" Uncontrolled variation is the enemy of quality."

– Edward Deming

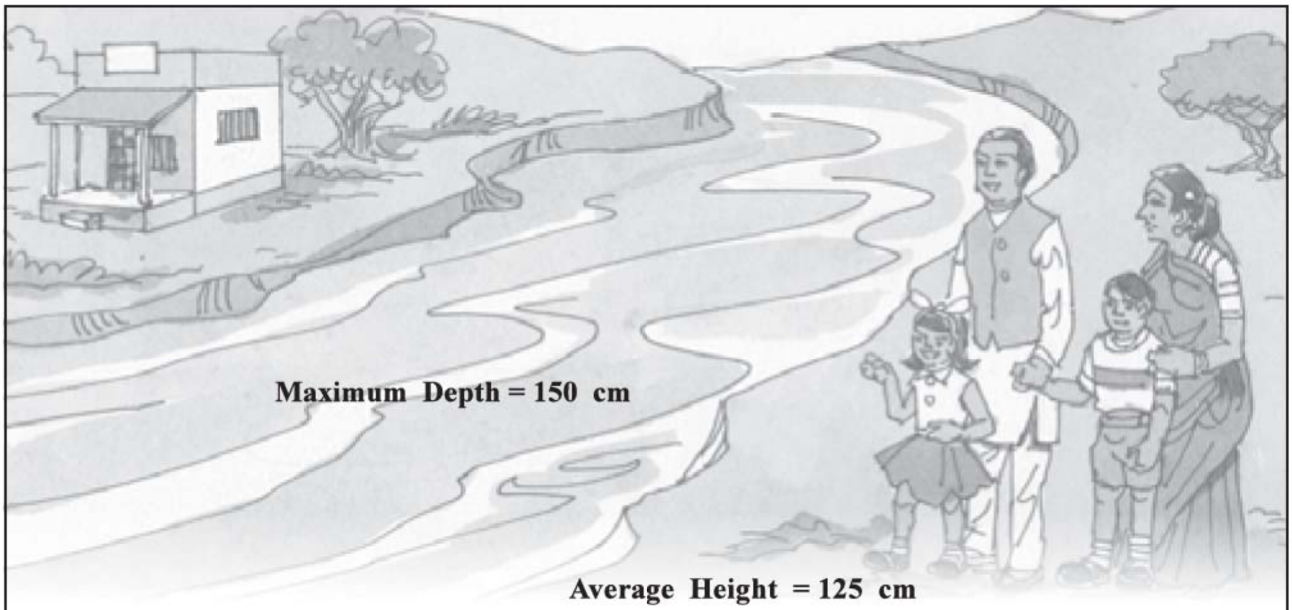
4

Measures of Dispersion

Contents :

4.1 Meaning of Dispersion

We have studied classification, tabulation and the measures of central tendency of the collected data in the preceding three chapters. As we know now, any measures of central tendency or averages of data represent summary or central value of the data. It may happen that some observations of the data are very near to the central value whereas some are very far. So, it is useful to know how the observations are scattered from the central values of the population. Though measures of central tendency are very useful in statistical analysis, only these measures are not sufficient. Let us understand it by the following figure and its details.



A person once visited a forest with his family. The family had to cross a river in the forest for further journey and boat was not available. He knew that the average depth of the river was l00 cms. He knew that the average height of his family was 125 cms. He, therefore thought that he and his family could cross the river on foot without any risk. But the maximum depth of the river was 150 cms and height of his daughter was only 90 cms. So, we can imagine what would happen if they tried to cross the river.

Thus, it is clear that without the knowledge of variability, only average (measures of central tendency) does not serve the purpose.

Similarly, let us take another example of average income of the people in a country, popularly known as per capita income. It is one of the key indicators, showing economic condition of the people of that country. But it can be easily understood that it does not throw any light on the distribution of income among various groups of people of the country. So, it is not possible to know inequality of income between the rich and the poor by just having average income (per capita income).

So, to study the data, we should know various characteristics of it. The measures of central tendency tell us only a part of what we need to know, but we should measure the spread or variability of the data for its better understanding. The measure of variability, along with the measures of central tendency gives such information. ln this chapter, we shall consider the different measures of variation based on internal variation of the observations and the scattering of the observations of data from the mean as measure of average.

