$

Arithmetic Mean, $\overline{X} = \dfrac{\Sigma X}{N}$

$= \dfrac{15}{5}$

$= 3$

Mean Deviation (M.D.) $= \dfrac{\Sigma|X-\overline{X}|}{N}$

$= \dfrac{6}{5}$

$= 1.20$

Note: If the coefficient of M.D. about Mean is required,

Coefficient of M.D. about Mean $= \dfrac{\text{M.D.}}{\text{Mean}}$

$= \dfrac{1.20}{3}$

$= 0.4000$

Example 7 : Calculate mean deviation about median for the items

7 4 10 9 15 12 7 9 7

(B.Com.Bharathidasan,A 01)

Solution :

Items X	$\|X-M\|$
4	5
7	2
7	2
7	2
9	0
9	0
10	1
12	3
15	6
Total	$\Sigma\|X-M\|=21$

Items are considered in ascending order.

Position of Median (M) is $\frac{N+1}{2}=\frac{9+1}{2}=5$

$\therefore$ Median (Item at 5[th] position) = 9

$$\text{M.D. about Median} = \frac{\Sigma|X-M|}{N} = \frac{21}{9} = 2.33$$

Note : If the Coefficient of M.D. about Median is required,

$$\text{Coefficient of M.D. about Median} = \frac{\text{M.D. about Median}}{\text{Median}} = \frac{2.33}{9} = 0.2589$$

Example 8: Daily earnings in Rs.(X) of 10 coolies are given. Calculate all the three mean deviations and the corresponding relative measures.

X: 32 51 23 46 20 78 57 56 57 30

Solution :

X (Rs.)	$\|X-\bar{X}\|$ $\bar{X}=45$	$\|X-M\|$ M=48.5	$\|X-Z\|$ Z=57
20	25	28.5	37
23	22	25.5	34
30	15	18.5	27
32	13	16.5	25
46	1	2.5	11
51	6	2.5	6
56	11	7.5	1
57	12	8.5	0
57	12	8.5	0
78	33	29.5	21
$\Sigma X=$ 450	$\Sigma\|X-\bar{X}\|=$ 150	$\Sigma\|X-M\|=$ 148.0	$\Sigma\|X-Z\|=$ 162

$$\overline{X} = \frac{\Sigma X}{N} = \frac{450}{10} = Rs.45$$

$$\text{M.D. about Mean} = \frac{\Sigma|X - \overline{X}|}{N} = \frac{150}{10} = Rs.15$$

$$\text{Coefficient of M.D about Mean} = \frac{M.D.}{Mean} = \frac{15}{45} = 0.3333$$

$$\text{Position of Median, M is } \frac{N+1}{2} = \frac{10+1}{2} = 5.5$$

$$\text{Median} = \frac{5^{th}\text{ item} + 6^{th}\text{ item}}{2} = \frac{46+51}{2} = Rs.48.50$$

$$\text{M.D. about Median} = \frac{\Sigma|X - M|}{N} = \frac{148}{10} = Rs.14.80$$

$$\text{Coefficient of M.D. about Median} = \frac{\text{M.D. about Median}}{\text{Median}} = \frac{14.80}{48.50} = 0.3052$$

Mode, Z = Rs.57

$$\text{Mean deviation about Mode} = \frac{\Sigma|X - Z|}{N} = \frac{162}{10} = Rs.\ 16.20$$

$$\text{Coefficient of M.D. about Mode} = \frac{\text{M.D. about Mode}}{\text{Mode}} = \frac{16.20}{57} = 0.2842$$

Note: Mean Deviation about Median is the least.

Discrete Series

The measure of central tendency Mean or Median or Mode is calculated first. The following formulae are used later.

$$\text{Mean Deviation (about Mean)} = \frac{\Sigma f|X - \overline{X}|}{N}$$

$$\text{Mean Deviation about Median} = \frac{\Sigma f|X - M|}{N}$$

$$\text{Mean Deviation about Mode} = \frac{\Sigma f|X - Z|}{N}$$

Example 9 :

X	2	4	6	8	10
f	1	4	6	4	1

Find mean deviation for the above data.

(B.Com. Bharathidasan, N 01)

Solution :

X	f	fX	$\lvert X-\bar{X}\rvert$ $\bar{X}=6$	$f\lvert X-\bar{X}\rvert$
2	1	2	4	4
4	4	16	2	8
6	6	36	0	0
8	4	32	2	8
10	1	10	4	4
Total	N = 16	ΣfX = 96	---	$\Sigma f\lvert X-\bar{X}\rvert$ = 24

Mean, $\bar{X} = \frac{\Sigma fX}{N}$

$= \frac{96}{16}$

$= 6$

Mean Deviation about Mean $= \frac{\Sigma f\lvert X-\bar{X}\rvert}{N}$

$= \frac{24}{16}$

$= 1.50$

Note : If the coefficient of M.D. about Mean is required,

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

$= \frac{1.50}{6}$

$= 0.2500$

Example 10 : Calculate all the three mean deviations and the corresponding coefficients of mean deviations :

Age (years) :	21	25	27	32	41	46	50	55
No.of Workers:	2	3	10	20	15	10	8	2

Solution :

Age X	No.of Workers f	fX	$\lvert X-\bar{X}\rvert$ $\bar{X}$=37.31	$f\lvert X-\bar{X}\rvert$	cf	$\lvert X-M\rvert$ M=36.5	$f\lvert X-M\rvert$	$\lvert X-Z\rvert$ Z=32	$f\lvert X-Z\rvert$
21	2	42	16.31	32.62	2	15.5	31.0	11	22
25	3	75	12.31	36.93	5	11.5	34.5	7	21
27	10	270	10.31	103.10	15	9.5	95.0	5	50
32	20	640	5.31	106.20	35	4.5	90.0	0	0
41	15	615	3.69	55.35	50	4.5	67.5	9	135
46	10	460	8.69	86.90	60	9.5	95.0	14	140
50	8	400	12.69	101.52	68	13.5	108.0	18	144
55	2	110	17.69	35.38	70	18.5	37.0	23	46
Total	N = 70	Σfx = 2612	--	$\Sigma f\lvert X-\bar{X}\rvert$ = 558.00	--	--	$\Sigma f\lvert X-M\rvert$ = 558.0	---	$\Sigma f\lvert X-Z\rvert$ = 558

Mean, $\bar{X} = \frac{\Sigma fX}{N} = \frac{2612}{70} = 37.31$

Mean Deviation about Mean $= \frac{\Sigma f\lvert X-\bar{X}\rvert}{N} = \frac{558.00}{70} = 7.97$

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

$= \frac{7.97}{37.31}$

$= 0.2136$

Position of Median is $\frac{N+1}{2} = \frac{70+1}{2} = 35.5$

Median (M) $= \frac{35^{th}\text{item} + 36^{th}\text{item}}{2}$

$= \frac{32+41}{2}$

$= 36.5$

Mean Deviation about Median $= \frac{\Sigma f\lvert X-M\rvert}{N}$

$= \frac{558}{70}$

$= 7.97$

Coefficient of M.D. about Median $= \frac{\text{M.D. about Median}}{\text{Median}}$

$= \frac{7.97}{36.5} = 0.2184$

Mode, $Z = 32$

Mean Deviation about Mode $= \frac{\Sigma f|X - Z|}{N} = \frac{558}{70} = 7.97$

Coefficient of M.D. about Mode $= \frac{\text{M.D. about Mode}}{\text{Mode}}$

$= \frac{7.97}{32} = 0.2491$

Continuous Series

The measure of central tendency, Mean or Median or Mode, is calculated first using the appropriate formula. The formulae considered in discrete series are used to find the necessary mean deviations by introducing m in the place of X.

Example 11: 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 Workers :	6	5	8	15	7	6	3

(C.A. Foundation, M 99)

Solution :

Marks	No.of Students f	Mid Value m	fm	$\lvert m - \overline{X} \rvert$ $\overline{X} = 33.40$	$f\lvert m - \overline{X} \rvert$
0-10	6	5	30	28.40	170.40
10-20	5	15	75	18.40	92.00
20-30	8	25	200	8.40	67.20
30-40	15	35	525	1.60	24.00
40-50	7	45	315	11.60	81.20
50-60	6	55	330	21.60	129.60
60-70	3	65	195	31.60	94.80
Total	N = 50	---	$\Sigma fm =$ 1670	---	$\Sigma f\lvert m - \overline{X} \rvert =$ 659.20

Mean, $\overline{X} = \frac{\Sigma fm}{N} = \frac{1670}{50} = 33.40$

Mean deviation about Mean $= \frac{\Sigma f|m - \overline{X}|}{N}$

$= \frac{659.20}{50}$

$= 13.18$

Example 12 : The following is the age distribution of 80 policy holders insured through an agent:

Age Group	Number of policy holders	Age Group	Number of policy holders
16-20	8	41-45	7
21-25	15	46-50	3
26-30	13	51-55	2
31-35	20	56-60	1
36-40	11		

Calculate mean deviation from the median.

