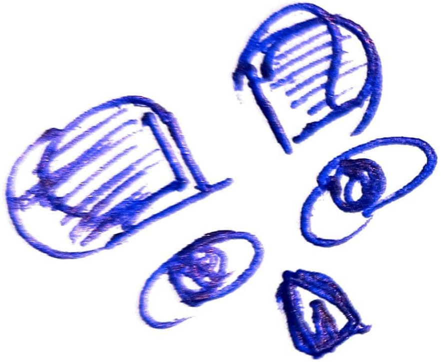


10

MEASURES OF DISPERSION

10.1 INTRODUCTION

In the preceding chapters, we have seen that averages like mean, median and mode, can be used to represent any series by a single number. There is no doubt that averages simplify the data to a large extent. However, they do not give the whole picture of the data.

There may exist different distributions whose averages are same, but which may differ widely from each other in the formation of the items. *Let us clear this point with the help of following example.*

Marks of Students in 5 Weekly Test

	X	Y	Z
1st Test	40	45	3
2nd Test	40	36	7
3rd Test	40	40	10
4th Test	40	35	85
5th Test	40	44	95
Total	200	200	200
Mean	40	40	40

In the given example, mean marks of each student is same at 40. It means, on the basis of averages, it can be said that status of all the three students is uniform. But, all the three students differ in respect of variability.

- *In case of X:* 40 marks is the perfect representative as all the values are concentrated at the same value, i.e. there is no deviation in any value with respect to the mean.
- *In case of Y:* The mean of 40 is a good representative as it is close to almost all the values, i.e. values are clustered around its mean.
- *In case of Z:* The mean of 40 may not be considered as a representative as all the values are scattered away from the value of the mean.

In this sense, we can say that even when the averages are same, the students differ from each other in the formation of the items. ***Hence averages are not sufficient to describe the characteristic of a statistical data.***

So, it is necessary to define some additional summary measures to adequately represent the characteristics of a distribution. One such measure is known as "Measure of Dispersion".

10.2 MEANING OF DISPERSION

Dispersion is the extent to which values in a distribution differ from the average of the distribution. It indicates lack of uniformity in the size of items.

Definitions of Dispersion

In the words of Connor, "Dispersion is a measure of the extent to which the individual items vary".

In the words of Spiegel, "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data".

In the words of Brooks and Dick, "Dispersion or spread is the degree of the scatter or variation of the variables about a central value".

In the words of W.L. King, "The term dispersion is used to indicate the facts that within a given group, the items differ from one another in size, or in other words, there is lack of uniformity in their sizes".

From the above definitions, it is clear that dispersion means the scatter of the values in a given distribution.

Average Vs Dispersion

An average is a single value which represents a set of values in a distribution. It is the central value which represents the entire distribution. On the other hand, dispersion indicates the extent to which the individual values fall away from the central value. Dispersion improves the understanding of a distribution.

For example, per capita income gives only the average income. A measure of dispersion can tell you about income inequalities, thereby improving the understanding of the relative standards of living enjoyed by different strata of society.

Averages of Second Order

Since measures of dispersion give an average of differences of various items from an average, they are termed as *"Averages of Second Order"*.

10.3 DISPERSION: ABSOLUTE OR RELATIVE

The measure of dispersion can be either 'absolute' or 'relative'.

1. Absolute Measure: ***The measures of dispersion which are expressed in terms of original units of a series are termed as Absolute Measures.***
   - *For example,* the dispersion of salaries about an average is measured in rupees.
   - Absolute measures are expressed in concrete units, i.e. units in terms of which the data has been expressed like rupees, centimetres, kilograms, etc.
   - Such measures are not suitable for comparing the variability of the two distributions which are expressed in different units of measurement.
2. Relative Measure: ***The measures of dispersion which are measured as a percentage or ratio of the average are termed as Relative Measures.***
   - Relative Measure is sometimes known as coefficient of dispersion as 'coefficient' means a pure number that is independent of the unit of measurement.
   - When two or more series have to be compared (whether expressed in same units or different units), relative dispersion is taken into account as the absolute dispersion may be erroneous or unfit for comparison if the series are originally in different units.

Absolute Vs Relative: *When two sets of data are expressed in different units (like quintals of sugar versus tonnes of sugarcane) or if the average size is very different (like manager's salary versus worker's wages), then the absolute measures of dispersion are not comparable. In such cases, measures of relative dispersion should be used.*

10.4 OBJECTIVES OF MEASURING DISPERSION

Following are some of the purposes for which measures of dispersion are needed.

1. To test reliability of an average: Measures of dispersion are used to test to what extent, an average represents the characteristic of a series.
   - A low value of dispersion implies that there is greater degree of homogeneity among various items and, consequently, their average can be taken as more reliable or representative of the distribution.
   - On the other hand, if the dispersion is large, then the data values are more deviated from the central value, thereby implying that the average is not representative of the data and hence not quite reliable.
2. To compare the extent of variability in two or more distributions: It aims to find out degree of uniformity or consistency in two or more sets of data. A high degree of variation would mean little uniformity or consistency, whereas a low degree of variation would mean greater uniformity or consistency.
3. To control the variability itself: The study of variation helps to analyse the reasons and causes of variations. This may be helpful in controlling the variation itself.
4. To facilitate and to serve as the basis for further statistical analysis: Measures of dispersions are used in computations of various important statistical techniques like correlation, regression, test of hypothesis, etc.

10.5 CHARACTERISTICS OF A GOOD MEASURE OF DISPERSION

A measure of dispersion is the average of second order. Therefore, all the qualities of a good average should be possessed by a good measure of dispersion as well. So, requirements for an ideal measure of dispersion are:

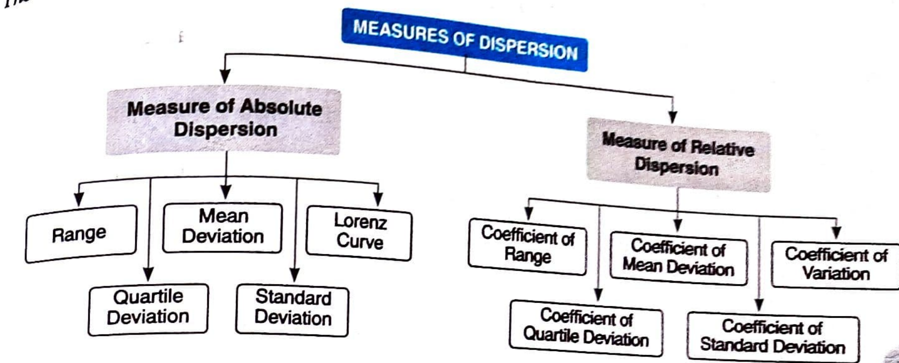
1. It should be *based on all observations.*
2. It should be *rigidly defined.*
3. It should be *easy to calculate and easy to understand.*
4. It is *not unduly affected by the fluctuations of sampling* and also by *extreme observations.*
5. It should be *capable of further mathematical/algebraic treatment.*

These requisites help in identifying the merits and demerits of individual measure of variation.

10.6 METHODS OF STUDYING DISPERSION

There are certain specific measures which help in determining the deviations from the central value. These are known as measures of dispersion.

The main measures are:



10.7 RANGE AND COEFFICIENT OF RANGE

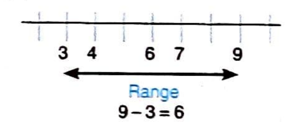
Range

The range is the simplest of all the measures of dispersion. ***It is defined as the difference between the largest and the smallest item in a distribution.***

Symbolically:

Range (R) = Largest item (L) – Smallest item (S)

- If the marks received by students of XI^{th} class are arranged in ascending (or descending) order, then the range of marks will be difference between the highest and the lowest marks.



- Range is an absolute measure of dispersion.
- A larger value of range indicates greater dispersion and a smaller value of range indicates lesser dispersion among the items.
- If all the items are the same, range will have a value of 0, i.e., there is no dispersion between the items.

As range is a measure of absolute dispersion, it cannot be usefully employed to compare the variability of two distributions expressed in different units. For example, if amount of dispersion is measured in rupees, then it is not comparable with dispersion measured in inches. In such cases, we need a relative measure (Coefficient of Range), which is independent of the units of measurement.

Coefficient of Range

Coefficient of range refers to the ratio of the difference between two extreme items (the largest and the smallest) of the distribution to their sum. Coefficient of Range is a relative measure of dispersion.

Symbolically:

$$\textbf{Coefficient of Range} = \frac{\text{Largest Item (L) – Smallest Item (S)}}{\text{Largest Item (L) + Smallest Item (S)}}$$

Range Vs Coefficient of Range

Range is an absolute measure of dispersion. So, if we want to compare the variability of two or more distributions with the same units of measurement, we may use range.

Coefficient of Range is a relative measure of dispersion, i.e., it is free from the units of measurement. If we want to compare the distributions given in different units, we make use of coefficient of range.

10.8 CALCULATION OF RANGE AND COEFFICIENT OF RANGE

Individual Series

Example 1. Calculate range and coefficient of range for values: 87, 92, 47, 58, 87, 62, 73, 73, 61

Solution:

It is always better to first arrange the observations in ascending or descending order. In an ascending order, the values are: 47, 58, 61, 62, 73, 73, 87, 87, 92

A visual examination shows that: Largest item (L) = 92; Smallest item (S) = 47

Range = L – S = 92 – 47 = 45

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{92-47}{92+47} = \frac{45}{139} = 0.32$$

Ans. Range = 45; Coefficient of range = 0.32

Example 2. The following data represents the marks of 10 students:

Marks	20	24	22	40	30	35	59	55	25	70

Calculate range and coefficient of range. Also calculate the percentage change in range, if the maximum marks is omitted.

Solution:

In an ascending order, the marks are: 20, 22, 24, 25, 30, 35, 40, 55, 59, 70

For the given values of marks, Largest item (L) = 70; Smallest item (S) = 20

Range = L – S = 70 – 20 = 50 marks

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{70-20}{70+20} = \frac{50}{90} = 0.55$$

When maximum marks of 70 is omitted, then New Range will be:

New Range = L – S = 59 – 20 = 39.

Change in Range = Original Range – New Range = 50 – 39 = 11

$$\text{Percentage Change in Range} = \frac{11}{50} \times 100 = 22\%$$

Ans. Range = 50 marks; Coefficient of range = 0.55; Percentage change in range (when maximum marks is omitted) = 22%

Discrete Series

In a discrete series, the values of largest (L) and smallest (S) item should not be mistaken as the largest and the smallest frequencies. They are the largest and smallest values of the variable.

So, range is calculated by subtracting the smallest item from the largest item, without taking into account their frequencies.