We have experienced that two or more groups having identical average may differ from each other in certain aspects. The spread or scattering of the observations from the average and the internal variations among individual observations could be different. Thus, instead of comparing groups only on the basis of averages, it is advisable to consider the variation within observations of each group. We now illustrate this fact by an example.

Suppose a financial analyst wants to study the performance of three companies A, B and C. He gets the following information of the profits of these three companies for last 5 years :

Year	1	2	3	4	5	Total
Profit of A (lakhs ₹)	30	30	30	30	30	150
Profit of B (lakhs ₹)	15	30	30	30	45	150
Profit of C (lakhs ₹)	–5	30	70	30	25	150

Now it is very obvious that financial analyst will first try to know the average annual profits of these three companies and then will try to know the fluctuations in them.

It is very clear from the data of profits of the three companies that for all three companies A, B and C their mean = median = mode = 30 (lakhs ₹). Now, observing individual yearly profits of three companies A, B and C, we can see that the profits of company A remain same in last five years, so the spread in its profit is 30 – 30 = 0; the spread in the profits of company B is 45 – 15 = 30 (lakhs ₹), whereas spread of the profits of company C is 70 – (–5) = 75 (lakhs ₹). Here, there is no variation in the profits for company A, as it remained same for all five years. The yearly profits of company B for last 5 years are nearer to its average of 30 (lakhs ₹); but for company C, its yearly profits are very far from its average profit of 30 (lakhs ₹). Hence though the mean, the median and the mode of these three companies are same, their profits differ significantly in terms of the variation. Thus, analyzing these three companies simply on the basis of equality of their averages may lead to erroneous and misleading conclusion that the three companies are same in terms of their profits.

So, the information about scatter or spread of the observations of a population is necessary not only for study of characteristics of that population but it is also necessary for comparative statistical study of two or more populations.

A measure which shows how far the observations of the data are scattered from the measure of average is termed as **dispersion.**

However, the term dispersion not only gives a general impression about the variability of a population but also a precise measure of this variation. There are several definitions available which are stated by different statisticians. One of the definitions given by **Spiegel** is as follows :

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

By having additional information of dispersion, the reliability of the measures of central tendency can be judged. If data are widely spread, the average represents only those observations which are nearer to the mean (average) value rather than representing all observations of the data.

Desirable Characteristics of a measure of dispersion :

The following are some desirable characteristics of a measure of dispersion.

(1) The definition of a measure of dispersion should be clear and unambiguous.

(2) It should be simple to understand and easy to compute.

(3) It should be based on all observations of the data.

(4) It should be suitable for further algebraic and statistical calculations.

(5) It should be a stable measure in the sense that if different samples of equal size are drawn and measures of dispersion are obtained, the values of these measures should be almost same for all samples.

(6) It should not be unduly affected by very small or very large observations of the data.

4.2 Concept of Absolute and Relative measures of dispersion

Absolute Measure :

A measure of dispersion which is expressed in the same (Statistical) units in which the observations of the data are expressed is called an **absolute measure** of dispersion. For example, if the original data are in kilogram, an absolute measure will also be in kilogram.

An absolute measure is not useful for comparison purpose for variability of two or more sets of data having different units of the measurement. Let us understand this from the following example :

Suppose weights (in kg) and heights (in cms) of students of a school are given. To know, which characteristic is having more variation, we obtain absolute measure of dispersion. Now, the unit of absolute measure of dispersion of weight is in kg. which is different from that of height which is in cms. So, it is not possible to compare their variability using absolute measures.

Relative Measure :

A measure of dispersion which is free from the unit of measurement is called **relative measure** of dispersion. The variability of two or more sets of data having different units of measurement can be compared only by relative measure of dispersion.

Generally, a measure of relative dispersion is the ratio of a measure of an absolute dispersion of observations of a data set to an appropriate average of the obselvations. The relative measure of dispersion is known as coefficient of dispersion as it is independent of units of measurement of the observations of data.