(B.B.A. Bharathidasan, N 98)

Solution :

Age Group	No. of Policy Holders f	True Class Intervals	Cum. Freq. cf	Mid Value m	$\|m - M\|$ M = 31.5	$f\|m - M\|$
16-20	8	15.5-20.5	8	18	13.5	108.0
21-25	15	20.5-25.5	23	23	8.5	127.5
26-30	13	25.5-30.5	36	28	3.5	45.5
31-35	20	30.5-35.5	56 ←	33	1.5	30.0
36-40	11	35.5-40.5	67	38	6.5	71.5
41-45	7	40.5-45.5	74	43	11.5	80.5
46-50	3	45.5-50.5	77	48	16.5	49.5
51-55	2	50.5-55.5	79	53	21.5	43.0
56-60	1	55.5-60.5	80	58	26.5	26.5
Total	N = 80	-------	--	--	----	$\Sigma f\|m - M\| = 582.0$

N/2 = 80/2 = 40; 30.5 - 35.5 is the median class.

∴ L = 30.5; f = 20; i = 35.5 - 30.5 = 5; cf = 36.

∴ Median,

$$M = L + \left[\frac{i(N/2 - cf)}{f}\right]$$

$$= 30.5 + \left[\frac{5(40-36)}{20}\right]$$

$$= 30.5 + \left[\frac{5 \times 4}{20}\right]$$

$$= 30.5 + 1$$

$$= 31.50$$

Mean deviation about median $= \dfrac{\Sigma f|m - M|}{N}$

$$= \frac{582}{80}$$

$$= 7.28$$

Example 13 : Calculate

(i) Mean Deviation about Mode and

(ii) Coefficient of Mean Deviation about Mode

Mid Value :	2.5	7.5	12.5	17.5	22.5	27.5
Frequency :	19	28	50	22	10	7

Solution :

Mid Value m	Frequency f	Class Interval	$\|m - Z\|$ Z = 12.2	$f\|m - Z\|$
2.5	19	0- 5	9.7	184.3
7.5	28	5-10	4.7	131.6
12.5	50	10-15	0.3	15.0
17.5	22	15-20	5.3	116.6
22.5	10	20-25	10.3	103.0
27.5	7	25-30	15.3	107.1
Total	N = 136	-------	----	$\Sigma f\|m - Z\| =$ 657.6

(i) Common difference between successive mid values = 5.
Half of the difference = 5/2 = 2.5.

By adding 2.5 to the mid values, upper boundaries are obtained. By subtracting 2.5 from the mid values, lower boundaries are obtained.

Greatest Frequency = 50

(It is the greatest frequency density also)

∴ Modal Class 10–15; ∴ L = 10; D_1 = 50 − 28 = 22, D_2 = 50 − 22 = 28; i = 15 − 10 = 5.

$$\therefore \text{Mode,} \quad Z = L + \left[\frac{iD_1}{(D_1 + D_2)}\right]$$

$$= 10 + \left[\frac{5 \times 22}{(22 + 28)}\right]$$

$$= 10 + \left[\frac{110}{50}\right]$$

$$= 10 + 2.20$$

$$= 12.20$$

$$\text{Mean Deviation about Mode} = \frac{\Sigma f|m - Z|}{N}$$

$$= \frac{657.6}{136}$$

$$= 4.84$$

$$\text{(ii) Coefficient of M.D. about Mode} = \frac{\text{M.D. about Mode}}{\text{Mode}}$$

$$= \frac{4.84}{12.20}$$

$$= 0.3967$$

Short cut Method

Calculation of mean deviation is difficult when the measure of central tendency is a fraction. Calculations in such cases are made easy by a short cut method. The formulae are then as follows:

4. They are simple to understand and not difficult to calculate.

5. They do not vary much from sample to sample.

6. They provide choice. Among the three mean deviations, the one that is suitable to a particular situation can be used.

7. Formation of different distributions can be compared on the basis of a mean deviation.

Demerits

1. Omission of negative sign of deviations makes them non-algebraic. It is pointed out as a great drawback.

2. They could not be manipulated. Combined mean deviation could not be found.

3. It is not widely used in business or economics.

STANDARD DEVIATION

Definition : *Standard Deviation is the root mean square deviation of the values from their arithmetic mean.*

S.D. is the abbreviation and σ (read, sigma) is the symbol. Mean square deviation of the values from their A.M. is **Variance** and is denoted by σ^2. S.D.is the positive square root of variance. Karl Pearson introduced the concept of standard deviation in 1893. S.D. is also called ***root mean square deviation.*** It is a mathematical deficiency of mean deviation to ignore negative sign. Standard deviation possesses most of the desirable properties of a good measure of dispersion.It is the most widely used absolute measure of dispersion. The corresponding relative measure is **Coefficient of Variation. It is very popular and so extensively used as** raise a doubt whether there is any other relative measure of dispersion.

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \times 100$$

Formulae

Method	Individual Observations	Discrete Series	Continuous Series
1. Actual Mean	$\sqrt{\frac{\Sigma(X-\overline{X})^2}{N}}$	$\sqrt{\frac{\Sigma f(X-\overline{X})^2}{N}}$	$\sqrt{\frac{\Sigma f(m-\overline{X})^2}{N}}$
2.Direct Method	$\sqrt{\frac{\Sigma X^2}{N}-\left[\frac{\Sigma X}{N}\right]^2}$	$\sqrt{\frac{\Sigma f X^2}{N}-\left[\frac{\Sigma f X}{N}\right]^2}$	$\sqrt{\frac{\Sigma f m^2}{N}-\left[\frac{\Sigma f m}{N}\right]^2}$
3.Asumed Mean	$\sqrt{\frac{\Sigma d^2}{N}-\left[\frac{\Sigma d}{N}\right]^2}$	$\sqrt{\frac{\Sigma f d^2}{N}-\left[\frac{\Sigma f d}{N}\right]^2}$	$\sqrt{\frac{\Sigma f d^2}{N}-\left[\frac{\Sigma f d}{N}\right]^2}$
4.Step Deviation	$Cx\sqrt{\frac{\Sigma d'^2}{N}-\left[\frac{\Sigma d'}{N}\right]^2}$	$Cx\sqrt{\frac{\Sigma f d'^2}{N}-\left[\frac{\Sigma f d'}{N}\right]^2}$	$Cx\sqrt{\frac{\Sigma f d'^2}{N}-\left[\frac{\Sigma f d'}{N}\right]^2}$

Individual Observations

Method 1 : Deviations taken from Actual Mean.

Standard Deviation, $\sigma = \sqrt{\frac{\Sigma x^2}{N}}$ where $x = X - \overline{X}$. This method is simple when $X - \overline{X}$ values are integers.

Steps:

1. Form a table with the given values, X in the first column.

2. Find out the arithmetic mean, $\overline{X}$ $(= \Sigma X / N)$.

3. Find out the deviation of each value from the actual mean and call it x. i.e., find $x = X - \overline{X}$. Enter those values in the next column.

4. Find out the squares of the deviations of the values from the actual mean, i.e., find x^2. Enter those values in the next column.

5. Find out the mean of the squared deviations of the values from their arithmetic mean. i.e., find $\Sigma X^2 / N$.

6. Find out the square root of $\Sigma X^2 / N$. It is the standard deviation.

Example 16 : 77, 73, 75, 70, 72, 76, 75, 72, 74, 76. Give standard deviation for the numbers given above.

(B.Com. Bharathidasan, N 01)

Solution :

X	$x = X - \overline{X}$ $\overline{X} = 74$	x^2
77	3	9
73	-1	1
75	1	1
70	-4	16
72	-2	4
76	2	4
75	1	1
72	-2	4
74	0	0
76	2	4
$\Sigma X = 740$	$\Sigma(X - \overline{X}) = 0$	$\Sigma x^2 = 44$

Arithmetic Mean, $\overline{X} = \frac{\Sigma X}{N}$

$= \frac{740}{10}$

$= 74.$

Standard Deviation, $\sigma = \sqrt{\frac{\Sigma x^2}{N}}$

$= \sqrt{\frac{44}{10}}$

$= \sqrt{4.4}$

$= 2.10$

Method 2 : Direct Method. Without taking any deviations, the standard deviation can directly be calculated by the formula.