Example 3. Find the range and coefficient of range of the following distribution:

Items	3	4	5	6	7	8	9	10
Frequency	35	30	20	10	8	3	2	1

Solution:

Range (R) = Largest item (L) – Smallest item (S) = 10 – 3 = 7

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{10-3}{10+3} = \frac{7}{13} = 0.54$$

Ans. Range = 7; Coefficient of range = 0.54

Continuous Series

In case of continuous frequency distributions, the range and coefficient of range can be calculated by two methods:

1. **First Method:** Take the difference between lower limit of the lowest class-interval and upper limit of the highest class-interval.
2. **Second Method:** Take the difference between the mid-points of the lowest class-interval and the highest class-interval.

Both the methods will give different answers. But, both the answers will be correct.

It must be noted that in case of a continuous series also, the frequencies of various class-intervals are immaterial since range depends only on the two extreme observations.

Example 4. Calculate the value of range and coefficient of range.

Marks	10–20	20–30	30–40	40–50	50–60	60–70
No. of Students	10	15	25	30	11	6

Solution:

Range and Coefficient of Range by First Method

Range (R) = Largest item (L) – Smallest item (S) = 70 – 10 = 60

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{70-10}{70+10} = \frac{60}{80} = 0.75$$

Ans. Range = 60; Coefficient of range = 0.75

Range and Coefficient of Range by Second Method

Range = Mid-point of highest class – Mid-point of lowest class = 65 – 15 = 50

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{65-15}{65+15} = \frac{50}{80} = 0.625$$

Ans. Range = 50; Coefficient of range = 0.625

Note: *In the above two methods, we get different answers, but both the answers are correct.*

Example 5. Calculate Range and its coefficient from the following data:

Marks (Below)	10	20	30	40	50	60	70	80	90	100
No. of Students	12	32	62	105	165	205	230	238	244	245

Solution:

In the given example, we will first calculate the simple class-intervals and convert the series into non-cumulative series, to determine the largest and smallest item.

Marks	No. of Students
0–10	12
10–20	20
20–30	30
30–40	43
40–50	60
50–60	22
60–70	40
70–80	55
80–90	10
90–100	6

Range (R) = Largest item (L) – Smallest item (S) = 100 – 0 = 100 marks

Coefficient of Range $= \frac{L-S}{L+S} = \frac{100-0}{100+0} = \frac{100}{100} = 1$

Ans. Range = 100 marks; Coefficient of range = 1

Example 6. Calculate Range and its coefficient for the age

Age (in Years)	5–7	8–10	11–13	14–16	17–19
No. of Students	20	18	10	8	4

Solution:

The inclusive series will be converted into exclusive series to determine the largest and smallest item.

Age (in Years)	No. of students
4.5–7.5	20
7.5–10.5	18
10.5–13.5	10
13.5–16.5	8
16.5–19.5	4

Range (R) = Largest item (L) – Smallest item (S) = 19.5 – 4.5 = 15 Years

Coefficient of Range $= \frac{L-S}{L+S} = \frac{19.5-4.5}{19.5+4.5} = \frac{15}{24} = 0.625$

Ans. Range = 15 Years; Coefficient of range = 0.625

Example 7. Find the range and coefficient of range.

Individual Series		Discrete Series		Continuous Series	
S. No.	Items	Marks	No. of students	Marks	No. of students
1	22	12	3	10–20	15
2	24	15	10	20–30	12
3	26	18	20	30–40	10
4	28	21	12	40–50	9
5	30	24	5	50–60	7
6	32	27	2	60–70	20

Solution:

Calculation of Range and Coefficient of range

Individual Series	Discrete Series	Continuous Series
Range (R) = L – S where, L = 32; S = 22 Therefore, R = 32 – 22 Range = 10 Coefficient of Range $= \frac{L-S}{L+S} = \frac{32-22}{32+22} = \frac{10}{54} = 0.18$ **Ans.** 1. Range = 10 2. Coefficient of range = 0.18	Range (R) = L – S where, L = 27; S = 12 Therefore, R = 27 – 12 Range = 15 Coefficient of Range $= \frac{L-S}{L+S} = \frac{27-12}{27+12} = \frac{15}{39} = 0.38$ **Ans.** 1. Range = 15 marks 2. Coefficient of Range = 0.38	Range (R) = L – S where, L = 70; S = 10 Therefore, R = 70 – 10 Range = 60 Coefficient of Range $= \frac{L-S}{L+S} = \frac{70-10}{70+10} = \frac{60}{80} = 0.75$ **Ans.** 1. Range = 60 marks 2. Coefficient of Range = 0.75

10.9 MERITS AND DEMERITS OF RANGE

Merits of Range

1. It is ***simple*** to understand and easy to compute.
2. It gives a ***quick measure*** of variability.
3. Range provides the ***broad picture of the data*** at a glance.

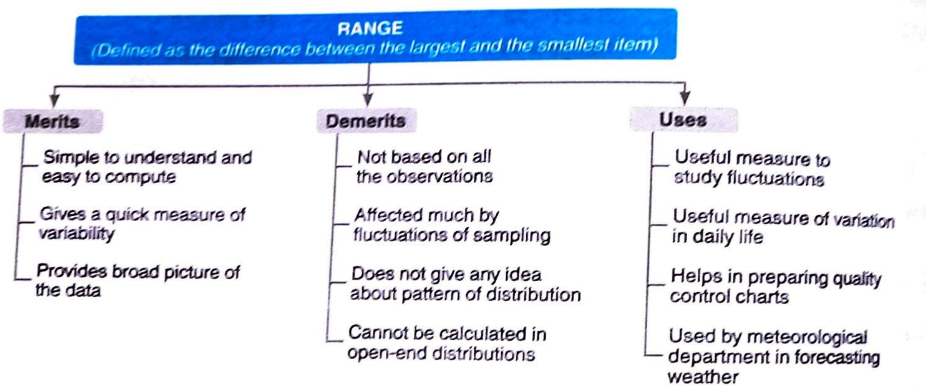
Demerits of Range

1. Range is ***not based on all the observations.*** If all the items of a distribution are replaced except the smallest and the largest item, then the range of the distribution remains same.
2. Range is very much ***affected by fluctuations of sampling.*** Its value varies widely from sample to sample.
3. It ***does not give any idea about the pattern of the distribution.*** There can be two distributions with the same range but different patterns of distribution.

4. The range *cannot be calculated in open-end distributions* because of the absence of highest and lowest class boundaries.

10.10 USES OF RANGE

1. Useful measure to study fluctuations: Fluctuations in the prices of common items and shares are studied effectively through range.
2. Day-to-day living: Range is a useful measure of variation in our daily life. *For example*, while purchasing readymade garments, we generally ask a shopkeeper about the maximum and minimum price of garments.
3. Quality Control: Range plays a very important role in preparing quality control charts. If the quality of the product remains within this prescribed range then quality is under control.
4. Weather Forecasts: The meteorological department makes use of range in forecasting weather conditions, humidity level, rainfalls, etc. These matters are of great public interest.



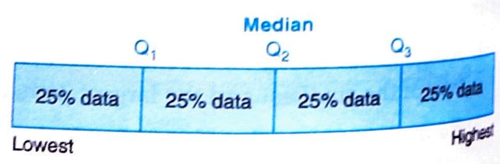
10.11 QUARTILE DEVIATION

Interquartile Range

Range is a crude measure because it takes into account only two extreme values, i.e. the largest and the smallest. The effect of extreme values on range can be avoided if we use the measure of interquartile range. ***Interquartile range refers to the difference between the values of two quartiles.***

Symbolically: **Interquartile Range = $Q_3 - Q_1$**

Quartiles and data distribution



Quartile Deviation (Semi-Interquartile Range)

Quartile Deviation is known as the half of difference of upper quartile (Q_3) and the lower quartile (Q_1). It is half of the inter-quartile range (range among the quartiles). So, it is also known as the Semi Interquartile Range.

Symbolically,

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

Coefficient of Quartile Deviation

Quartile deviation is an absolute measure of dispersion. *For comparative studies of variability of two distributions, we make use of relative measure, known as Coefficient of Quartile Deviation.*

Symbolically,

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

Coefficient of quartile deviation is studied to compare the degree of variation in the series.

10.12 CALCULATION OF QUARTILE DEVIATION

Individual Series

Example 8. Find the interquartile range, quartile deviation and coefficient of quartile deviation from the data given below: 200, 210, 208, 160, 220, 250, 300

Solution:

Calculation of lower quartile (Q_1) and upper quartile (Q_3)

S. No.	Items arranged in ascending order
1	160
2	200
3	208
4	210
5	220
6	250
7	300
N = 7	

$Q_1 = \text{Size of} \left[\frac{N+1}{4}\right]^{th} \text{item} = \text{Size of} \left[\frac{7+1}{4}\right]^{th} \text{item} = \text{Size of } 2^{nd} \text{ item}$

$Q_1 = 200$

$Q_3 = \text{Size of } 3\left[\frac{N+1}{4}\right]^{th} \text{item} = \text{Size of } 3\left[\frac{7+1}{4}\right]^{th} \text{item} = \text{Size of } 6^{th} \text{ item}$

$Q_3 = 250$

Interquartile Range = $Q_3 - Q_1 = 250 - 200 = 50$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{250 - 200}{2} = 25$

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{250 - 200}{250 + 200} = \dfrac{50}{450} = 0.11$

Ans. Interquartile range = 50; Quartile Deviation = 25; Coefficient of Quartile Deviation = 0.11

Example 9. Calculate quartile deviation and its coefficient from the following data: 55, 60, 70, 90, 90, 110, 120, 130, 145, 145, 155, 170

Solution:

The values are already arranged in an ascending order.

Q_1 = Size of $\left[\dfrac{N+1}{4}\right]^{th}$ item = Size of $\left[\dfrac{12+1}{4}\right]^{th}$ item = Size of 3.25^{th} item

Size of 3.25^{th} item = Size of 3^{rd} item + .25 times (Size of 4^{th} item – Size of 3^{rd} item) = 70 + .25 (90 – 70) = 70 + 5

$Q_1 = 75$

Q_3 = Size of $3\left[\dfrac{N+1}{4}\right]^{th}$ item = Size of $3\left[\dfrac{12+1}{4}\right]^{th}$ item = Size of 9.75^{th} item

Size of 9.75^{th} item = Size of 9^{th} item + .75 times (Size of 10^{th} item – Size of 9^{th} item) = 145 + .75 (145 – 145)

$Q_3 = 145$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{145 - 75}{2} = 35$

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{145 - 75}{145 + 75} = \dfrac{70}{220} = 0.318$

Ans. Quartile Deviation = 35; Coefficient of Quartile Deviation = 0.318

Example 10. Calculate lower and upper quartiles, when Quartile deviation = 10 and Coefficient of quartile deviation = 0.5.