4.3 Measures of Dispersion

We shall study the following absolute and relative measures of dispersion :

(1) Range

(2) Quartile Deviation

(3) Mean Deviation

(4) Standard Deviation

From the above measures, range and quartile deviation are known as positional measures of dispersion as these measures depend on the positions of observations of the data arranged in increasing order of their magnitudes, while the mean deviation and standard deviation are known as the measures of dispersion representing the summary of deviations of observations from the measure of central tendency.

4.3.1 Range

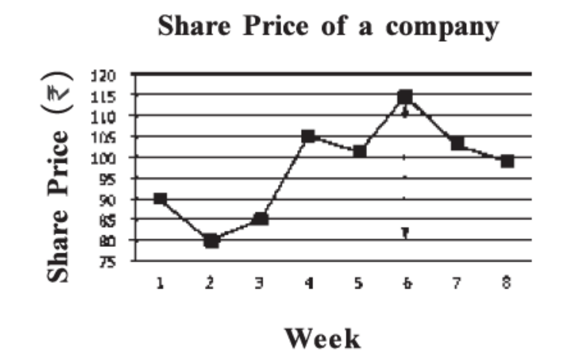
The difference between the highest and lowest observation of the data is called the Range and it is denoted by the symbol R.

$\therefore$ Range $R = x_H - x_L$

where x_H = the highest observation

x_L = the lowest observation

Range R is an absolute measure of dispersion having same unit of measurement as that of the observations.



The data of weekly closing share prices of a company for 8 weeks are given.

From the graph, it is clear that the maximum price is ₹ 115 and its minimum price is ₹ 80. So the range will be 115 – 80 = ₹ 35.

It is obvious from the definition of range that frequency does not play any role in determining the range, even for grouped data. For any grouped frequency distribution, the range can be obtained as the difference between the upper limit of the highest class interval and the lower limit of the lowest class interval.

If we divide the range R of the data by the sum $x_H + x_L$, we get the relative range.

$$\therefore \quad \text{Relative Range} = \frac{R}{x_H + x_L} = \frac{x_H - x_L}{x_H + x_L}$$

The relative range is also known as **coefficient of range.** It is free from the unit of measurement.

If the coefficient of range for a population is small, then it can be said that variability is less in the observations of the population i.e. the values of the observations are not far from each other. But if the coefficient of range is high then it can be said that variability is more in the observations of the population, i.e. the values of the observations are very far from each other.

Illustration 1 : The runs scored by a batsman in his last 10 innings of cricket matches are 48, 75, 37, 52, 93, 81, 25, 72, 18 and 60. Find the range and the coefficient of range of his runs.

Here, $x_H = 93$, $x_L = 18$

Hence Range $= x_H - x_L = 93 - 18 = 75$

$\therefore$ $R = 75$ runs

Thus, the range of runs scored in last ten innings is 75 runs.

$$\text{Coefficient of range} = \frac{R}{x_H + x_L}$$

$$= \frac{75}{93+18} = \frac{75}{111} = 0.6757$$

$\therefore$ Coefficient of range ≈ 0.68

Thus, the coefficient of range of runs is 0.68.

Illustration 2 : From the following information of monthly salary (in ₹) of workers of a factory, find the range and the coefficient of range of the salary.

Monthly Salary (₹)	3500	4000	5000	7500	10,000	12,000
No. of workers	3	21	30	19	6	5

The variable (monthly salary) of the frequency distribution is discrete and from the observations of salary it is clear that $x_H = 12{,}000$ and $x_L = 3500$

Range $= x_H - x_L$

$= 12{,}000 - 3500$

$= 8500$

$\therefore \quad R =$ ₹ 8500

Thus, the range of the monthly salary of workers is ₹ 8500.

$$\text{Coefficient of range} = \frac{R}{x_H + x_L}$$

$$= \frac{8500}{12000+3500}$$

$$= \frac{8500}{15500}$$

$$= 0.5484$$

$\therefore \quad$ Coefficient of range $\doteq 0.55$

Thus, the coefficient of range of the monthly salary of workers is 0.55.