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$$

This method can be used for all kinds of data. This formula is used later for correcting the mistakes in the calculations.

Steps:

1. Form a table with the given values, X, in the first column.

2. Find the square of each X and write it in the next column under the title X^2.

3. Find the totals, ΣX and ΣX^2 and N, the number of values.

4. Substitute in the above formula and simplify.

Example 17: 10 students of B.Com. class of a College have obtained the following marks in Statistics out of 100 marks.

Calculate the standard deviation.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	5	10	20	25	40	42	45	48	70	80

(B.Com. (C.A.) Bharathiar, A 01)

Solution:

S. No.	Marks X	X^2
1	5	25
2	10	100
3	20	400
4	25	625
5	40	1600
6	42	1764
7	45	2025
8	48	2304
9	70	4900
10	80	6400
Total	$\Sigma X =$ 385	$\Sigma X^2 =$ 20143

Standard Deviation,

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$$

$$= \sqrt{\frac{20143}{10} - \left(\frac{385}{10}\right)^2}$$

$$= \sqrt{2014.3 - (38.5)^2}$$

$$= \sqrt{2014.30 - 1482.25}$$

$$= \sqrt{532.05}$$

$$= 23.07$$

Method 3 : Deviations taken from Assumed Mean.

This is same as the one followed in the calculation of arithmetic mean. But the formula is as follows:

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

$d = X - A$ is preferred when $X - \overline{X}$ are fractions.

Steps :

1. Form a table with the given values, X, in the first column.

2. Assume any value as A if it is not specified in a problem. It is preferable to assume a value in between the minimum value and the maximum value of X as A.

3. Find out the deviation of each value from the assumed mean A and call it d. i.e., find $d = X - A$ and write them in the next column.

4. Write the squares of the deviations, d^2, in the next column.

5. Find Σd and Σd^2 and identify N, the number of values. Substitute them in the above formula and simplify.

Example 18: For the data below, calculate standard deviation.

40 50 60 70 80 90 100

(B.Com. Bharathiar, A 02)

Solution :

X	d = X–A A = 70	d^2
40	–30	900
50	–20	400
60	–10	100
70	0	0
80	10	100
90	20	400
100	30	900
Total	$\Sigma d = 0$	$\Sigma d^2 = 2800$

Standard Deviation,

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

$$= \sqrt{\frac{2800}{7} - \left(\frac{0}{7}\right)^2}$$

$$= \sqrt{400 - 0^2}$$

$$= \sqrt{400}$$

$$= 20$$

Method 4 : Step Deviation Method. This is same as the one followed in the calculation of arithmetic mean. But the formula is as follows:

$$\text{Standard Deviation, } \sigma = C \times \sqrt{\frac{\Sigma d'^2}{N} - \left[\frac{\Sigma d'}{N}\right]^2}$$

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

This method is preferred when $X - \bar{X}$ are fractions and there is common difference between X.

Steps :

1. Form a table with the given values X, in the first column.
2. Choose A and C as mentioned under Arithmetic Mean.
3. Find out the step deviation corresponding to each X.

i.e., find $d' = \frac{X - A}{C}$ and write those values in the next column.

4. Write the squares of d'. i.e., d'^2, in the next column.

5. Find $\Sigma d'$ and $\Sigma d'^2$ and identify N, the number of values. Substitute them in the above formula and simplify.

Example 19: Given below are the marks obtained by 5 B.Com. students.

Roll No.	101	102	103	104	105
Marks	10	30	20	25	15

Calculate standard deviation.

(B.Com.Bharathiar, A 2K)

Solution:

Roll No.	Marks X	$d' = \frac{X-A}{C}$ A=20; C=5	d'^2
101	10	-2	4
102	30	2	4
103	20	0	0
104	25	1	1
105	15	-1	1
Total	-	$\Sigma d' = 0$	$\Sigma d'^2 = 10$

Standard Deviation,

$$\sigma = C \times \sqrt{\frac{\Sigma d'^2}{N} - \left[\frac{\Sigma d'}{N}\right]^2}$$

$$= 5 \times \sqrt{\frac{10}{5} - \left(\frac{0}{5}\right)^2}$$

$$= 5 \times \sqrt{2 - 0^2}$$

$$= 5 \times \sqrt{2 - 0}$$

$$= 5 \times \sqrt{2}$$

$$= 5 \times 1.4142$$

$$= 7.07$$

Note : A problem can be solved by any one method.

Discrete Series

Method 1 : Deviations taken from Actual Mean.

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma fx^2}{N}}$$

where $x = X - \overline{X}$ and $N = \Sigma f$

Steps :

1. Find out the arithmetic mean, $\overline{X}$.
2. Find out the deviation of each X from the actual mean and call it x. i.e., find $x = X - \overline{X}$.
3. Find out x^2 values.
4. Multiply each x^2 by the corresponding f and get fx^2
5. Find Σfx^2
6. Divide Σfx^2 by N and find the square root of the quotient. i.e., find $\sqrt{\frac{\Sigma fx^2}{N}}$

Example 20: Calculate the standard deviation of the following series.

X	6	9	12	15	18
f	7	12	13	10	8

Solution :

X	f	fX	$x = X - \bar{X}$ $\bar{X} = 12$	x^2	fx^2
6	7	42	–6	36	252
9	12	108	–3	9	108
12	13	156	0	0	0
15	10	150	3	9	90
18	8	144	6	36	288
Total	N= 50	ΣfX= 600	---	---	Σfx^2= 738

Arithmetic Mean, $\bar{X} = \frac{\Sigma f X}{N}$

$= \frac{600}{50}$

$= 12.00$

Standard Deviation, $\sigma = \sqrt{\frac{\Sigma f x^2}{N}}$

$= \sqrt{\frac{738}{50}}$

$= \sqrt{14.76}$

$= 3.84$

Method 2 : Direct Method. Under this method, the formula becomes the following.

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma f X^2}{N} - \left[\frac{\Sigma f X}{N}\right]^2}$$

Steps :

1. Form a table with the given values, X and the frequencies, f in the first two columns.

2. Multiply each X by the corresponding f to find f X. Write all such f X values in the next column.

3. Multiply each fX by the corresponding X to find fX^2 (It is not $(fX)^2$. That is, fX should not be squared) Write all such fX^2 values in the next column.

4. Find N (= Σf), ΣfX and ΣfX^2.

5. Substitute in the above formula and simplify.

Example 21 : Calculate the standard deviation.

No.of Goals Scored in a Match	(X)	0	1	2	3	4	5
No.of Matches	(f)	1	2	4	3	0	2

Solution :

X	f	fX	fX^2
0	1	0	0
1	2	2	2
2	4	8	16
3	3	9	27
4	0	0	0
5	2	10	50
Total	N =12	ΣfX =29	ΣfX^2 =95

Standard Deviation,

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left[\frac{\Sigma fX}{N}\right]^2}$$

$$= \sqrt{\frac{95}{12} - \left(\frac{29}{12}\right)^2}$$

$$= \sqrt{7.9167 - (2.4167)^2}$$

$$= \sqrt{7.9167 - 5.8404}$$

$$= \sqrt{2.0763}$$

$$= 1.44$$

Method 3 : Deviations taken from Assumed Mean. The formula is as follows:

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left[\frac{\Sigma fd}{N}\right]^2}$$

$d = x - A$ A - Assumed Mean $N = \Sigma f$

Steps:

1. Form a table with the given values, X and the frequencies, f in the first two columns.

2. Choose the value for A, assumed mean, if it is not specified.

3. Subtract A from each X and form the next column with $d = X - A$ values.

4. Multiply each d by the corresponding f and enter all such products in the next column under the title fd.

5. Multiply each fd by the corresponding d and enter all such products in the next column under the title fd^2. (These are not the squares of fd values)

6. Find N (= Σf), Σfd and Σfd².
7. Substitute in the above formula and simplify.

Example 22: Calculate Standard Deviation from the following data:

X	6	9	12	15	18
f	7	12	19	10	2

(B.B.M. Bharathiar, N 01)

Solution: Let X : 6, 9, 12, 15 and 18.