Solution:

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = 10$

$Q_3 - Q_1 = 20$ ...(1)

Coefficient Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = 0.5$

$= \dfrac{20}{Q_3 + Q_1} = 0.5$ (As $Q_3 - Q_1 = 20$)

So, $Q_3 + Q_1 = 40$

or, $Q_3 = 40 - Q_1$

Putting the value of Q_3 in (1) we get:

$(40 - Q_1) - Q_1 = 20$

$-2Q_1 = -20$

Or, $Q_1 = 10$

Also, $Q_3 = 40 - Q_1 = 40 - 10 = 30$

Ans. Lower Quartile (Q_1) = 10; Upper Quartile (Q_3) = 30

Discrete Series

Example 11. From the following table giving height of students, calculate the interquartile range, quartile deviation and coefficient of quartile deviation:

Height (in cm.)	153	155	157	159	161	163	165	167	169
No. of Students	25	21	28	20	18	24	22	18	23

Solution:

Height (in cm) (X)	*No. of students (f)*	*c.f.*
153	25	25
155	21	46
157	28	74
159	20	94
161	18	112
163	24	136
165	22	158
167	18	176
169	23	199

Q_1 = Size of $\left[\dfrac{N+1}{4}\right]^{th}$ item = Size of $\left[\dfrac{199+1}{4}\right]^{th}$ item = Size of 50^{th} item

Q_1 = 157 centimetres

Q_3 = Size of $3\left[\dfrac{N+1}{4}\right]^{th}$ item = Size of $3\left[\dfrac{199+1}{4}\right]^{th}$ item = Size of 150^{th} item

Q_3 = 165 centimetres

Interquartile range = $Q_3 - Q_1 = 165 - 157 = 8$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{165 - 157}{2} = 4$

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{165 - 157}{165 + 157} = \dfrac{8}{322} = 0.025$

Ans. Interquartile range = 8 centimetres; Quartile Deviation = 4 centimetres; Coefficient of Quartile Deviation = 0.025

Example 12. From the following particulars, calculate the range of marks obtained by middle 50% of the students. Also calculate quartile deviation.

Marks	2	4	6	8	10	12
No. of students	3	5	10	12	6	4

Solution:

Marks (X)	No. of students (f)	c.f.
2	3	3
4	5	8
6	10	18
8	12	30
10	6	36
12	4	40

To calculate marks of middle 50% of students, we will have to calculate the difference between marks of 10[th] student (i.e. Q_1) and 30[th] student (i.e. Q_3), i.e. we have to calculate interquartile range.

$$Q_1 = \text{Size of} \left[\frac{N+1}{4}\right]^{th} \text{item} = \text{Size of} \left[\frac{40+1}{4}\right]^{th} \text{item} = \text{Size of } 10.25^{th} \text{ item}$$

$Q_1 = 6$

$$Q_3 = \text{Size of } 3\left[\frac{N+1}{4}\right]^{th} \text{item} = \text{Size of } 3\left[\frac{40+1}{4}\right]^{th} \text{item} = \text{Size of } 30.75^{th} \text{ item}$$

$Q_3 = 10$

Interquartile range $= Q_3 - Q_1 = 10 - 6 = 4$

Thus, the range of marks obtained by middle 50% of the students = 4

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{10-6}{2} = 2$$

Ans. Range of marks obtained by middle 50% of the students (Interquartile range) = 4 marks; Quartile Deviation = 2 marks

Continuous Series

Example 13. Calculate the interquartile range, quartile deviation and coefficient of quartile deviation from the following figures:

Size	0–5	5–10	10–15	15–20	20–25	25–30
Frequency	3	9	15	23	30	20

Solution:

Size (X)	Frequency (f)	c.f.
0–5	3	3
5–10	9	12
10–15	15	27
15–20	23	50
20–25	30	80
25–30	20	100
	$N = \Sigma f = 100$	

$$Q_1 = \frac{N}{4} = \frac{100}{4} = 25^{th} \text{ item}$$

25[th] item lies in the group 10 – 15

$l_1 = 10$, c.f. = 12, f = 15, i = 5

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 10 + \frac{25-12}{15} \times 5 = 14.33$$

$$Q_3 = \frac{3N}{4} = \frac{3(100)}{4} = 75^{th} \text{ item}$$

75[th] item lies in the group 20 – 25

$l_1 = 20$, c.f. = 50, f = 30, i = 5

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 20 + \frac{75-50}{30} \times 5 = 24.17$$

Interquartile range $= Q_3 - Q_1 = 24.17 - 14.33 = 9.84$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{24.17 - 14.33}{2} = 4.92$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{24.17 - 14.33}{24.17 + 14.33} = \frac{9.84}{38.50} = 0.25$$

Ans. Interquartile range = 9.84; Quartile Deviation = 4.92; Coefficient of Quartile Deviation = 0.25

Example 14. Calculate the value of interquartile range, quartile deviation and coefficient of quartile deviation:

Marks	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50	51–55
No. of students	10	17	22	31	42	32	26	19	14

Solution:

This is a case of inclusive class-intervals. It should be first converted into exclusive series.

Marks (X)	No. of Students (f)	c.f.
10.5–15.5	10	10
15.5–20.5	17	27
20.5–25.5	22	49
25.5–30.5	31	80
30.5–35.5	42	122
35.5–40.5	32	154

40.5–45.5	26	180
45.5–50.5	19	199
50.5–55.5	14	213
	N = Σf = 213	

$$Q_1 = \frac{N}{4} = \frac{213}{4} = 53.25^{th} \text{ item}$$

53.25^{th} item lies in the group 25.5 – 30.5

$l_1 = 25.5$, c.f. = 49, f = 31, i = 5

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 25.5 + \frac{53.25 - 49}{31} \times 5 = 26.18$$

$$Q_3 = \frac{3N}{4} = \frac{3(213)}{4} = 159.75^{th} \text{ item}$$

159.75^{th} item lies in the group 40.5 – 45.5

$l_1 = 40.5$, c.f. = 154, f = 26, i = 5

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 40.5 + \frac{159.75 - 154}{26} \times 5 = 41.6$$

Interquartile range = $Q_3 - Q_1 = 41.6 - 26.18 = 15.42$ marks

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{41.6 - 26.18}{2} = 7.71 \text{ Marks}$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{41.6 - 26.18}{41.6 - 26.18} = \frac{15.42}{67.78} = 0.22$$

Ans. Interquartile range = 15.42 marks; Quartile Deviation = 7.71 marks; Coefficient of Quartile Deviation = 0.22

Example 15. Calculate the appropriate measure of dispersion from the following data:

Marks	Below 20	20–30	30–40	Above 40
No. of students	7	10	14	9

Solution:

Since we have the frequency distribution with open-end classes, the appropriate (rather the only) measure of dispersion, that we can compute is Quartile Deviation and its coefficient.

Marks (X)	*No of students (f)*	*c.f.*
Below 20	7	7
20–30	10	17
30–40	14	31
Above 40	9	40
	N = Σf = 40	

$$Q_1 = \frac{N}{4} = \frac{40}{4} = 10^{th} \text{ item}$$

10^{th} item lies in the group 20 – 30

$l_1 = 20$, c.f. = 7, f = 10, i = 10

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 20 + \frac{10 - 7}{10} \times 10 = 23$$

$$Q_3 = \frac{3N}{4} = \frac{3(40)}{4} = 30^{th} \text{ item}$$

30^{th} item lies in the group 30 – 40

$l_1 = 30$, c.f. = 17, f = 14, i = 10

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 30 + \frac{30 - 17}{14} \times 10 = 39.28$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{39.28 - 23}{2} = 8.14 \text{ Marks}$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{39.28 - 23}{39.28 + 23} = \frac{16.28}{62.28} = 0.26$$

Ans. Quartile Deviation = 8.14 marks; Coefficient of Quartile Deviation = 0.26

Note: *In the given example, extreme items at either end of the series cannot influence the value of quartile deviation. It is only the middle half of the data i.e., $Q_3 - Q_1$ which is needed for calculating quartile deviation. So, in case of frequency distribution with open-end classes, the appropriate (rather the only) measure of dispersion, that we can compute is Quartile Deviation.*

10.13 MERITS AND DEMERITS OF QUARTILE DEVIATION

Merits of Quartile Deviation

1. It is quite ***easy to understand and calculate.***
2. It is the only measure of dispersion which can be ***used to deal with a distribution having open-end classes.***
3. In comparison to range, it is ***less affected by extreme values.***

Demerits of Quartile Deviation

1. It is ***not based on all the observations*** as it ignores the first 25% and the last 25% of the items. Thus, it cannot be regarded as a reliable measure of variability.
2. It is ***not capable of further algebraic treatment.*** It is in a way a positional average and does not study variation of the values of a variable from any average.

3. It is considerably ***affected by fluctuations*** in the sample. A change in the value of a single item, may, in many cases, affect its value considerably.

QUARTILE DEVIATION
(Refers to half of difference of Q_3 and Q_1)

Merits
- Easy to understand and calculate
- Can be used in distributions having open-end classes
- Less affected by Extreme Values

Demerits
- Not based on all observations.
- Not capable of further Algebraic Treatment
- Affected by fluctuations in the sample

10.14 MEAN DEVIATION

Introduction

Range, Interquartile range and Quartile deviation suffer from a common defect i.e., they are calculated by taking into account only two values of a series: either the extreme values as in case of range, or the values of the quartiles as in case of quartile deviation.

So, it is always better to have such a measure of dispersion which takes into account all the observations of a series and is calculated in relation to a central value. *Mean deviation is such a measure of dispersion.*

Meaning

Mean deviation of a series is the arithmetic average of the deviations of various items from a measure of central tendency (mean, median or mode). Mean deviation is also known as 'First moment of dispersion'.

- Mean deviation is based on all the items of the series.
- Theoretically, Mean deviation can be calculated by taking deviations from any of the three averages. But in actual practice, mean deviation is calculated either from mean or from median.
  - Mode is usually not considered as its value is indeterminate and it gives erroneous conclusions.
  - Between mean and median, the latter is supposed to be better than the former, because the sum of deviations from Median is less than the sum of the deviations from mean. So, ***if the choice is to be made between mean and median, median will be more appropriate.***
- While calculating deviations from the selected average, the signs (+ or –) of deviations are ignored and ***all the deviations are taken as positive.***

Why Plus or Minus sign is ignored!!!

If algebraic signs are considered, then the sum of deviations from the mean should be zero and from the median would be nearly zero (in case of moderately asymmetrical series). So, to study the variation of items from a central value, plus and minus signs are ignored.

Coefficient of Mean Deviation

Mean deviation is an absolute measure of dispersion. In order to transform it into a relative measure, it is divided by the average, from which it has been calculated. It is then known as the Coefficient of Mean Deviation.