Illustration 3 : The items produced in a factory are packed into different boxes according to their weight. Using the following information, find the range and the relative range of weight of boxes :

Weight (kg)	10 - 15	15 - 20	20 - 25	25 - 30	30 - 35
No. of boxes	8	15	26	47	4

The given frequency distribution is continuous. The upper limit of the last class and lower limit of the first class will be the highest and lowest observations of the data respectively.

Hence $x_H = 35$ and $x_L = 10$

Range $= x_H - x_L$

$= 35 - 10$

$= 25$

$\therefore \quad R = 25$ kgs.

Thus, range of the weight of the boxes is 25 kgs.

$$\text{Relative range} = \frac{R}{x_H + x_L}$$

$$= \frac{25}{35+10}$$

$$= \frac{25}{45}$$

$$= 0.5556$$

$\therefore \quad$ Relative range $\doteq 0.56$

Thus, relative range of the weight of the boxes is 0.56.

Illustration 4 : Find the range and the coefficient of range of the marks from the following frequency distribution of marks scored by 50 students of a school in a certain examination.

Marks	50 - 59	60 - 69	70 - 79	80 - 89	90 - 99
No. of Students	2	15	23	6	4

The upper limit of the last class in 99 and the lower limit of the first class is 50.

Hence, $x_H = 99$ and $x_L = 50$

Range $= x_H - x_L = 99 - 50 = 49$

$\therefore$ $R = 49$ Marks

Thus, range of the marks of students is 49 marks.

$$\text{Coefficient of range} = \frac{R}{x_H + x_L}$$

$$= \frac{49}{99+50}$$

$$= \frac{49}{149}$$

$$= 0.3289$$

$\therefore$ Coefficient of range ≈ 0.33

Thus, coefficient of range of marks obtained by students is 0.33.

Advantages and Disadvantages of Range

Advantages :

(1) The range is very clearly defined.

(2) Its computation is simple.

(3) Range is a useful measure especially when variability among the observations of the data is less.

Disadvantages :

(1) All observations of the data are not used in the computation of range.

(2) Range is very sensitive about sampling fluctuations.

(3) It is not a suitable measure for algebraic operations.

(4) It cannot be calculated for the frequency distribution having open-ended classes.

Note :

Range is useful in statistical quality control for the construction of control charts which measure the variability within the samples taken from the production. If variation is not very large then range is useful in measuring variation in money rates, exchange rates, share prices etc. In our day-to-day problems like 'daily sales in a supermarket', 'temperature of a city', 'expense of petrol by two wheeler or car', etc. are generally expressed in the form of the interval in which it lies, and range of the data can be known from it.

Activity

Collect the information about height and weight of boys and girls whose age is from 15 to 25 years and obtain the interval of height and weight from it and find range. Find the relative range for height and weight and compare them.

EXERCISE 4.1

1. The following data refer to the heights in cms. of 10 students of a class. Find range and coefficient of range of height of the students :

   162, 145, 170, 181, 167, 151, 175, 185, 169, 156

2. A bus company has 77 buses for travelling in the city. The information of number of passengers in bus on a particular day at a particular time is given below. Find the range and coefficient of range of number of passengers.

No. of Passengers	2	7	10	18	25	30	37
No. of Buses	1	4	11	17	23	16	5

3. Using the following frequency distribution of marks of students of a school, find range and relative range of the marks.

Marks	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
No. of Students	8	20	25	60	45	10

4. The frequency distribution of daily income (in thousand ₹) of 80 shops of an area is as follows. Find the absolute and the relative measure of range of daily income from it.

Daily income (thousand ₹)	5-9	10-14	15-19	20-24	25-29	30-34
No. of shops	11	20	17	13	12	7

*

4.3.2 Quartile Deviation

We know that only extreme observations i.e. the largest and the smallest observations are used in the computation of range. Similarly by using positional measures known as first quartile Q_1 and third quartile Q_3, another measure of dispersion namely quartile deviation is obtained. The measure of variation defined by using the middle 50% of the observations arranged in increasing order of magnitudes is called the **quartile deviation.**

The measure of quartile deviation is obtained by dividing the difference between Q_3 and Q_1 by 2. Symbolically, it is denoted by Q_d. We can write the formula of quartile deviation as follows.