X	f	d=X–A A=12	fd	fd²
6	7	–6	–42	252
9	12	–3	–36	108
12	19	0	0	0
15	10	3	30	90
18	2	6	12	72
Total	N= 50	-	Σfd= –36	Σfd² =522

Standard Deviation, $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left[\frac{\Sigma fd}{N}\right]^2}$

$= \sqrt{\frac{522}{50} - \left(\frac{-36}{50}\right)^2}$

$= \sqrt{10.44 - (0.72)^2}$

$= \sqrt{10.4400 - 0.5184}$

$= \sqrt{9.9216}$

$= 3.15$

Method 4: Step Deviation Method. The following formula is used.

Standard Deviation, $\sigma = C \times \sqrt{\frac{\Sigma fd'^2}{N} - \left[\frac{\Sigma fd'}{N}\right]^2}$

$d' = \frac{X-A}{C}; N = \Sigma f$

Steps :

1. Form a table with the given values, X and the frequencies, f in the first two columns.

2. Choose A and C as mentioned under Arithmetic Mean.

3. Find out $d' = \frac{X-A}{C}$ corresponding to each X and enter them in the next column.

4. Multiply each d' by the corresponding f to get fd'. Enter them in the next column.

5. Multiply each fd' by the corresponding d' and get fd'^2. Enter them in the next column.

6. Find N $(=\Sigma f)$, $\Sigma fd'$ and $\Sigma fd'^2$

7. Substitute in the above formula and simplify.

Example 23: The weekly salaries of a group of employees are given in the following table. Find the mean and standard deviation of the salaries.

Salary (in Rs.)	:	75	80	85	90	95	100
No.of Persons	:	3	7	18	12	6	4

(B.Com. Bharathiar, N 2K)

Solution:

Salary (in Rs.) X	No.of Persons f	$d' = \frac{X-A}{C}$ A=85;C=5	fd'	fd'^2
75	3	−2	−6	12
80	7	−1	−7	7
85	18	0	0	0
90	12	1	12	12
95	6	2	12	24
100	4	3	12	36
Total	N= 50	-	$\Sigma fd'$ = 23	$\Sigma fd'^2$ = 91

Arithmetic Mean,
$$\bar{X} = A + \left(\frac{C\Sigma fd'}{N}\right)$$
$$= 85 + \left(\frac{5 \times 23}{50}\right)$$
$$= 85 + 2.3 = \text{Rs. } 87.30$$

Standard Deviation, $\sigma = C \times \sqrt{\frac{\Sigma f d'^2}{N} - \left[\frac{\Sigma f d'}{N}\right]^2}$

$= 5 \times \sqrt{\frac{91}{50} - \left(\frac{23}{50}\right)^2}$

$= 5 \times \sqrt{1.82 - (0.46)^2}$

$= 5 \times \sqrt{1.8200 - 0.2116}$

$= 5 \times \sqrt{1.6084}$

$= 5 \times 1.27$

= Rs. 6.35

Continuous Series

When X in the formulae for the calculation of standard deviation of discrete series is replaced by m the corresponding formulae for continuous series are obtained.

The calculations start with the mid values (m) of the class intervals and class frequencies (f). When less than or more than frequencies are given, class interval and class frequencies are to be found first.

Method 1 : Deviations taken from Actual Mean. The formula is as follows:

Standard Deviation, $\sigma = \sqrt{\frac{\Sigma f\left(m - \overline{X}\right)^2}{N}}$

Steps :

1. Form a table with class intervals and class frequencies in the first two columns.

2. Find the mid values (m) and write them in the next column.

3. Find the products of f and m and write them in the next column.

4. Find $\overline{X} = \Sigma fm/N$ where $N = \Sigma f$. $\overline{X}$ may be found by other formulae also.

5. Subtract $\overline{X}$ from each m. Enter the resulting $m - \overline{X}$ values in the next column.

6. Write $(m - \overline{X})^2$ in the next column.

7. Find $f(m-\overline{X})^2$ and write them in the next column.

8. Find $\Sigma f(m-\overline{X})^2$

9. Divide $\Sigma f(m-\overline{X})^2$ by N and take the square root to get the standard deviation.

Example 24: Find the standard deviation.

Class Interval	0-10	10-20	20-30	30-40	40-50	Total
Frequency	2	5	9	3	1	20

Solution:

Class Interval	Frequency f	Mid Value m	fm	$m-\overline{X}$ $\overline{X}=23$	$(m-\overline{X})^2$	$f(m-\overline{X})^2$
0-10	2	5	10	−18	324	648
10-20	5	15	75	− 8	64	320
20-30	9	25	225	2	4	36
30-40	3	35	105	12	144	432
40-50	1	45	45	22	484	484
Total	N= 20	-	$\Sigma fm=$ 460	-	-	$\Sigma f(m-\overline{X})^2=$ 1920

Arithmetic Mean, $\overline{X} = \frac{\Sigma fm}{N}$

$= \frac{460}{20}$

$= 23$

Standard Deviation, $\sigma = \sqrt{\frac{\Sigma f\left(m-\overline{X}\right)^2}{N}}$

$= \sqrt{\frac{1920}{20}}$

$= \sqrt{96}$

$= 9.80$

Method 2 : Direct Method. Under this method, the form of the formula is as follows:

$$\text{Standard Deviation}, \sigma = \sqrt{\frac{\Sigma fm^2}{N} - \left(\frac{fm}{N}\right)^2}$$

Steps :

1. Form a table with class intervals and frequencies in the first two columns.

2. Find the mid values (m) and write them in the next column.

3. Find the products of f and m and write those fm values in the next column.

4. Find the products of m and fm and write those fm^2 values in the next column.

5. Find N (= Σf), Σfm and Σfm^2

6. Substitute in the formula and simplify.

Example 25: The following data were obtained while observing the life span of a few neon lights of a company. Calculate S.D.

Life Span (Years) :	4-6	6-8	8-10	10-12	12-14	Total
No. of Neon Lights :	10	17	32	21	20	100

Solution:

Life Span (Years)	No.of Neon Lights (f)	Mid Value(m)	fm	fm^2
4- 6	10	5	50	250
6- 8	17	7	119	833
8-10	32	9	288	2592
10-12	21	11	231	2541
12-14	20	13	260	3380
Total	N= 100	-	Σfm = 948	Σfm^2 = 9596

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma fm^2}{N} - \left(\frac{\Sigma fm}{N}\right)^2}$$

$$= \sqrt{\frac{9596}{100} - \left(\frac{948}{100}\right)^2}$$

$$= \sqrt{95.96 - (9.48)^2}$$

$$= \sqrt{95.9600 - 89.8704}$$

$$= \sqrt{6.0896} = 2.47.$$

Method 3: Deviations taken from Assumed Mean. The following formula is used:

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

$d = m - A$ and $N = \Sigma f$.

Steps:

1. Form a table with class intervals and frequencies in the first two columns.

2. Find the mid values (m) of the class intervals and write them in the next column.

3. Choose a suitable value for A, the assumed mean. Subtract A from each m. Write the d (= m–A) values in the next column.

4. Multiply d by the corresponding f and write the products fd in the next column.

5. Multiply fd by the corresponding d and write the products fd^2 in the next column.

6. Find N (= Σf), Σfd and Σfd^2.

7. Substitute in the above formula and simplify.

Example 26: Calculate the standard deviation of the following series.

No.of Students in 00 (Below)	2	6	10	14	18	22	26
No.of Colleges	0	7	19	42	61	72	80

Solution:

No.of Students in 00 (Below)	No.of Colleges	No.of Students in 00	No.of Colleges f	Mid Value m	d = m–A A=12	fd	fd^2
2	0	2- 6	7	4	–8	–56	448
6	7	6-10	12	8	–4	–48	192
10	19	10-14	23	12	0	0	0
14	42	14-18	19	16	4	76	304
18	61	18-22	11	20	8	88	704
22	72	22-26	8	24	12	96	1152
26	80						
Total	---	---	N= 80	---	---	Σfd =156	Σfd^2= 2800

Standard Deviation, σ $= \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

$= \sqrt{\frac{2800}{80} - \left(\frac{156}{80}\right)^2}$

$= \sqrt{35 - (1.95)^2}$

$= \sqrt{35.0000 - 3.8025}$

$= \sqrt{31.1975}$

$= 5.59$

Method 4 : Step Deviation Method. The formula used in this method is as follows:

Standard Deviation, σ $= C \times \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$

$d' = \frac{m - A}{C}$; $N = \Sigma f$

Steps :

1. Form a table with class intervals and frequencies in the first two columns.

2. Find the mid values (m) of the class intervals and write them in the next column.

3. Choose values for A and C. Calculate $d' = \frac{m - A}{C}$ corresponding to each m. Write the d' values in the next column.