Symbolically:

$$\textbf{Coefficient of Mean Deviation from Mean } (\bar{X}) = \frac{MD_{\bar{X}}}{\bar{X}}$$

$$\textbf{Coefficient of Mean Deviation from Median (Me)} = \frac{MD_{Me}}{Me}$$

10.15 CALCULATION OF MEAN DEVIATION AND ITS COEFFICIENT

Individual Series

In case of individual series, the mean deviation is calculated by totalling the deviations from the mean or median (ignoring plus or minus signs) and dividing the total by the number of items.

Steps for Calculation of Mean Deviation

Step 1. Calculate the specific average (Mean or Median) from which mean deviation is to be found out.

Step 2. Obtain absolute (positive) deviations of each observation from the specific average.

Step 3. Absolute deviations are totalled up to find out $\Sigma|D|$

Step 4. Apply the formula:

$$\text{Mean Deviation from Mean } (MD_{\bar{X}}) = \frac{\Sigma|X-\bar{X}|}{N} = \frac{\Sigma|D|}{N}$$

$$\text{Mean Deviation from ~~Mean~~ Median } (MD_{Me}) = \frac{\Sigma|X-Me|}{N} = \frac{\Sigma|D|}{N}$$

Example 16. Calculate mean deviation from mean and median for the given data: 5, 8, 11, 12, 14.

Solution:

Calculation of Mean Deviation from Mean

Values (X)	Deviations from Mean $|X-\bar{X}|$ $|D|$
5	5
8	2
11	1
12	2
14	4
$\Sigma X = 50$	$\Sigma|D| = 14$

$$\text{Mean } (\bar{X}) = \frac{\Sigma X}{N} = \frac{50}{5} = 10$$

$$\text{Mean Deviation from Mean } (MD_{\bar{X}}) = \frac{\Sigma|D|}{N} = \frac{14}{5} = 2.8$$

$$\text{Coefficient of Mean deviation} = \frac{MD_{\bar{X}}}{\bar{X}} = \frac{2.8}{10} = 0.28$$

Ans. Mean deviation from mean = 2.8; Coefficient of mean deviation = 0.28

Calculation of Mean Deviation from Median

$$Me = \text{Size of } \left[\frac{N+1}{2}\right]^{th} \text{ item} = \text{Size of } \left[\frac{5+1}{2}\right]^{th} \text{ item} = \text{Size of } 3^{rd} \text{ item}$$

Median = 11

Values (X)	Deviations from Median $\|X - M\|$ $\|D\|$
5	6
8	3
11	0
12	1
14	3
$\Sigma X = 50$	$\Sigma\|D\| = 13$

$$\text{Mean Deviation from Median } (MD_{Me}) = \frac{\Sigma|D|}{N} = \frac{13}{5} = 2.6$$

$$\text{Coefficient of Mean deviation} = \frac{MD_{Me}}{\text{Median}} = \frac{2.6}{11} = 0.23$$

Ans. Mean deviation from median = 2.6; Coefficient of mean deviation = 0.23

Alternate Method

Mean Deviation can also be calculated by calculating deviations from an assumed mean. This method is adopted especially when the actual mean is a fractional number.

Mean deviation from assumed mean is calculated by the following formula:

$$\textbf{Mean deviation from assumed mean} = \frac{\Sigma|D| + (\bar{X} - A)(\Sigma f_B - \Sigma f_A)}{N}$$

Where:

$\Sigma|D|$ = Sum of absolute deviations from Assumed Mean

$\bar{X}$ = Actual Mean

A = Assumed Mean

Σf_B = Number of values above actual mean

Σf_A = Number of values below actual mean including actual mean

N = Number of observations

Example 17. Calculate mean deviation in the following example by taking 11 as assumed mean
5, 8, 11, 12, 14

Solution:

Calculation of Mean Deviation from Assumed Mean

Values (X)	Deviations from Assumed Mean $\|X - A\|$ $\|D\|$
5	6
8	3
11 (A)	0
12	1
14	3
$\Sigma X = 50$	$\Sigma\|D\| = 13$

$$\text{Mean deviation from assumed mean} = \frac{\Sigma|D| + (\bar{X} - A)(\Sigma f_B - \Sigma f_A)}{N}$$

$$= \frac{13 + (10 - 11)(2 - 3)}{5} = \frac{14}{5} = 2.8$$

Ans. Mean deviation from Assumed Mean = 2.8

Example 18. With mean as the base, calculate mean deviation and compare the variability of the two series A and B.

Series A	10	12	16	20	25	27	30
Series B	10	20	22	25	27	31	40

Solution:

Calculation of Mean Deviation from Mean

Series A (X_A)	Deviations from Mean $\|D\| = (X_A - \bar{X}_A)$	Series B (X_B)	Deviations from Mean $\|D\| = (X_B - \bar{X}_B)$
10	10	10	15
12	8	20	5
16	4	22	3
20	0	25	0
25	5	27	2
27	7	31	6
30	10	40	15
$\Sigma X_A = 140$	$\Sigma\|D\| = 44$	$\Sigma X_B = 175$	$\Sigma\|D\| = 46$

Series A

$$\text{Mean } (\bar{X}_A) = \frac{\Sigma X_A}{N} = \frac{140}{7} = 20$$

$$\text{Mean deviation (Series A)} = \frac{\Sigma|D|}{N} = \frac{44}{7} = 6.28$$

$$\text{Coefficient of M.D. (Series A)} = \frac{M.D.}{\bar{X}_A} = \frac{6.28}{20} = 0.31$$

Series B

$$\text{Mean } (\bar{X}_B) = \frac{\Sigma X_B}{N} = \frac{175}{7} = 25$$

$$\text{Mean deviation (Series B)} = \frac{\Sigma|D|}{N} = \frac{46}{7} = 6.57$$

$$\text{Coefficient of M.D. (Series B)} = \frac{M.D.}{\bar{X}_B} = \frac{6.57}{25} = 0.26$$

Ans. Since coefficient of mean deviation for series A is more than that of series B, we can say that series A has greater variability as compared to series B.

Discrete Series

The calculation of mean deviation in a discrete series involves the following steps:

Step 1. Calculate specific average from which mean deviation is to be found out.

Step 2. Obtain the absolute (positive) deviations $|D|$ of each observation from the specific average.

Step 3. Multiply absolute deviations $|D|$ with respective frequencies (f) and obtain the sum of products to get. $\Sigma f|D|$

Step 4. Divide $\Sigma f|D|$ by number of items to get mean deviation.

$$\text{Mean Deviation from Mean } (MD_{\bar{X}}) = \frac{\Sigma f|X-\bar{X}|}{N} = \frac{\Sigma f|D|}{N}$$

$$\text{Mean Deviation from Median } (MD_{Me}) = \frac{\Sigma f|X-Me|}{N} = \frac{\Sigma f|D|}{N}$$

Example 19. Calculate mean deviation about arithmetic mean and coefficient of mean deviation.

Values (X)	10	11	12	13
Frequency (f)	3	12	18	12

Solution:

Calculation of Mean Deviation about Arithmetic mean

Values (X)	Frequency (f)	(fX)	$\lvert D\rvert = \lvert X-\bar{X}\rvert$ $\lvert X-11.87\rvert$	$f\lvert D\rvert$
10	3	30	1.87	5.61
11	12	132	0.87	10.44
12	18	216	0.13	2.34
13	12	156	1.13	13.56
	Σf = 45	**ΣfX = 534**		**$\Sigma f\lvert D\rvert$ = 31.95**

$$\text{Mean } (\bar{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{534}{45} = 11.87$$

$$\text{Mean deviation from Mean } (MD_{\bar{X}}) = \frac{\Sigma f|D|}{N} = \frac{31.95}{45} = 0.71$$

$$\text{Coefficient of Mean deviation} = \frac{MD_{\bar{X}}}{\bar{X}} = \frac{0.71}{11.87} = 0.059$$

Ans. Mean deviation from mean = 0.71; Coefficient of mean deviation = 0.059

Example 20. Calculate absolute and relative measures of range and quartile deviation. Also calculate mean deviation from median and Coefficient of mean deviation for the following data:

Daily Income (₹)	100	150	80	200	250	180
No. of Persons	24	26	16	20	6	30

Solution:

Calculation of Absolute and Relative Measures of Range

Absolute Measure (Range)

Range = Largest item (L) – Smallest item (S) = 250 – 80= ₹ 170

Relative Measure (Coefficient of Range)

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{250-80}{250+80} = \frac{170}{330} = 0.515$$

Calculation of Absolute and Relative Measures of Quartile Deviation

Arranging the data in ascending order of magnitude and calculating cumulative frequencies, we get:

Daily Income (₹) (X)	No. of persons (f)	c.f.
80	16	16
100	24	40
150	26	66
180	30	96
200	20	116
250	6	122
	N = Σf = 122	

$$Q_1 = \text{Size of} \left[\frac{N+1}{4}\right]^{th} \text{item} = \text{Size of} \left[\frac{122+1}{4}\right]^{th} \text{item} = \text{Size of } 30.75^{th} \text{ item}$$

30.75^{th} item falls in the cumulative frequency of 40 and the size against this cumulative frequency is 100. Therefore, $Q_1 = 100$

$$Q_3 = \text{Size of } 3\left[\frac{N+1}{4}\right]^{th} \text{item} = \text{Size of } 3\left[\frac{122+1}{4}\right]^{th} \text{item} = \text{Size of } 92.25^{th} \text{ item}$$

92.25^{th} item falls in the cumulative frequency of 96 and the size against this cumulative frequency is 180. So, $Q_3 = 180$

Absolute Measure (Quartile Deviation)

$$\text{Quartile Deviation} = \frac{Q_3-Q_1}{2} = \frac{180-100}{2} = ₹\,40$$

Relative Measure (Coefficient of Quartile Deviation)

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3-Q_1}{Q_3+Q_1} = \frac{180-100}{180+100} = \frac{80}{280} = 0.285$$

Calculation of Mean Deviation from Median and Its Coefficient

Daily Income (₹) (X)	No. of persons (f)	c.f.	$\lvert D\rvert = \lvert X - Me\rvert$ $\lvert X - 150\rvert$	$f\lvert D\rvert$
80	16	16	70	1,120
100	24	40	50	1,200
150	26	66	0	0
180	30	96	30	900
200	20	116	50	1,000
250	6	122	100	600
	N = Σf = 122			$\Sigma f\lvert D\rvert = 4{,}820$

Median = Size of $\left[\frac{N+1}{2}\right]^{th}$ item = Size of $\left[\frac{122+1}{2}\right]^{th}$ item = Size of 61.5th item

Median = 150

Absolute Measure (Mean Deviation)

Mean deviation from Median (MD_{Me}) = $\frac{\Sigma f\lvert D\rvert}{N} = \frac{4{,}820}{122}$ = ₹ 39.51

Relative Measure (Coefficient of Mean Deviation)

Coefficient of Mean deviation = $\frac{MD_{Me}}{Median} = \frac{39.51}{150} = 0.26$

Ans. Absolute Measures: Range = ₹ 170; Quartile Deviation = ₹ 40; Mean Deviation from median = ₹ 39.51

Relative Measure: Coefficient of Range = 0.515; Coefficient of Quartile Deviation = 0.285; Coefficient of Mean Deviation = 0.26

Continuous Series

In case of continuous series, the formulae of mean deviation are same as that for a discrete series. *For the given continuous frequency distribution, the mid-points of class intervals have to be found out and they are taken as 'm'.* In this way, a continuous series assumes the shape of a discrete series. After that, all the steps of discrete series are applied.