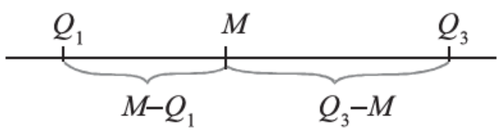
$$Q_d = \frac{Q_3 - Q_1}{2}$$

The quartile deviation is also known as **semi-inter-quartile range.**

Additional Information for Understanding

$(Q_3 - Q_1)$ is called inter-quartile range, but mostly it is reduced to the semi-inter- quartile range which is nothing but the mid-point of inter-quartile range.

Quartile deviation shows the average value by which two quartile (i.e. Q_1 and Q_3) differ from the median. It is clear from the following figure.



$$\text{Quartile Deviation} = \frac{(Q_3 - M) + (M - Q_1)}{2} = \frac{Q_3 - Q_1}{2}$$

If Q_d is divided by the mean of Q_1 and Q_3 we get the relative measure of quartile deviation.

$\therefore$ Relative Quartile Deviation = $\dfrac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

The relative quartile deviation is also known as **coefficient of quartile deviation.** Note that Q_d has unit of measurement in which the observations are expressed but coefficient of quartile deviation is independent of such unit i.e coefficient of quartile deviation is unit free measure.

Illustration 5 : A bus operator gets the following number of passengers in 10 trips on a day. Find the quartile deviation and the coefficient of quartile deviation of number of passengers from the following data :

19, 25, 35, 10, 24, 8, 12, 5, 20, 30

Arranging the observations of given data in increasing order, we get

5, 8, 10, 12, 19, 20, 24, 25, 30, 35

Here, $n = 10$, $\dfrac{n-1}{4} = 2.75$ and $3\left(\dfrac{n-1}{4}\right) = 8.25$

Here, Quartile Q_1 = Value of the $\left(\dfrac{n+1}{4}\right)$ th observation

= Value of the 2.75th observation

Hence, $Q_1 = 8 + 0.75(10 - 8)$

$= 8 + 1.5$

$\therefore \quad Q_1 = 9.5$ passengers

Third Quartile Q_3 = Value of the $3\left(\dfrac{n+1}{4}\right)$th observation

= Value of the 8.25th observation

Hence, $Q_3 = 25 + 0.25(30 - 25)$

$= 25 + 1.25$

$\therefore \quad Q_3 = 26.25$ passengers

Quartile deviation $Q_d = \dfrac{Q_3 - Q_1}{2}$

$= \dfrac{26.25 - 9.5}{2}$

$= 8.38$

$\therefore \quad Q_d = 8.38$ passengers

The quartile deviation of the passengers is 8.38 passengers.

Coefficient of quartile deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \dfrac{16.75}{26.25 + 9.5}$

$= \dfrac{16.75}{35.75}$

$= 0.4685$

$\therefore$ Coefficient of quartile deviation ≈ 0.47

The coefficient of quartile deviation for the passengers is 0.47.

Illustration 6 : The information regarding time (in minutes) taken to solve a puzzle by 50 students is given below. Find the quartile deviation and the coefficient of quartile deviation of the time taken to solve a puzzle by the children from it.

Times (minutes)	2	4	6	8	10
No. of children	3	12	18	12	5

Time (minutes) x	**No. of students** f	**Cumulative frequency** cf
2	3	3
4	12	15
6	18	33
8	12	45
10	5	50
Total	n **= 50**	

Here, $n = 50$, $\frac{n+1}{4} = 12.75$, $3\left(\frac{n+1}{4}\right) = 38.25$

First Quartile Q_1 = Value of the $\left(\frac{n+1}{4}\right)$ th observation

= Value of the 12.75th observation

$\therefore \quad Q_1 = 4$ minutes

Third Quartile Q_3 = Value of the $3\left(\frac{n+1}{4}\right)$th observation

= Value of the 38.25 th observation

$\therefore \quad Q_3 = 8$ minutes

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

$= \frac{8-4}{2}$

$= 2$

$\therefore \quad Q_d = 2$ minutes

The quartile deviation of the time taken to solve a puzzle by the student is 2 minutes.