4. Multiply f and d' in pairs and enter fd' values in the next column.

5. Multiply fd' values by the respective d' and enter those fd'^2 values in the next column. (They are not the squares of fd').

6. Find N ($= \Sigma f$), $\Sigma fd'$ and $\Sigma fd'^2$.

7. Substitute in the above formula and simplify.

Example 27: Calculate the standard deviation of the following frequency distribution.

Annual Profit (Rs.Crores):	20-40	40-60	60-80	80-100
No.of Banks:	10	14	25	48

Annual Profit (Rs.Crores):	100-120	120-140	140-160
No.of Banks:	33	24	16

Solution:

Annual Profit (Rs.Crores)	No.of Banks	Mid Value m	$d' = \frac{m-A}{C}$ A=90;C=20	fd'	fd'²
20- 40	10	30	-3	-30	90
40- 60	14	50	-2	-28	56
60- 80	25	70	-1	-25	25
80-100	48	90	0	0	0
100-120	33	110	1	33	33
120-140	24	130	2	48	96
140-160	16	150	3	48	144
Total	N = 170	-	-	Σ fd' =46	Σ fd'² =444

Standard Deviation, σ $= C \times \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$

$$= 20 \times \sqrt{\frac{444}{170} - \left(\frac{46}{170}\right)^2}$$

$$= 20 \times \sqrt{2.6118 - (0.2706)^2}$$

$$= 20 \times \sqrt{2.6118 - 0.0732}$$

$$= 20 \times \sqrt{2.5386}$$

$$= 20 \times 1.5933$$

= Rs. 31.87 Crores.

Combined Standard Deviation

When two or three groups merge, the mean and standard deviation of the combined group are calculated as follows.

Case 1. Merger of Two Groups

	Size	Mean	S.D.
Group I	N_1	$\overline{X}_1$	σ_1
Group II	N_2	$\overline{X}_2$	σ_2

COEFFICIENT OF VARIATION

Definition:

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \times 100$$

C.V. is the abbreviation. ∴ $C.V. = \frac{S.D.}{A.M.} \times 100$

$$= \frac{\sigma}{\bar{X}} \times 100$$

Karl Pearson gave this definition. Like all other relative measures of dispersion, it is a pure number. All relative measures of dispersion are free from units of measurements such as Kg., metre, litre, etc. The variations in two or more series (groups or sets of data) are compared on the basis of a relative measure of dispersion.

For example, an Indian may have different income at various periods of time. His income is quoted in Rs. An American may also have different income. His income is quoted in Dollars. The variations in their incomes can be compared by using any relative measure of dispersion.

Coefficient of variation is the most widely used relative measure of dispersion. It is based on the best absolute measure of dispersion and the best measure of central tendency. It is a percentage. While comparing two or more groups, the group which has less coefficient of variation is less variable or more consistent or more stable or more uniform or more homogeneous. On the contrary, the group which has more coefficient of variation is more variable or less consistent or less stable or less uniform or less homogeneous.

Example 40: Calculate the coefficient of variation of the following:

40 41 45 49 50 51 55 59 60 60

S.D., $\sigma = C \times \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$

$= 10 \times \sqrt{\frac{140}{80} - \left(\frac{12}{80}\right)^2}$

$= 10 \times \sqrt{1.75 - (0.15)^2}$

$= 10 \times \sqrt{1.7500 - 0.0225}$

$= 10 \times \sqrt{1.7275}$

$= 10 \times 1.3143$

$= 13.14$

C.V. $= \frac{\sigma}{\overline{X}} \times 100$

$= \frac{13.14}{26.50} \times 100 = 49.58$

Example 42: The means and standard deviation values for the number of runs of two players A and B are 55; 65 and 4.2; 7.8 respectively. Who is the more consistent player?

(B.Com. (C.A.) Bharathiar, A 02)

Solution: Given:

Player A: Mean = 55; Standard Deviation = 4.2

Player B: Mean = 65; Standard Deviation = 7.8

∴ Coefficient of Variation of Player A $= \frac{\text{S.D.}}{\text{A.M.}} \times 100$

$= \frac{4.2}{55} \times 100 = 7.64$

Coefficient of Variation of Player B $= \frac{7.8}{65} \times 100 = 12.00$

Coefficient of Variation of Player A is less. Therefore, Player A is the more consistent player.

Example 43: From the following data, find

(i) which firm pays more amount as monthly wages

(ii) which firm has greater variability in individual wages and

(iii) what are the mean monthly wage and the standard deviation of wages if the two firms merge.

	Firm I	Firm II
Number of Workers	100	200
Mean monthly wage(Rs.000)	7	8
S.D. of individual wages (Rs.000)	2	2.5

Solution: Given:

$N_1 = 100$ $\qquad$ $N_2 = 200$

$\overline{X}_1 = 7$ $\qquad$ $\overline{X}_2 = 8$

$\sigma_1 = 2$ $\qquad$ $\sigma_2 = 2.5$

(i) Total wage = Number × Mean

Total wage in firmI $= N_1\overline{X}_1 = 100 \times 7 =$ Rs. 700 thousands.

Total wage in firmII $= N_2\overline{X}_2 = 200 \times 8 =$ Rs.1,600 thousands.

Hence, firm II pays more amount as monthly wages.

(ii) Coefficient of Variation $= \frac{\sigma}{\overline{X}} \times 100$

Coefficient of Variation of firmI $= \frac{\sigma_1}{\overline{X}_1} \times 100$

$= \frac{2}{7} \times 100$

$= 28.57$

Coefficient of Variation of firmII $= \frac{\sigma_2}{\overline{X}_2} \times 100$

$= \frac{2.5}{8} \times 100$

$= 31.25$

Coefficient of Variation of firmII is more. Hence, there is greater variability in individual wages in firm II.

(iii) Combined mean monthly wage,

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

$$= \frac{100 \times 7 + 200 \times 8}{100 + 200}$$

$$= \frac{700 + 1600}{300}$$

$$= \frac{2300}{300}$$

= Rs. 7.67 (thousands)

Combined S.D., $\sigma_{12} = \sqrt{\dfrac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$

$$= \sqrt{\frac{100 \times 2^2 + 200 \times (2.5)^2 + 100 \times (-0.67)^2 + 200 \times (0.33)^2}{100 + 200}}$$

$\because d_1 = \bar{X}_1 - \bar{X}_{12} = 7.00 - 7.67 = -0.67$

$d_2 = \bar{X}_2 - \bar{X}_{12} = 8.00 - 7.67 = 0.33$

$$= \sqrt{\frac{400.00 + 1250.00 + 44.89 + 21.78}{300}}$$

$$= \sqrt{\frac{1716.67}{300}}$$

$$= \sqrt{5.7222}$$

$$= 2.39$$

Example 44: From the following price of gold in a week, find the city in which the price was more stable.

Day	Mon	Tues	Wed	Thu	Fri	Sat
City A	498	500	505	504	502	509
City B	500	505	502	498	496	505

Solution :

City A X_1	$X_1 - \bar{X}_1$ $\bar{X}_1 = 503$	$(X_1 - \bar{X}_1)^2$	City B X_2	$X_2 - \bar{X}_2$ $\bar{X}_2 = 501$	$(X_2 - \bar{X}_2)^2$
498	−5	25	500	−1	1
500	−3	9	505	4	16
505	2	4	502	1	1
504	1	1	498	−3	9
502	−1	1	496	−5	25
509	6	36	505	4	16
$\Sigma X_1 = 3018$	-	$\Sigma(X_1 - \bar{X}_1)^2 = 76$	$\Sigma X_2 = 3006$	-	$\Sigma(X_2 - \bar{X}_2)^2 = 68$

City A

$$\overline{X}_1 = \frac{\Sigma X_1}{N_1}$$

$$= \frac{3018}{6}$$

$$= \text{Rs. } 503$$

$$\sigma_1 = \sqrt{\frac{\Sigma(X_1 - \overline{X}_1)^2}{N_1}}$$

$$= \sqrt{\frac{76}{6}}$$

$$= \sqrt{12.6667}$$

$$= \text{Rs.} 3.56$$

$$C.V._1 = \frac{\sigma_1}{\overline{\overline{X}}_1} \times 100$$

$$= \frac{3.56}{503} \times 100$$

$$= 0.71$$

City B

$$\overline{X}_2 = \frac{\Sigma X_2}{N_2}$$

$$= \frac{3006}{6}$$

$$= \text{Rs. } 501$$

$$\sigma_2 = \sqrt{\frac{\Sigma(X_2 - \overline{X}_2)^2}{N_2}}$$

$$= \sqrt{\frac{68}{6}}$$

$$= \sqrt{11.3333}$$

$$= \text{Rs. } 3.37$$

$$C.V._2 = \frac{\sigma_2}{\overline{\overline{X}}_2} \times 100$$

$$= \frac{3.37}{501} \times 100$$

$$= 0.67$$

Coefficient of Variation of price in City B is less. Hence, the price was more stable in City B.