Symbolically,

Mean Deviation from Mean ($MD_{\bar{X}}$) = $\frac{\Sigma f\lvert m - \bar{X}\rvert}{N} = \frac{\Sigma f\lvert D\rvert}{N}$

Mean Deviation from Median (MD_{Me}) = $\frac{\Sigma f\lvert m - Me\rvert}{N} = \frac{\Sigma f\lvert D\rvert}{N}$

(Where, m = Mid-point; $\bar{X}$ = Arithmetic Mean; Me = Median)

The following examples would illustrate the calculation of mean deviation in case of continuous series.

Example 21. Calculate mean deviation from mean and its coefficient for the following data:

Class-interval	2–4	4–6	6–8	8–10
Frequency	3	4	2	1

Solution:

Calculation of Mean Deviation from Mean

Class-interval (X)	Frequency (f)	Mid-point (m)	fm	$\lvert D\rvert = \lvert m - \bar{X}\rvert$ $\lvert m - 5.2\rvert$	$f\lvert D\rvert$
2–4	3	3	9	2.2	6.6
4–6	4	5	20	0.2	0.8
6–8	2	7	14	1.8	3.6
8–10	1	9	9	3.8	3.8
	N = Σf = 10		Σfm = 52		$\Sigma f\lvert D\rvert = 14.8$

$(\bar{X}) = \frac{\Sigma fm}{\Sigma f} = \frac{52}{10} = 5.2$

Mean deviation from Mean ($MD_{\bar{X}}$) = $\frac{\Sigma f\lvert D\rvert}{N} = \frac{14.8}{10} = 1.48$

Coefficient of Mean deviation = $\frac{MD_{\bar{X}}}{\bar{X}} = \frac{1.48}{5.2} = 0.28$

Ans. Mean deviation from mean = 1.48; Coefficient of mean deviation = 0.28

Example 22. Calculate the following: (i) Range and Coefficient of range; (ii) Quartile Deviation and coefficient of Quartile Deviation; (iii) Mean deviation from median and its coefficient.

X	0–5	5–10	10–15	15–20	20–25	25–30
f	4	7	8	2	6	3

Solution:

(i) *Calculation of Range and Coefficient of Range*

Range (R) = Largest item (L) – Smallest item (S) = 30 – 0 = 30

Coefficient of Range = $\frac{L-S}{L+S} = \frac{30-0}{30+0} = \frac{30}{30} = 1$

(ii) *Calculation of Quartile Deviation and Coefficient of Quartile Deviation*

X	f	c.f.
0–5	4	4
5–10	7	11
10–15	8	19
15–20	2	21
20–25	6	27
25–30	3	30
	N = Σf = 30	

$Q_1 = \frac{N}{4} = \frac{30}{4} = 7.5^{th}$ item

7.5^{th} item lies in the group 5 – 10

$l_1 = 5$, c.f = 4, f = 7, i = 5

By applying formula:

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 5 + \frac{7.5 - 4}{7} \times 5 = 7.5$$

$Q_3 = \frac{3N}{4} = \frac{90}{4} = 22.5^{th}$ item

22.5^{th} item lies in the group 20 – 25

$l_1 = 20$, c.f = 21, f = 6, i = 5

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 20 + \frac{22.5 - 21}{6} \times 5 = 21.25$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{21.25 - 7.5}{2} = 6.875$$

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{21.25 - 7.5}{21.25 + 7.5} = \frac{13.75}{28.75} = 0.478$$

(iii) Calculation of Mean deviation from median and its coefficient

X	f	c.f.	Mid-point (m)	$\|D\| = \|m - 12.5\|$	$f\|D\|$
0–5	4	4	2.5	10	40
5–10	7	11	7.5	5	35
10–15	8	19	12.5	0	0
15–20	2	21	17.5	5	10
20–25	6	27	22.5	10	60
25–30	3	30	27.5	15	45
	N = Σf = 30				$\Sigma f\|D\| = 190$

$Me = \frac{N}{2} = \frac{30}{2} = 15^{th}$ item

15^{th} item lies in the group 10 – 15

$l_1 = 10$, c.f. = 11, f = 8, i = 5

$$Me = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 10 + \frac{15 - 11}{8} \times 5 = 12.5$$

$$\text{Mean deviation from median } (MD_{Me}) = \frac{\Sigma f|D|}{N} = \frac{190}{30} = 6.33$$

$$\text{Coefficient of Mean deviation} = \frac{MD_{Me}}{Median} = \frac{6.33}{12.5} = 0.506$$

Ans.
(i) Range = 30; Coefficient of range = 1;
(ii) Quartile Deviation = 6.875; Coefficient of Quartile Deviation = 0.478;
(iii) Mean deviation from median = 6.33; Coefficient of mean deviation = 0.506.

Example 23. Calculate mean deviation and coefficient of mean deviation from the following figures:

Size	Frequency
0 and up to 10	1
0 and up to 20	3
0 and up to 30	7
0 and up to 40	8
0 and up to 50	10

Solution:

In the given question, we are not given the average, from which mean deviation is to be calculated. **In such cases, we calculate mean deviation by median as median is a constant and representative value.** We shall first convert the given 'less than' cumulative frequency into continuous frequency distribution:

Calculation of Mean Deviation from Median

X	f	c.f.	Mid-point (m)	$\|D\| = \|m - 25\|$	$f\|D\|$
0–10	1	1	5	20	20
10–20	2	3	15	10	20
20–30	4	7	25	0	0
30–40	1	8	35	10	10
40–50	2	10	45	20	40
	N = Σf = 10				$\Sigma f\|D\| = 90$

$Me = \frac{N}{2} = \frac{10}{2} = 5^{th}$ item

5^{th} item lies in the group 20 – 30

$l_1 = 20$, c.f = 3, f = 4, i = 10

$$Me = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 20 + \frac{5 - 3}{4} \times 10 = 25$$

$$\text{Mean deviation from Median } (MD_{Me}) = \frac{\Sigma f|D|}{N} = \frac{90}{10} = 9$$

Coefficient of Mean deviation = $\frac{MD_{Me}}{Median} = \frac{9}{25} = 0.36$

Ans. Mean deviation from median = 9; Coefficient of mean deviation = 0.36

Example 24. Calculate mean deviation from median and Coefficient of mean deviation from the following data:

Mid-Points	10	30	50	70	90
Frequency (f)	10	16	30	32	12

Solution:

In the given example, we are given mid-points. Hence we will first convert the mid-points into class-intervals to calculate median.

Calculation of Median

Mid-points (m)	*Class-Interval (X)*	*Frequency (f)*	*c.f.*
10	0–20	10	10
30	20–40	16	26
50	40–60	30	56
70	60–80	32	88
90	80–100	12	100
		N = Σf = 100	

$Me = \frac{N}{2} = \frac{100}{2} = 50^{th}$ item

50^{th} item lies in the group 40 – 60

$l_1 = 40$, c.f = 26, f = 30, i = 20

$$Me = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 40 + \frac{50 - 26}{30} \times 20 = 56$$

Calculation of Mean Deviation from Median

Mid-points (m)	*Frequency (f)*	$\lvert D \rvert = \lvert m - 56 \rvert$	$f\lvert D \rvert$
10	10	46	460
30	16	26	416
50	30	6	180
70	32	14	448
90	12	34	408
	N = Σf = 100		$\Sigma f\lvert D \rvert = 1,912$

Mean deviation from Median (MD_{Me}) = $\frac{\Sigma f\lvert D \rvert}{N} = \frac{1,912}{100} = 19.12$

Coefficient of Mean deviation = $\frac{MD_{Me}}{Median} = \frac{19.12}{56} = 0.34$

Ans. Mean deviation from median = 19.12;
Coefficient of mean deviation = 0.34.

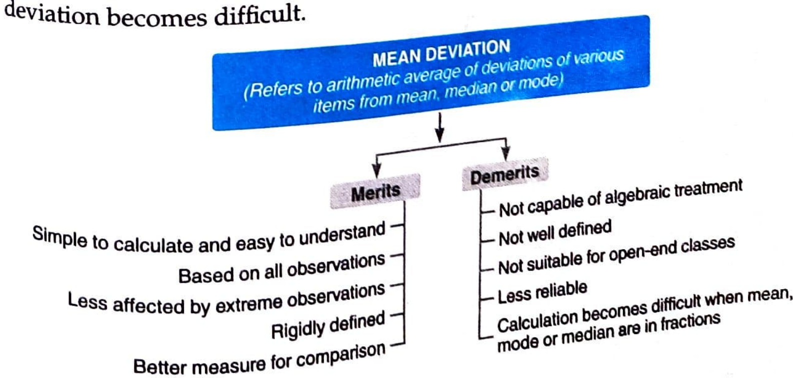
10.16 MERITS AND DEMERITS OF MEAN DEVIATION

Merits of Mean Deviation

1. Simplicity: It is simple to calculate and easy to understand.
2. Based on all observations: Mean deviation is a more comprehensive measure of dispersion (compared to range and quartile deviation) as it is based upon all items of the series.
3. Less effect of extreme values: As compared with standard deviation, it is less affected by extreme observations.
4. Rigidly defined: Mean deviation is rigidly defined and its value is precise and definite.
5. Better Measure for comparison: Mean deviation is based on the deviations from an average. So, it provides a better measure for comparison about formation of different distributions.

Demerits of Mean Deviation

1. Not capable of algebraic treatment: Mean deviation ignores the positive and negative signs of deviations. As a result, this method cannot be used for further algebraic treatment.
2. Not well defined: It is not well defined of dispersion since deviations can be taken from any measure of central tendency and mean deviation calculated from different averages (mean, median, mode) will not be same.
3. Not suitable for open-end classes: Mean deviation cannot be computed for distribution with open-end classes.
4. Less reliable: Mean deviation when calculated from mode is not reliable because in many cases mode has no fixed value.
5. Difficult calculations: If mean, mode and median are in fractions, then calculation of mean deviation becomes difficult.