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{4}{8+4}$

$= \frac{4}{12}$

$= 0.3333$

$\therefore$ Coefficient of quartile deviation ≈ 0.33

Thus, the coefficient of quartile deviation of the time taken to solve a puzzle is 0.33.

Illustration 7 : Using the following distribution of income of 1000 persons of a city, calculate the quartile deviation and coefficient of quartile deviation of income of the persons :

Income (thousand ₹)	Less then 50	50 - 70	70 - 90	90 - 110	110 - 130	130 - 150	above 150
No. of persons	54	100	140	300	230	125	51

Income (Thousand ₹)	**No. of Persons** f	**Cumulative freuquency** cf
Less than 50	54	54
50 - 70	100	154
70 - 90	140	294
90 - 110	300	594
110 - 130	230	824
130 - 150	125	949
above 150	51	1000
Total	n **= 1000**	–

Here, $n = 1000$, $\frac{n}{4} = 250$ and $3\left(\frac{n}{4}\right) = 750$

First Quartile Q_1 = Value of the $\left(\frac{n}{4}\right)$th observation

= Value of the 250 th observation

Referring to the column of cumulative frequencies (cf), we see that the 250 th observation lies in the class 70 - 90. Hence, the Q_1 -class is 70 - 90.

First Quartile $Q_1 = L + \frac{\frac{n}{4} - cf}{f} \times c$

Here, $L = 70$, $\frac{n}{4} = 250$, $cf = 154$, $f = 140$, $c = 20$

Hence, $Q_1 = 70 + \frac{250 - 154}{140} \times 20$

$= 70 + \frac{1920}{140}$

$= 70 + 13.7143$

$= 83.7143$

$\therefore \quad Q_1 \approx 83.71$ (thousand ₹)

Third deviation Q_3 = Value of the $3\left(\frac{n}{4}\right)$th observation

= Value of the 750th observation

Referring to the column of cumulative frequencies (cf), we see that the 750th observation lies in the class 110 – 130. Hence the Q_3 class is 110 – 130.

$$\text{Third deviation } Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$$

Here, $L = 110$, $3\left(\frac{n}{4}\right) = 750$, $cf = 594$, $f = 230$, $c = 20$

$$\text{Hence, } Q_3 = 110 + \frac{750 - 594}{230} \times 20$$

$$= 110 + \frac{3120}{230}$$

$$= 110 + 13.5652$$

$$= 123.5652$$

$\therefore$ $Q_3 \approx 123.57$ (thousand ₹)

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

$$= \frac{123.57 - 83.71}{2}$$

$$= \frac{39.86}{2}$$

$\therefore$ $Q_d = 19.93$

Thus, the quartile deviation of income of the persons is 19.93 (thousand ₹).

$$\text{Coefficient of quartile deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{123.57 - 83.71}{123.57 + 83.71}$$

$$= \frac{39.86}{207.28}$$

$$= 0.1923$$

$\therefore$ Coefficient of quartile deviation ≈ 0.19

Thus, coefficient of quartile deviation of income of the persons is 0.19.

Advantages and Disadvantages of Quartile Deviation

Advantages :

(1) The quartile deviation is a clearly defined measure of dispersion.

(2) Its computation is simple.

(3) Its value is not affected by unusally small or large observations of the data, as only middle 50% observations are taken into account for the computation of the quartile deviation.

(4) It is the only measure of dispersion which can be computed if frequency distribution has open-ended classes.

Disadvantages :

(1) The first 25% and the last 25% observations are ignored for obtaining quartile deviation. Thus, all the observations are not used in the computation of this measure.

(2) It is not a suitable measure for algebraic operations.

(3) This is not a stable measure with respect to sampling.

(4) This measure is less applicable in the advanced study of statistics.

EXERCISE 4.2

1. A shooter missed his target in the last 10 trials by the following distance (mm) during the practice session.

20, 32, 24, 41, 18, 27, 15, 36, 35, 25

Find the quartile deviation and coefficient of quartile deviation of such distance missed by the shooter.

2. Find the quartile deviation and coefficient of quartile deviation of the marks from the following frequency distribution of marks of 43 students of a school.