Example 45: Goals scored by two teams A and B in a series of football matches were observed as follows:

No.of Goals Scored in a Match	No.of Matches Team A	Team B
0	5	4
1	7	5
2	5	5
3	3	4
4	2	3
5	3	3

Which team, A or B, may be considered as a more consistent

Solution: Goals(X) are common. No.of matches(f) differ between the teams.

Goals	Matches		Team A		Team B	
	Team A	Team B				
X	f_1	f_2	f_1X	f_1X^2	f_2X	f_2X^2
0	5	4	0	0	0	0
1	7	5	7	7	5	5
2	5	5	10	20	10	20
3	3	4	9	27	12	36
4	2	3	8	32	12	48
5	3	3	15	75	15	75
Total	N_1= 25	N_2= 24	Σf_1X =49	Σf_1X^2 =161	Σf_2X =54	Σf_2X^2 =184

Team A

$$\text{Mean, } \overline{X}_1 = \frac{\Sigma f_1 X}{N_1} = \frac{49}{25} = 1.96$$

$$\text{S.D.,}\sigma_2 = \sqrt{\frac{\Sigma f_1X^2}{N_1} - \left(\frac{\Sigma f_1X}{N_1}\right)^2} = \sqrt{\frac{161}{25} - \left(\frac{49}{25}\right)^2} = \sqrt{6.44 - (1.96)^2} = \sqrt{6.4400 - 3.8416} = \sqrt{2.5984} = 1.61$$

$$\text{C.V.}_1 = \frac{\sigma_1}{\overline{X}_1} \times 100 = \frac{1.61}{1.96} \times 100 = 82.14$$

Team B

$$\text{Mean, } \overline{X}_2 = \frac{\Sigma f_2 X}{N_2} = \frac{54}{24} = 2.25$$

$$\text{S.D.,}\sigma_2 = \sqrt{\frac{\Sigma f_2X^2}{N_2} - \left(\frac{\Sigma f_2X}{N_2}\right)^2} = \sqrt{\frac{184}{24} - \left(\frac{54}{24}\right)^2} = \sqrt{7.6667 - (2.25)^2} = \sqrt{7.6667 - 5.0625} = \sqrt{2.6042} = 1.61$$

$$\text{C.V.}_2 = \frac{\sigma_2}{\overline{X}_2} \times 100 = \frac{1.61}{2.25} \times 100 = 71.56$$

Coefficient of variation of Team B is less. Hence, Team B is the more consistent team.

Example 46: The marks in Business Mathematics of two sections of students of a College are given below. Find in which section, the marks are more variable.

Marks	20-30	30-40	40-50	50-60	60-70
No. of Students: Section A	5	13	24	5	3
No. of Students: Section B	7	14	25	12	2

Solution:

Marks	No.of Students Sec.A f_1	No.of Students Sec.B f_2	Mid Value m	$d'=\frac{m-A}{C}$ A=45; C=10	d'^2	Section A f_1d'	Section A $f_1d'^2$	Section B f_2d'	Section B $f_2d'^2$
20-30	5	7	25	−2	4	−10	20	−14	28
30-40	13	14	35	−1	1	−13	13	−14	14
40-50	24	25	45	0	0	0	0	0	0
50-60	5	12	55	1	1	5	5	12	12
60-70	3	2	65	2	4	6	12	4	8
Total	$N_1=$ 50	$N_2=$ 60	-	-	---	$\Sigma f_1d'$ =−12	$\Sigma f_1d'^2$ =50	$\Sigma f_2d'$ =−12	$\Sigma f_2d'^2$ =62.

Section A

$$\text{Mean, } \overline{X}_1 = A+\left(\frac{C\Sigma f_1d'}{N_1}\right)$$

$$= 45+\left(\frac{10\times -12}{50}\right)$$

$$= 45-2.40$$

$$= 42.60$$

$$\text{S.D., } \sigma_1 = C\times\sqrt{\frac{\Sigma f_1d'^2}{N_1}-\left(\frac{\Sigma f_1d'}{N_1}\right)^2}$$

$$= 10\times\sqrt{\frac{50}{50}-\left(\frac{-12}{50}\right)^2}$$

$$= 10\times\sqrt{1-(-0.24)^2}$$

$$= 10\times\sqrt{1.0000-0.0576}$$

$$= 10\times\sqrt{0.9424}$$

$$= 10\times 0.9708 = 9.71$$

Section B

$$\text{Mean, } \overline{X}_2 = A+\left(\frac{C\Sigma f_2d'}{N_2}\right)$$

$$= 45+\left(\frac{10\times -12}{60}\right)$$

$$= 45-2.00$$

$$= 43.00$$

$$\text{S.D., } \sigma_2 = C\times\sqrt{\frac{\Sigma f_2d'^2}{N_2}-\left(\frac{\Sigma f_2d'}{N_2}\right)^2}$$

$$= 10\times\sqrt{\frac{62}{60}-\left(\frac{-12}{60}\right)^2}$$

$$= 10\times\sqrt{1.0333-(-0.2)^2}$$

$$= 10\times\sqrt{1.0333-0.0400}$$

$$= 10\times\sqrt{0.9933}$$

$$= 10\times 0.9966 = 9.97$$

$$C.V. = \frac{\sigma_1}{\overline{X}_1} \times 100 \qquad C.V. = \frac{\sigma_2}{\overline{X}_2} \times 100$$

$$= \frac{9.71}{42.60} \times 100 \qquad = \frac{9.97}{43.00} \times 100$$

$$= 22.79 \qquad = 23.19$$

Coefficient of Variation of section Bis more. Hence, in section B, the marks are more variable.

VARIANCE

Definition: *Variance is the mean square deviation of the values from their arithmetic mean.*

σ^2 (read, sigma square) is the symbol. Standard deviation is the positive square root of variance and is denoted by σ. The term of variance was introduced by R.A.Fisher in the year 1913. It is used much in sampling, analysis of variance, etc. In analysis of variance, total variation is split into a few components. Each component is ascribable to one factor of variation. The significance of the variation is then tested.

Formulae :

These formulae can be compared with those under standard deviation.

Method	Individual Observations	Discrete Series	Continuous Series
1.Actual Mean			
	$\frac{\Sigma(X-\overline{X})^2}{N}$	$\frac{\Sigma f(X-\overline{X})^2}{N}$	$\frac{\Sigma f(m-\overline{X})^2}{N}$
2.Direct Method			
	$\frac{\Sigma X^2}{N}-\left(\frac{\Sigma X}{N}\right)^2$	$\frac{\Sigma f X^2}{N}-\left(\frac{\Sigma f X}{N}\right)^2$	$\frac{\Sigma fm^2}{N}-\left(\frac{\Sigma fm}{N}\right)^2$
3.Assumed Mean			
	$\frac{\Sigma d^2}{N}-\left(\frac{\Sigma d}{N}\right)^2$	$\frac{\Sigma fd^2}{N}-\left(\frac{\Sigma fd}{N}\right)^2$	$\frac{\Sigma fd^2}{N}-\left(\frac{\Sigma fd}{N}\right)^2$
4. Step Deviation			
	$C^2\left[\frac{\Sigma d'^2}{N}-\left(\frac{\Sigma d'}{N}\right)^2\right]$	$C^2\left[\frac{\Sigma fd'^2}{N}-\left(\frac{\Sigma fd'}{N}\right)^2\right]$	$C^2\left[\frac{\Sigma fd'^2}{N}-\left(\frac{\Sigma fd'}{N}\right)^2\right]$

9

SKEWNESS, MOMENTS, KURTOSIS

In the previous two chapters two aspects of a series are considered. Measures of central tendency provide a central value. It is a representative value of a series. Measures of dispersion furnish the degree of vartiation in a series. They indirectly describe the extent to which an average is typical of a series. The third aspect of a series is skewness. Whether or not the individual items are equally distributed on either side of mode is of interest. The fourth and the final aspect is kurtosis. Under that the peakedness of the frequency curve of a series is examined. All these four aspects can be studied in terms of moments.