10.17 STANDARD DEVIATION

Introduction

The methods of measuring dispersion discussed so far, are not universally adopted for want of adequacy and accuracy.

- *Range* is not satisfactory as its magnitude is determined by most extreme cases in the entire group.
- *Quartile deviation* has no algebraic properties and its interpretation is difficult.
- *Mean Deviation* is also an unsatisfactory measure as it ignores the algebraic signs of deviation.

Therefore, we need a measure of dispersion, which is free from these shortcomings. *To some extent, standard deviation is one such measure.*

Meaning

The concept of standard deviation was first used by Karl Pearson in the year 1893. ***Standard deviation is the square root of the arithmetic average of the squares of the deviations measured from the mean.***

- It is the most commonly used measure of dispersion as it satisfies most of the properties laid down for an ideal measure of dispersion.
- The standard deviation is denoted by the Greek letter σ.

10.18 CALCULATION OF STANDARD DEVIATION

We discuss below the calculation of standard deviation for different series:

Individual Series

There are three main methods of calculating standard deviation in case of individual series:

(i) Actual Mean Method *(Refer Example 25)*

(ii) Direct method *(Refer Example 26)*

(iii) Short-cut method or Assumed Mean Method *(Refer Example 27)*

Actual Mean Method

In this method, deviations are taken from the actual mean. *The steps involved in calculation of standard deviation are:*

Step 1. Calculate actual mean ($\overline{X}$) of the observation

Step 2. Find out deviations of each item of the series from mean, i.e., calculate ($X-\overline{X}$) and denote the deviations by x.

Step 3. Square the deviations and obtain the total, i.e., Σx^2

Step 4. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma x^2}{N}}$

Where:

σ = Standard Deviation

Σx^2 = Sum total of the squares of deviations from actual mean

N = Number of pair of observations

Let us understand the Actual Mean method with the help of Example 25.

Example 25. Calculate the standard deviation from the following data:

5	8	7	11	14

Solution:

Calculation of Standard Deviation (Actual Mean Method)

Values (X)	$x = X - \overline{X}$	x^2
5	5 − 9 = −4	16
8	8 − 9 = −1	1
7	7 − 9 = −2	4
11	11 − 9 = +2	4
14	14 − 9 = +5	25
ΣX = 45		Σx^2 = 50

Arithmetic mean ($\overline{X}$) $= \frac{\Sigma X}{N} = \frac{45}{5} = 9$

Standard Deviation (σ) $= \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{50}{5}} = \sqrt{10} = 3.16$

Ans. Standard deviation = 3.16

Direct Method

In this method, standard deviation is calculated without finding out the deviations from the actual mean. *The steps involved in the Direct Method are:*

Step 1. Calculate the actual mean ($\overline{X}$) of the observations

Step 2. Square the observations and obtain the total, i.e., ΣX^2

Step 3. Apply the following formula:

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

Where:

σ = Standard Deviation

ΣX^2 = Sum total of the squares of observations

$\overline{X}$ = Actual Mean

N = Number of pair of observations

Example 26 will make the direct method more clear:

Example 26. Calculate the Standard Deviation of the data given in Example 25 by the Direct method.

Solution:

Calculation of Standard Deviation (Direct Method)

Values (X)	*X²*
5	25
8	64
7	49
11	121
14	196
ΣX = 45	**ΣX² = 455**

$$\text{Arithmetic mean } (\bar{X}) = \frac{\Sigma X}{N} = \frac{45}{5} = 9$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2} = \sqrt{\frac{455}{5} - (9)^2} = \sqrt{10} = 3.16$$

Ans. Standard deviation = 3.16

Short-Cut Method (Assumed Mean Method)

At times, actual mean may come in fractions. In such cases, calculation of standard deviation may become complicated and somewhat difficult. In such a situation, it is advisable to use short-cut method to simplify the calculations.

The following steps are taken under this method:

Step 1. Take any value of the X in the series as assumed mean (A).

Step 2. Find out deviations of the items from an assumed mean and denote these by *d, i.e.,* $d = X - A$.

Step 3. Calculate the sum of deviations to obtain Σd

Step 4. Square the deviations and denote the total as Σd²

Step 5. Substitute the values of Σd² and Σd in the following formula:

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

Where:

σ = Standard Deviation
Σd = Sum total of deviations from assumed mean
Σd² = Sum total of squares of deviations
N = Number of pair of observations

Example 27 will illustrate the computation of standard deviation by short-cut method:

Example 27. Calculate the Standard Deviation of the data given in *Example 25* by the Short-Cut method (Assumed Mean Method).

Solution:

Calculation of Standard Deviation (Short-Cut Method)

Values (X)	*d = X – A* *A = 7*	*d²*
5	5 – 7 = – 2	4
8	8 – 7 = + 1	1
7 (A)	7 – 7 = 0	0
11	11 – 7 = + 4	16
14	14 – 7 = + 7	49
N = 5	**Σd = 10**	**Σd² = 70**

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} = \sqrt{\frac{70}{5} - \left(\frac{10}{5}\right)^2} = \sqrt{14 - 4} = \sqrt{10} = 3.16$$

Ans. Standard deviation = 3.16

Example 28. Nine students have obtained the following marks in statistics out of 100 marks. Calculate the standard deviation of marks obtained.

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

Solution:

Calculation of Standard Deviation

Values (X)	*d = X – A* *A = 40*	*d²*
5	– 35	1,225
10	– 30	900
20	– 20	400
25	– 15	225
40 (A)	0	0
45	+ 5	25
48	+ 8	64
70	+ 30	900
80	+ 40	1,600
N = 9	**Σd = –17**	**Σd² = 5,339**

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} = \sqrt{\frac{5.339}{9} - \left(\frac{-17}{9}\right)^2}$$

$$\text{Standard Deviation } (\sigma) = \sqrt{593.22 - 3.57} = \sqrt{589.65} = 24.28$$

Ans. Standard Deviation = 24.28

Example 29. Calculate the standard deviation by: (i) Actual Mean Method; (ii) Direct Method; (iii) Short-Cut Method.

3	4	6	7	10

Solution:

Calculation of Standard Deviation

Actual Mean Method			Direct Method		Short-Cut Method		
X	$X-\overline{X}$ (x)	x^2	X	X^2	X	$d=X-A$ $(A=6)$	d^2
3	−3	9	3	9	3	−3	9
4	−2	4	4	16	4	−2	4
6	0	0	6	36	6 (A)	0	0
7	+1	1	7	49	7	+1	1
10	+4	16	10	100	10	+4	16
$\Sigma X = 30$		$\Sigma x^2 = 30$	$\Sigma X = 30$	$\Sigma X^2 = 210$	$N = 5$	$\Sigma d = 0$	$\Sigma d^2 = 30$

Actual Mean Method:

$$\overline{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{30}{5}}$$

$$\sigma = \sqrt{6} = 2.45$$

Direct Method:

$$\overline{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

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

$$= \sqrt{\frac{210}{5} - (6)^2}$$

$$\sigma = \sqrt{6} = 2.45$$

Short-Cut Method:

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

$$\sigma = \sqrt{\frac{30}{5} - \left(\frac{0}{5}\right)^2}$$

$$\sigma = \sqrt{6} = 2.45$$

Ans. Standard Deviation = 2.45

Discrete Series

The standard deviation in discrete series can be calculated by the following four methods:

(i) Actual Mean Method *(Refer Example 30)*
(ii) Direct method *(Refer Example 31)*
(iii) Short-cut method or Assumed Mean Method *(Refer Example 32)*
(iv) Step Deviation method *(Refer Example 33)*

Actual Mean Method

In this method, deviations are taken from the actual mean. The *steps involved in calculation of standard deviation are:*

Step 1. Calculate the actual mean $(\overline{X})$ of the series as: $\overline{X} = \frac{\Sigma fX}{\Sigma f}$

Step 2. Find out deviations of the items from the actual mean, i.e., calculate $(X - \overline{X})$ and denote the deviations by x.

Step 3. Square the deviations and multiply them by their respective frequencies (f) and obtain the total, i.e., Σfx^2

Step 4. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$

Where:

σ = Standard Deviation;
Σfx^2 = Sum total of the squared deviations multiplied by frequency;
N = Number of pair of observations

Example 30 will make the Actual Mean method more clear.

Example 30. Calculate standard deviation by the actual mean method:

Size	5	10	15	20
Frequency	2	1	4	3

Solution:

Calculation of Standard Deviation (Actual Mean Method)

Size (X)	Frequency (f)	fX	$x = X - \overline{X}$	x^2	fx^2
5	2	10	−9	81	162
10	1	10	−4	16	16
15	4	60	+1	1	4
20	3	60	+6	36	108
	$N = \Sigma f = 10$	$\Sigma fX = 140$			$\Sigma fx^2 = 290$

Arithmetic mean $(\overline{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{140}{10} = 14$

Standard Deviation $(\sigma) = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{290}{10}} = \sqrt{29} = 5.38$

Ans. Standard deviation = 5.38

Direct Method

The steps involved in the Direct Method are:

Step 1. Calculate the actual mean $(\overline{X})$ of the series as: $\overline{X} = \frac{\Sigma fX}{\Sigma f}$

Step 2. Square the observations to get X^2

Step 3. Multiply frequency (f) to X^2 and obtain the total, i.e., ΣfX^2

Step 4. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fX^2}{N} - (\overline{X})^2}$

Where:

σ = Standard Deviation

$\bar{X}$ = Actual Mean

ΣfX^2 = Sum total of the squared observations multiplied by frequency

N = Number of pair of observations

Example 31 will illustrate the calculation of standard deviation by using Direct method.

Example 31. Calculate the Standard Deviation of the data given in *Example 30* by the Direct method.

Solution:

Calculation of Standard Deviation (Direct Method)

Size (X)	Frequency (f)	fX	X^2	fX^2
5	2	10	25	50
10	1	10	100	100
15	4	60	225	900
20	3	60	400	1,200
	N = Σf = 10	**ΣfX = 140**		**Σfx² = 2,250**

$$\text{Arithmetic mean } (\bar{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{140}{10} = 14$$

$$\text{Standard deviation } (\sigma) = \sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2}$$

$\Sigma fX^2 = 2{,}250$, N = 10, $\bar{X} = 14$

$$(\sigma) = \sqrt{\frac{2{,}250}{10} - (14)^2} = \sqrt{225 - 196} = 5.38$$

Ans. Standard deviation = 5.38

Short-Cut Method (Assumed Mean Method)

Calculation of standard deviation by short-cut method involves the following steps:

Step 1. Take any value of the X in the series as assumed mean (A).

Step 2. Find out deviations of items from the assumed mean and denote these by d, i.e., d = X – A.

Step 3. Multiply these deviations by the respective frequencies and obtain the total, i.e. Σfd

Step 4. Calculate the square of deviations, i.e. d^2

Step 5. Multiply the squared deviations by respective frequencies and obtain the total to get Σfd^2

Step 6. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

Where:

σ = Standard Deviation

Σfd^2 = Sum total of the squared deviations multiplied by frequency

Σfd = Sum total of deviations multiplied by frequency

N = Number of pair of observations

Let us illustrate this method with the help of Example 32.

Example 32. Calculate the Standard Deviation of the data given in Example 30 by the Short-Cut method (Assumed Mean Method).

Solution:

Calculation of Standard Deviation (Short-Cut Method)

Size (X)	Frequency (f)	d = X – A A = 10	fd	d^2	fd^2
5	2	– 5	– 10	25	50
10 A	1	0	0	0	0
15	4	+ 5	20	25	100
20	3	+ 10	30	100	300
	N = Σf = 10		**Σfd = 40**		**Σfd² = 450**

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

$\Sigma fd^2 = 450$; N = 10 Σfd = 40

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{450}{10} - \left(\frac{40}{10}\right)^2} = \sqrt{45 - 16} = \sqrt{29} = 5.38$$

Ans. Standard deviation = 5.38

Step Deviation Method

The step deviation method involves the following steps:

Step 1. Take any value of the X in the series as assumed mean (A).

Step 2. Find out deviations (d) of the items from an assumed mean.

Step 3. Divide these deviations by common factor (C) to obtain step deviations (d′).

Step 4. Multiply the step deviations by the respective frequencies and obtain the total, i.e. Σfd′

Step 5. Calculate the square of step deviations, i.e. d'^2

Step 6. Multiply these squared step deviations by the respective frequencies and obtain the total to get $\Sigma fd'^2$

Step 7. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$

Where:

σ = Standard Deviation

$\Sigma fd'^2$ = Sum total of the squared step deviations multiplied by frequency

$\Sigma fd'$ = Sum total of step deviations multiplied by frequency

C = Common Factor

N = Number of pair of observations

Example 33 will illustrate the computation of standard deviation by step deviation method:

Example 33. Calculate the Standard Deviation of the data given in Example 30 by the Step deviation method.

Solution:

Calculation of Standard Deviation (Step deviation method)

Size (X)	Frequency (f)	$d = X - A$, $A = 10$	$d' = \frac{X-A}{C}$, $C = 5$	fd'	d'^2	fd'^2
5	2	−5	−1	−2	1	2
10 (A)	1	0	0	0	0	0
15	4	5	+1	4	1	4
20	3	10	+2	6	4	12
	N = Σf = 10			Σfd′ = 8		Σfd′² = 18

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

$\Sigma fd'^2 = 18;\ N = 10;\ \Sigma fd' = 8;\ C = 5$

$(\sigma) = \sqrt{\frac{18}{10} - \left(\frac{8}{10}\right)^2} \times 5 = \sqrt{1.8 - .64} \times 5 = 5.38$

Ans. Standard deviation = 5.38

Example 34. Calculate the standard deviation from the given data by: (i) Actual Mean Method; (ii) Direct Method; (iii) Short-Cut Method; (iv) Step Deviation Method.

X	2	4	6	8
f	3	1	4	2

Solution:

Calculation of Standard Deviation

Actual Mean Method						Direct Method				
X	f	fX	$x = X - \bar{X}$	x^2	fx^2	X	f	fX	X^2	fX^2
2	3	6	−3	9	27	2	3	6	4	12
4	1	4	−1	1	1	4	1	4	16	16
6	4	24	+1	1	4	6	4	24	36	144
8	2	16	+3	9	18	8	2	16	64	128
	N = Σf = 10	ΣfX = 50			Σfx² = 50		N = Σf = 10	ΣfX = 50		ΣfX² = 300

$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{50}{10} = 5$

$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{50}{10}}$

$\sigma = \sqrt{5} = 2.236$

$\bar{X} = \frac{\Sigma fX}{\Sigma f} = \frac{50}{10} = 5$

$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2} = \sqrt{\frac{300}{10} - (5)^2}$

$\sigma = \sqrt{5} = 2.236$

Short-Cut Method						Step Deviation Method				
X	f	$d = X - A$ (A = 4)	fd	d^2	fd^2	$d = X - A$ (A = 4)	$d' = \frac{X-A}{C}$ (C = 5)	fd'	d'^2	fd'^2
2	3	−2	−6	4	12	−2	−1	−3	1	3
4 (A)	1	0	0	0	0	0	0	0	0	0
6	4	+2	+8	4	16	+2	+1	+4	1	4
8	2	+4	+8	16	32	+4	+2	+4	4	8
	N = 10		Σfd = 10		Σfd² = 60			Σfd′ = 5		Σfd′² = 15

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

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

$\sigma = \sqrt{5} = 2.236$

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

$\sigma = \sqrt{\frac{15}{10} - \left(\frac{10}{10}\right)^2} \times 2$

$\sigma = \sqrt{1.25} \times 2 = 2.236$

Continuous Series

The method of calculating standard deviation in a continuous series, is the same as it is in a discrete series, because classes are represented by their mid-values. Standard deviation in continuous series can also be calculated by following four methods:

(i) Actual Mean Method *(Refer Example 35)*

(ii) Direct method *(Refer Example 36)*

(iii) Short-cut method or Assumed Mean Method *(Refer Example 37)*

(iv) Step Deviation method *(Refer Example 38)*

Actual Mean Method

The steps involved in calculation of standard deviation are:

Step 1. Calculate the actual mean ($\bar{X}$) of the series as: $\bar{X} = \frac{\Sigma fm}{\Sigma f}$

Step 2. Find out deviations of the mid-points (m) from the actual mean, i.e., calculate $(m - \bar{X})$ and denote the deviations by x.

Step 3. Square the deviations and multiply them by their respective frequencies (f) and obtain the total, i.e., Σfx^2

Step 4. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$

Where:

σ = Standard Deviation

Σfx^2 = Sum total of the squared deviations multiplied by frequency

N = Number of pair of observations

Example 35 will make the Actual Mean method more clear.

Example 35. Calculate standard deviation by the actual mean method:

Marks	0–10	10–20	20–30	30–40	40–50
No. of students	4	3	6	5	2

Solution:

Calculation of Standard Deviation (Actual Mean Method)

Marks (X)	No. of Students (f)	Mid-point (m)	fm	$x = m - \overline{X}$	x^2	fx^2
0–10	4	5	20	– 19	361	1,444
10–20	3	15	45	– 9	81	243
20–30	6	25	150	+ 1	1	6
30–40	5	35	175	+ 11	121	605
40–50	2	45	90	+ 21	441	882
	N = Σf = 20		Σfm = 480			Σfx^2 = 3,180

Arithmetic Mean ($\overline{X}$) $= \frac{\Sigma fm}{\Sigma f} = \frac{480}{20} = 24$

Standard Deviation (σ) $= \sqrt{\frac{\Sigma fx^2}{N}}$

Σfx^2= 3,180 and N = 20

$\sigma = \sqrt{\frac{3,180}{20}} = \sqrt{159} = 12.61$ Marks

Ans. Standard deviation = 12.61 marks

Direct Method

The steps involved in the Direct Method are:

Step 1. Calculate the actual mean ($\overline{X}$) of the series as: $\overline{X} = \frac{\Sigma fm}{\Sigma f}$

Step 2. Square the mid-points to get m^2

Step 3. Multiply frequency (f) to m^2 and obtain the total, i.e., Σfm^2

Step 4. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fm^2}{N} - (\overline{X})^2}$

Where:

σ = Standard Deviation

$\overline{X}$ = Actual Mean

Σfm^2 = Sum total of the squared mid-points multiplied by frequency

N = Number of pair of observations

Example 36 will illustrate the calculation of standard deviation by using Direct method.

Example 36. Calculate the Standard Deviation for data given in Example 35 by the Direct method.

Solution:

Calculation of Standard Deviation (Direct Method)

Marks (X)	No. of students (f)	Mid-point (m)	fm	m^2	fm^2
0–10	4	5	20	25	100
10–20	3	15	45	225	675
20–30	6	25	150	625	3,750
30–40	5	35	175	1,225	6,125
40–50	2	45	90	2,025	4,050
	N = Σf = 20		Σfm = 480		Σfm^2 = 14,700

Arithmetic Mean ($\overline{X}$) $= \frac{\Sigma fm}{\Sigma f} = \frac{480}{20} = 24$

Standard Deviation (σ) $= \sqrt{\frac{\Sigma fm^2}{N} - (\overline{X})^2}$

Σfm^2= 14,700, N = 20, Σfm = 480

$\sigma = \sqrt{\frac{14,700}{20} - (24)2} = \sqrt{735 - 576} = \sqrt{159} = 12.61$ Marks

Ans. Standard deviation = 12.61 Marks

Short-Cut Method (Assumed Mean Method)

The short-cut method involves the following steps:

Step 1. Take any value of the mid-point in the series as assumed mean (A).

Step 2. Find out deviations of mid-points from the assumed mean and denote it by d, i.e., d = m – A.

Step 3. Multiply these deviations by the respective frequencies and obtain the total, i.e. Σfd.

Step 4. Calculate the square of deviations, i.e., d^2

Step 5. Multiply the squared deviations by respective frequencies and obtain the total to get Σfd^2

Step 6. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$

Where:

σ = Standard Deviation

Σfd^2 = Sum total of the squared deviations multiplied by frequency

Σfd = Sum total of deviations multiplied by frequency

N = Number of pair of observations

Short-Cut method will be clear with the help of Example 37.

Example 37. Calculate Standard deviation for the data given in *Example 35* by the Short-Cut method.

Solution:

Calculation of Standard Deviation (Short-cut Method)

Marks (X)	No. of students (f)	Mid-point (m)	d = m − A (A = 25)	fd	d^2	fd^2
0–10	4	5	− 20	− 80	400	1,600
10–20	3	15	− 10	− 30	100	300
20–30	6	25 (A)	0	0	0	0
30–40	5	35	+ 10	+ 50	100	500
40–50	2	45	+ 20	+ 40	400	800
	N = Σf = 20			Σfd = −20		Σfd^2 = 3,200

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

$\Sigma fd^2 = 3{,}200$, $N = 20$, $\Sigma fd = -20$

$(\sigma) = \sqrt{\frac{3{,}200}{20} - \left(\frac{-20}{20}\right)^2} = \sqrt{160 - 1} = \sqrt{159} = 12.61$ Marks

Ans. Standard deviation = 12.61 Marks

Step Deviation Method

The steps involved in calculating standard deviation by step deviation are:

Step 1. Take any mid-point (m) in the series as assumed mean (A).