SKEWNESS

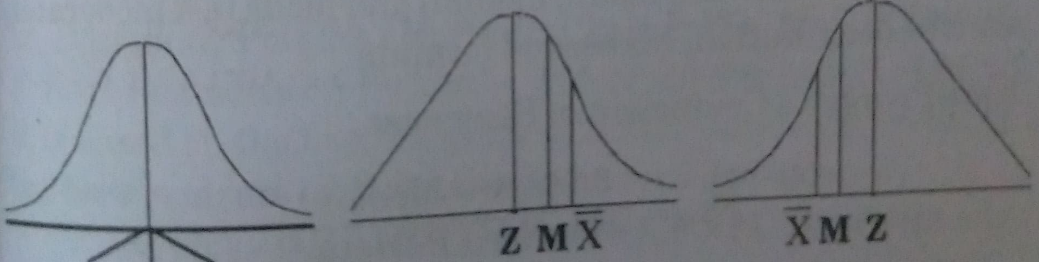
Definition: *Skewness is the degree of asymmety, or departure from symmetry, of a distribution.* (Murray R. Spiegel : Theory and Problems of Statistics - Page 90)

Consider the following three continuous series with common mid values.

Coefficients of skewness are calculated later. But the values of averages and quartiles are presented in a tabular form now. The frequency curves are also drawn at the bottom of the table.

How a symmetric curve looks, what is the relation between the averages in such a case and how are the quartiles related then are a few questions for which the answers are being found. These aspects of skewed curves are also known.

Mid Value	Series A (Frequency)	Series B (Frequency)	Series C (Frequency)
20	1	1	1
30	12	12	27
40	55	91	40
50	91	55	55
60	55	40	91
70	12	27	12
80	1	1	1
Nature of Skewness	Symmetry No Skewness	Asymmetry (Positive Skewness)	Asymmetry (Negative Skewness)
Averages	$\bar{X}=M=Z$ 50=50=50	$\bar{X}>M>Z$ 49.07>46.73>41.87	$\bar{X}<M<Z$ 50.93<53.27<58.13
Quartiles	$Q_3-M=M-Q_1$ $Q_1=42.95$ M=50.00 $Q_3=57.05$	$Q_3-M>M-Q_1$ $Q_1=39.81$ M=46.73 $Q_3=57.81$	$Q_3-M<M-Q_1$ $Q_1=42.19$ M=53.27 $Q_3=60.19$
Nature of the Curve	Bell shaped	Longer tail in the right side (Skewed to the right)	Longer tail in the left side (Skewed to the left)



$\bar{X}=M=Z$ Symmetric Curve | Posively Skewed | Negatively Skewed

(Refer to : Example 14 | Example 15 | Example 16)

Fig.1. Curves with Different Kinds of Skewness.

Absolute Measures: The following are the two absolute measures of skewness. They are of no practiccal use. They

Example 5 : Calculate Karl Pearson's coefficient of skewness for the following data:

25 15 23 40 27 25 23 25 20

(B.B.M. Bharathiar, A 02)

Solution:

X	X^2
25	625
15	225
23	529
40	1600
27	729
25	625
23	529
25	625
20	400
$\Sigma X =$ 223	$\Sigma X^2 =$ 5887

Mean, $\overline{X} = \frac{\Sigma X}{N}$

$= \frac{223}{9}$

$= 24.78$

S.D., $\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$

$= \sqrt{\frac{5887}{9} - (24.78)^2}$

$= \sqrt{654.1111 - 614.0484}$

$= \sqrt{40.0627}$

$= 6.33$

Mode, Z $= 25.00$

Karl Pearson's coefficient of skewness,

$$Sk_P = \frac{\overline{X} - Z}{\sigma} = \frac{24.78 - 25.00}{6.33} = -\frac{0.22}{6.33} = -0.0348$$

Example 6: Calculate Karl-Pearson's coefficient of skewness for the following data:

Wage per Item Rs.	12	15	20	25	30	40	50
Number of Items	10	25	40	70	32	13	10

Solution :

Wage per Item (Rs.)(X)	No.of Items(f)	d=X–A A=25	fd	fd^2
12	10	–13	–130	1690
15	25	–10	–250	2500
20	40	–5	–200	1000
25	70	0	0	0
30	32	5	160	800
40	13	15	195	2925
50	10	25	250	6250
Total	N= 200	---	Σfd = 25	Σfd^2 =15165

Mean, $\overline{X} = A + \left(\frac{\sum fd}{N}\right) = 25 + \left(\frac{25}{200}\right) = 25+0.125 = \text{Rs. } 25.13$

S.D., $\sigma = \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} = \sqrt{\frac{15165}{200} - \left(\frac{25}{200}\right)^2} = 8.71$

Mode, Z = 25

Karl Pearson's coefficient of skewness, $Sk_p = \frac{\overline{X} - Z}{\sigma}$

$$= \frac{25.13 - 25}{8.71}$$

$$= \frac{0.13}{8.71}$$

$$= 0.0149$$

Example 7: Calculate coefficient of skewness by Karl Pearson's method:

Profit (Rs. lakhs)	10-20	20-30	30-40	40-50	50-60
No. of Companies	18	20	30	22	10

(B.Com. Madras, A 99)

Solution :

Profit (Rs.lakhs)	No.of Companies f	Mid Value m	$d' = \frac{m-A}{C}$ A=35; C=10	fd'	fd'^2
10-20	18	15	–2	–36	72
20-30	20	25	–1	–20	20
30-40	30	35	0	0	0
40-50	22	45	1	22	22
50-60	10	55	2	20	40
Total	N=100	---	---	$\sum fd' =$ –14	$\sum fd'^2 =$ 154

A.M., $\overline{X} = A + \left(\frac{C\sum fd'}{N}\right) = 35 + \left(\frac{10 \times -14}{100}\right) = 35 - 1.40 = 33.60$

S.D., $\sigma = C \times \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} = 10 \times \sqrt{\frac{154}{100} - \left(\frac{-14}{100}\right)^2} = 12.33$

Mode, $Z = L + \left[\frac{iD_1}{(D_1 + D_2)}\right]$

30 is the greatest frequency and 30-40 is the modal class interval. $\therefore L = 30; i = 40-30=10; D_1 = 30-20=10; D_2 = 30-22=8$

$\therefore \quad Z = 30 + \left[\frac{10 \times 10}{(10+8)}\right]$

$= 30 + \left[\frac{100}{18}\right]$

$= 30 + 5.56$

$= 35.56$

Karl Pearson's coefficient of skewness, $Sk_P = \frac{\overline{X} - Z}{\sigma}$

$= \frac{33.60 - 35.56}{12.33}$

$= \frac{-1.96}{12.33}$

$= -0.1590$

Example 8: Calculate Karl Pearson's coefficient of skewness from the following data:

Weight(lbs)	No.of Students	Weight (lbs)	No.of Students
90-100	4	140-150	19
100-110	2	150-160	10
110-120	18	160-170	3
120-130	22	170-180	2
130-140	21		

(B.Com. (C.A.) Bharathiar, A 01

Solution :

Weight (lbs)	No.of Students f	Mid Value m	$d' = \frac{m-A}{C}$ A=135; C=10	fd'	fd'^2	Cumulative frequency cf
90-100	4	95	–4	–16	64	4
100-110	2	105	–3	– 6	18	6
110-120	18	115	–2	–36	72	24
120-130	22	125	–1	–22	22	46
130-140	21	135	0	0	0	67
140-150	19	145	1	19	19	86
150-160	10	155	2	20	40	96
160-170	3	165	3	9	27	99
170-180	2	175	4	8	32	101
Total	N= 101	--	--	$\sum fd'$ =–24	$\sum fd'^2$ =294	--

$$\text{A.M., } \overline{X} = A + \left(\frac{C\sum fd'}{N}\right) = 135 + \left(\frac{10 \times -24}{101}\right) = 132.62$$

$$\text{S.D., } \sigma = C \times \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} = 10 \times \sqrt{\frac{294}{101} - \left(\frac{-24}{101}\right)^2} = 16.90$$

Grouping and analysis tables indicate that the class intervals 120-130 and 130-140 have equal frequency densities. Hence, mode is ill defined.

N/2 = 101/2 = 50.5; ∴ 130-140 is the median class interval.

L = 130; i = 140–130 =10; f = 21; cf = 46

$$\therefore \text{Median,} \quad M = L + \left[\frac{i(N/2 - cf)}{f}\right]$$

$$= 130 + \left[\frac{10(50.5 - 46)}{21}\right]$$

$$= 130 + \left[\frac{10 \times 4.5}{21}\right]$$

$$= 130 + 2.14$$

$$= 132.14$$

Karl Pearson's coefficient of skewness,

$$Sk_P = \frac{3(\overline{X} - M)}{\sigma}$$

$$= \frac{3(132.62 - 132.14)}{16.90}$$

$$= \frac{3 \times 0.48}{16.90} = 0.0852$$

BOWLEY'S COEFFICIENT

Example 9: Compare the skewness of A and B.

	Q_1	M	Q_3
Series A	40	60	80
Series B	62.85	65.25	72.15

Solution : Series A

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

$$= \frac{80 + 40 - 2 \times 60}{80 - 40}$$

$$= \frac{0}{40} = 0$$

Series B

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

$$= \frac{72.15 + 62.85 - 2 \times 65.25}{72.15 - 62.85}$$

$$= \frac{4.50}{9.30} = 0.4839$$

In series A, there is no skewness. In series B, there is moderate positive skewness.

Example 10: Calculate Bowley's coefficient of skewness

No.of Children per Family :	0	1	2	3	4	5	6
No. of Families :	7	10	16	25	18	11	8

(B.Com. Madras, 001)

Solution :

No. of Children per Family(X)	No. of Families f	Cum. Freq. cf
0	7	7
1	10	17
2	16	33 ←
3	25	58 ←
4	18	76 ←
5	11	87
6	8	95
Total	N=95	---

Position of Q_1 is $\frac{N+1}{4}$

$$= \frac{95+1}{4} = 24$$

$\therefore$ $Q_1 = 2$

Position of M is $\frac{N+1}{2}$

$$= \frac{95+1}{2} = 48$$

$\therefore$ $M = 3$

Position of Q_3 is $3\left(\frac{N+1}{4}\right) = 3 \times 24 = 72$

$\therefore Q_3 = 4$

Bowley's coefficient of skewness, $Sk_B = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

$= \frac{4 + 2 - 2 \times 3}{4 - 2}$

$= \frac{0}{2}$

$= 0$

Example 11: The following table gives the distribution of weekly wages of 500 workers in a factory.

Weekly Wages Rs.	Below 200	200-250	250-300
No. of Workers	10	25	145
Weekly Wages Rs.	300-350	350-400	400 & above
No. of Workers	220	70	30

(a) Obtain the limits of income of central 50% of the observed workers.

(b) Calculate Bowley's coefficient of skewness.

(B.Com. (C.A.) Periyar, N 01)

Solution:

Weekly Wages (Rs.)	No. of Workers f	Cum. Freq. cf
Below 200	10	10
200-250	25	35
250-300	145	180 ←
300-350	220	400 ←
350-400	70	470
400 and above	30	500
Total	N=500	---

(a) $N/4 = 500/4 = 125$. $\therefore$ 250–300 is the Q_1 class.

$\therefore L_1 = 250; i_1 = 300-250 = 50;$

$f_1 = 145; cf_1 = 35$

$$Q_1 = L_1 + \left[\frac{i_1(N/4 - cf_1)}{f_1}\right]$$

$$= 250 + \left[\frac{50(125-35)}{145}\right]$$

$$= 250 + \left[\frac{50 \times 90}{145}\right]$$

$$= 250 + 31.03 = \text{Rs. } 281.03$$

$3N/4 = 3 \times 125 = 375$. $\therefore$ 300–350 is the Q_3 class.

$\therefore L_3 = 300; i_3 = 350-300 = 50; f_3 = 220; cf_3 = 180.$

$$Q_3 = L_3 + \left[\frac{i_3(3N/4 - cf_3)}{f_3}\right]$$

$$= 300 + \left[\frac{50(375 - 180)}{220}\right]$$

$$= 300 + \left[\frac{50 \times 195}{220}\right]$$

$$= 300 + 44.32$$

$$= \text{Rs. } 344.32$$

∴ the limits of weekly wages of central 50% of the observed workers are Rs. 281.03 and Rs. 344.32

(b) $N/2 = 500/2 = 250$. ∴ 300-350 is the median class.

∴ $L = 300$; $i = 350-300=50$; $f = 220$; $cf = 180$.

$$M = L + \left[\frac{i(N/2 - cf)}{f}\right]$$

$$= 300 + \left[\frac{50(250 - 180)}{220}\right]$$

$$= 300 + \left[\frac{50 \times 70}{220}\right]$$

$$= 300 + 15.91$$

$$= \text{Rs. } 315.91$$

Bowley's coefficient of skewness,

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

$$= \frac{344.32 + 281.03 - 2 \times 315.91}{344.32 - 281.03}$$

$$= \frac{-6.47}{63.29}$$

$$= -0.1022$$

Example 12: Calculate Bowley's coefficient of skewness.

Annual Sales (in Rs.000)	0- 20	20-50	50-100	100-250	250-500	500-1000
No. of Items	20	50	69	30	25	19

(B.B.M. Periyar, N 01)

Solution:

Annual Sales (in Rs.000)	No. of Items f	Cum. Freq. cf
0- 20	20	20
20- 50	50	70 ←
50- 100	69	139 ←
100- 250	30	169 ←
250- 500	25	194
500-1000	19	213
Total	N=213	----

$N/4 = 213/4 = 53.25$

$\therefore$ 20-50 is the Q_1 class.

$\therefore L_1 = 20; i_1 = 50-20 = 30;$
$f_1 = 50; cf_1 = 20$

$$Q_1 = L_1 + \left[\frac{i_1(N/4 - cf_1}{f_1}\right]$$

$$= 20 + \left[\frac{30(53.25 - 20)}{50}\right]$$

$$= 20 + \left[\frac{30 \times 33.25}{50}\right]$$

$$= 20 + 19.95$$

$$= 39.95$$

$N/2 = 213/2 = 106.5$. $\therefore$ 50-100 is the median class.

$\therefore L = 50; i = 100-50=50; f = 69; cf = 70.$

$$M = L + \left[\frac{i(N/2 - cf)}{f}\right]$$

$$= 50 + \left[\frac{50(106.5 - 70)}{69}\right]$$

$$= 50 + \left[\frac{50 \times 36.5}{69}\right]$$

$$= 50 + 26.45 = 76.45$$

$3N/4 = 3 \times 53.25 = 159.75$. $\therefore$ 100–250 is the Q_3 class.

$\therefore L_3 = 100; i_3 = 250-100 = 150; f_3 = 30; cf_3 = 139.$

$$Q_3 = L_3 + \left[\frac{i_3(3N/4 - cf_3)}{f_3}\right]$$

$$= 100 + \left[\frac{150(159.75 - 139)}{30}\right]$$

$$= 100 + \left[\frac{150 \times 20.75}{30}\right]$$

$$= 100 + 103.75$$

$$= 203.75$$

Bowley's coefficient of skewness,

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

$$= \frac{203.75 + 39.95 - 2 \times 76.45}{203.75 - 39.95}$$

$$= \frac{90.80}{163.80} = 0.5543$$

KARL-PEARSON'S & BOWLEY'S COEFFICIENTS

Example 13: From the information given below calculate Karl Pearson's coefficient of skewness and also quartile coefficient of skewness.

Measures	Place A	Place B
Mean	256.5	240.8
Median	201.0	201.6
S.D.	215.4	181.1
Third Quartile	260.0	242.0
First Quartile	157.0	164.2

(B.Com. (C.A.) Bharathiar, N 01)

Solution :

Place A

$$Sk_P = \frac{3(\overline{X} - M)}{\sigma}$$

$$= \frac{3(256.5 - 201.0)}{215.4}$$

$$= 0.7730$$

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

$$= \frac{260.0 + 157.0 - 2 \times 201.0}{260.0 - 157.0}$$

$$= \frac{15.0}{103.0}$$

$$= 0.1456$$

Place B

$$Sk_P = \frac{3(\overline{X} - M)}{\sigma}$$

$$= \frac{3(240.8 - 201.6)}{181.1}$$

$$= 0.6494$$

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

$$= \frac{242.0 + 164.2 - 2 \times 201.6}{242.0 - 164.2}$$

$$= \frac{3.0}{77.8}$$

$$= 0.0386$$