Step 2. Find out deviations (d) of the mid-point from the assumed mean.

Step 3. Divide these deviations by common factor (C) to obtain step deviations (d').

Step 4. Multiply step deviations by respective frequencies and obtain the total, i.e. $\Sigma fd'$

Step 5. Calculate the square of step deviations, i.e., d'^2

Step 6. Multiply these squared step deviations by the respective frequencies and obtain the total to get $\Sigma fd'^2$

Step 7. Apply the following formula: $\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$

Where:

σ = Standard Deviation

$\Sigma fd'^2$ = Sum total of the squared step deviations multiplied by frequency

$\Sigma fd'$ = Sum total of step deviations multiplied by frequency

C = Common Factor

N = Number of pair of observations

The computation of standard deviation by step deviation method will be more clear by Example 38:

Example 38. Calculate the Standard deviation of the data given in *Example 35* by the Step deviation method.

Solution:

Calculation of Standard Deviation (Step deviation method)

Marks (X)	No. of Students (f)	Mid-point (m)	d = m − A (A = 25)	$d' = \frac{m-A}{C}$ C = 10	fd'	d'^2	fd'^2
0–10	4	5	− 20	− 2	− 8	4	16
10–20	3	15	− 10	− 1	− 3	1	3
20–30	6	25 (A)	0	0	0	0	0
30–40	5	35	+ 10	+ 1	+ 5	1	5
40–50	2	45	+ 20	+ 2	+ 4	4	8
	N = Σf = 20				Σfd' = −2		$\Sigma fd'^2$ = 32

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

$\Sigma fd'^2 = 32$; $N = 20$; $\Sigma fd' = -2$; $C = 10$

$(\sigma) = \sqrt{\frac{32}{20} - \left(\frac{-2}{20}\right)^2} \times 10 = \sqrt{1.6 - .01} \times 10 \; \sqrt{1.59} \times 10 = 12.61$ Marks

Ans. Standard deviation = 12.61 Marks

Example 39. Calculate standard deviation by the actual mean method:

Class	1–3	3–5	5–7	7–9	9–11	11–13	13–15
Frequency	1	9	25	35	17	10	3

Solution:

Calculation of Standard Deviation (Actual mean Method)

Class (X)	Frequency (f)	Mid-points (m)	fm	$x = m - \bar{X}$	x^2	fx^2
1–3	1	2	2	– 6	36	36
3–5	9	4	36	– 4	16	144
5–7	25	6	150	– 2	4	100
7–9	35	8	280	0	0	0
9–11	17	10	170	+ 2	4	68
11–13	10	12	120	+ 4	16	160
13–15	3	14	42	+ 6	36	108
	N = Σf = 100		Σfm = 800			$\Sigma fx^2 = 616$

Arithmetic Mean $(\bar{X}) = \frac{\Sigma fm}{\Sigma f} = \frac{800}{100} = 8$

Standard Deviation $(\sigma) = \sqrt{\frac{\Sigma fx^2}{N}}$

$\Sigma fx^2 = 616$ and N = 100

$\sigma = \sqrt{\frac{616}{100}} = \sqrt{6.16} = 2.48$

Ans. Standard deviation = 2.48

Example 40. Calculate standard deviation from the following data:

Marks (more than)	0	10	20	30	40	50	60	70
No. of students	100	90	75	50	25	15	5	0

Solution:

Since we are given the cumulative frequencies, we first find the simple frequency

Calculation of Standard Deviation

Marks (X)	No. of students (f)	Mid-points (m)	d = m – A A = 25	$d' = \frac{m-A}{C}$ C = 10	fd'	d'^2	fd'^2
0–10	10	5	– 20	– 2	– 20	4	40
10–20	15	15	– 10	– 1	– 15	1	15
20–30	25	25 (A)	0	0	0	0	0
30–40	25	35	+ 10	+ 1	+ 25	1	25
40–50	10	45	+ 20	+ 2	+ 20	4	40
50–60	10	55	+ 30	+ 3	+ 30	9	90
60–70	5	65	+ 40	+ 4	+ 20	16	80
	N = Σf = 100				Σfd' = 60		$\Sigma fd'^2 = 290$

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

$\Sigma fd'^2 = 290$, N = 100, Σfd' = 60, C = 10

$(\sigma) = \sqrt{\frac{290}{100} - \left(\frac{60}{100}\right)^2} \times 10 = \sqrt{2.9 - .36} \times 10 = \sqrt{2.54} \times 10 = 15.94$ Marks

Ans. Standard deviation = 15.94 marks

Example 41. Calculate the standard deviation of the following series:

Expenditure (Below ₹)	5	10	15	20	25
No. of students	6	16	28	38	46

Solution:

In the given example, we are given the cumulative frequencies. So, we will first calculate the simple frequency

Calculation of Standard Deviation

Expenditure (₹)	No. of Students (f)	Mid-points (m)	d = m – A (A = 12.5)	$d' = \frac{m-A}{C}$ C = 5	fd'	d'^2	fd'^2
0–5	6	2.5	– 10	– 2	– 12	4	24
5–10	10	7.5	– 5	– 1	– 10	1	10
10–15	12	12.5 (A)	0	0	0	0	0
15–20	10	17.5	+ 5	+ 1	+ 10	1	10
20–25	8	22.5	+ 10	+ 2	+ 16	4	32
	N = Σf = 46				Σfd' = 4		$\Sigma fd'^2 = 76$

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

$\Sigma fd'^2 = 76$, N = 46, Σfd' = 4, C = 5

$(\sigma) = \sqrt{\frac{76}{46} - \left(\frac{4}{46}\right)^2} \times 5 = \sqrt{1.65 - .007} \times 5 = \sqrt{1.643} \times 5 = 6.41$

Ans. Standard deviation = ₹ 6.41

Example 42. Calculate the standard deviation from the following data by: (i) Actual Mean; Method; (ii) Direct Method; (iii) Short-Cut Method; (iv) Step Deviation Method.

X	0–10	10–20	20–30	30–40
f	2	3	4	1

Solution:

Calculation of Standard Deviation

m = mid-point

Actual Mean Method							Direct Method				
X	f	m	fm	$x = m - \bar{X}$	x^2	fx^2	f	m	fm	m^2	fm^2
0–10	2	5	10	– 14	196	392	2	5	10	25	50
10–20	3	15	45	– 4	16	48	3	15	45	225	675
20–30	4	25	100	+ 6	36	144	4	25	100	625	2,500
30–40	1	35	35	+ 16	256	256	1	35	35	1,225	1,225
	N = 10		Σfm = 190			Σfx² = 840	N = 10		Σfm = 190		Σfm² = 4,450

Actual Mean Method:

$$\bar{X} = \frac{\Sigma fm}{\Sigma f} = \frac{190}{10} = 19$$

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} = \sqrt{\frac{840}{10}}$$

$$\sigma = \sqrt{84} = 9.165$$

Direct Method:

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

$$\sigma = \sqrt{\frac{4{,}450}{10} - (19)^2}$$

$$\sigma = \sqrt{84} = 9.165$$

m = mid-points; **d** = m – A where, A = 15; $d' = \frac{m - A}{C}$ where, C = 10

Short-Cut Method							Step Deviation Method						
X	f	m	d	fd	d^2	fd^2	f	m	d	d'	fd'	d'^2	fd'^2
0–10	2	5	– 10	– 20	100	200	2	5	– 10	– 1	– 2	1	2
10–20	3	15 (A)	0	0	0	0	3	15 (A)	0	0	0	0	0
20–30	4	25	+ 10	+ 40	100	400	4	25	+ 10	+ 1	+ 4	1	4
30–40	1	35	+ 20	+ 20	400	400	1	35	+ 20	+ 2	+ 2	4	4
	N = 10			Σfd = 40		Σfd² = 1,000	N = 10				Σfd' = 4		Σfd'² = 10

Short-Cut Method:

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

$$\sigma = \sqrt{\frac{1{,}000}{10} - \left(\frac{40}{10}\right)^2}$$

$$\sigma = \sqrt{84} = 9.165$$

Step Deviation Method:

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

$$\sigma = \sqrt{\frac{10}{10} - \left(\frac{4}{10}\right)^2} \times 10$$

$$\sigma = \sqrt{0.84} \times 10 = 9.165$$

10.19 VARIANCE

Variance is another measure based on standard deviation. *By variance, we mean the square of the standard deviation.* The term was first used by R.A. Fisher in 1913.

Symbolically,

$$\text{Variance} = \sigma^2$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\text{Variance}}$$

Smaller the value of variance (σ), lesser is the variability or greater the consistency and vice-versa.

Standard deviation and variance are measures of variability and they are closely related. The only difference between the two is that standard deviation is the square root of variance and variance is the average squared deviations from mean.

10.20 RELATIVE MEASURES OF STANDARD DEVIATION

Standard deviation is an absolute measure of dispersion. As a result, it cannot be used to compare variability of two or more series, when the two series are expressed in different units. *Thus, for comparing the dispersions of two or more series with different units, it is necessary to compute the relative measures of standard deviation.*

Let us discuss some of the important relative measures:

Coefficient of Standard Deviation

It is computed by dividing standard deviation (σ) by the mean ($\bar{X}$) of the data.

$$\text{Coefficients of Standard Deviation } (\sigma) = \frac{\sigma}{\bar{X}}$$

It is also known as 'standard coefficient of dispersion'.

Coefficient of Variation

This measure was introduced by Karl Pearson. So, it is also known as 'Karl Pearson's Coefficient of Variation'. ***When two or more groups of similar data are to be compared with respect to stability (or uniformity or consistency or homogeneity), coefficient of variation is the most appropriate measure.***

It indicates the relationship between the standard deviation and the arithmetic mean expressed in terms of percentage.

$$\text{Coefficients of Variation (C.V.)} = \frac{\sigma}{\bar{X}} \times 100$$

Where: C.V. = Coefficient of Variation; σ = Standard Deviation; $\bar{X}$ = Arithmetic Mean

The series for which coefficient of variation is greater is said to be more variable or conversely less stable, less uniform, less consistent, less homogeneous. *Suppose, we want to compare stability in runs of two or more batsmen over a period of time, then we should calculate the coefficient of variation. The batsman having the least value of coefficient of variation is taken as the most stable batsman.*

10.21 MISCELLANEOUS PRACTICALS

Example 43. The number of goals scored by two teams in a football session were as under:

No. of goals scored	0	1	2	3	4	5
No. of Matches (Team A)	15	10	7	5	3	2
No. of Matches (Team B)	20	10	5	4	2	1

Which team is more consistent?