Marks	10	20	30	40	50	60
No. of students	4	7	15	8	7	2

3. The distribution of amount paid by 200 customers coming for snacks at a restaurant on a particular day is given below :

Amount (₹)	0-50	50-100	100-150	150-200	200-250
No. of customers	25	40	80	30	25

Find the quartile deviation and coefficient of quartile deviation of the amount paid by customers on the day.

*

4.3.3 Average Deviation

The range and quartile deviation are two measures of dispersion in which all observations of the data are not used and they do not show variation of the observations from any average. This limitation can be overcome by a measure of dispersion in which all observations are considered and variation of each observation from the average is also taken into account. These requirements are fullfilled by average deviation (also known as mean deviation). The difference between value of the observation and its mean is known as **deviation.** These deviations can be negative, zero or positive. The sum of all such deviations is zero as seen in chapter 3. To overcome this situation, the absolute values of these deviations are taken. It means that the negative signs of the negative deviations are ignored and the measure based on absolute deviations of the data is defined.

Thus, the mean of the absolute deviations of the observations of a data from its mean is called Mean Deviation and it is denoted by *MD*.

The relative measure of mean deviation obtained by dividing *MD* by the mean $\bar{x}$ i.e. $\frac{MD}{\bar{x}}$ is known as coefficient of mean deviation.

$\therefore$ Coefficient of mean deviation $= \frac{MD}{\bar{x}}$

Additional Information for Understanding

Since mean is the most widely used measure of central tendency in statistics, we take deviations only from mean in computation of average deviation. The measure of average deviation can also be obtained by taking absolute deviations from the median or the mode as per the need of the data.

Method of computing Mean Deviation

We shall now discuss the method and the formulae of computing mean deviation for ungrouped (raw) and grouped (classified) data.

Ungrouped (raw) Data

Suppose $x_1, x_2, ..., x_n$ are observations of ungrouped data and $\bar{x}$ is their mean. First, the absolute differences of each observation x_i from its mean $\bar{x}$ (i.e. $|x_i - \bar{x}|$) are obtained. Now, the sum of all such absolute deviations is obtained and dividing it by the total number of observations, we get mean deviation. Thus, we define the mean deviation *MD* as

$$MD = \frac{\Sigma|x_i - \bar{x}|}{n}$$

Where, $\bar{x} = \frac{\Sigma x_i}{n}$

$|x_i - \bar{x}|$ = absolute value of $(x_i - \bar{x})$ which is the deviation of observation x_i from mean.

n = Total number of observations

Note : For simplicity at the time of computation in examples, we shall not use suffix '*i*'. We shall write x in place x_i, d in place of d_i, and f in place of f_i.

Grouped Data

Discrete Frequency Distribution :

Suppose $x_1, x_2, ..., x_k$ are the values assumed by the discrete variable x of the discrete frequency distribution with corresponding frequencies $f_1, f_2 ...., f_k$. The mean deviation of the discrete frequency distribution is computed by using the following formula.

$$MD = \frac{\Sigma f_i|x_i - \bar{x}|}{n}$$

Where, x_i = *i*-th value of the variable x

f_i = frequency corresponding to x_i

$n = \Sigma f_i$ = total frequency or sum of all frequencies

$\bar{x} = \frac{\Sigma f_i x_i}{n}$ = Mean

Continuous Frequency Distribution

Suppose the mid-values of k classes of the continuous frequency distribution are $x_1, x_2, ..., x_k$ with corresponding frequencies $f_1, f_2, ..., f_k$ then the mean deviation of the continuous frequency distribution is computed using the following formula :

$$MD = \frac{\Sigma f_i |x_i - \bar{x}|}{n}$$

Where x_i = mid value of the i-th class

f_i = frequency of the i-th class

$n = \Sigma f_i$ = total frequency or sum of all frequencies

$\bar{x} = \frac{\Sigma f_i x_i}{n}$ = mean

After obtaining the mean deviation (MD) by any of the above formula, the coefficient of mean deviation can be obtained as follows :

Coefficient of mean deviation $= \frac{MD}{\bar{x}}$

Illustration 8 : The measurements of weight (in kg) of 8 students of a class of a school are given below:

46, 58, 60, 43, 75, 66, 51, 81

Find the mean deviation and coefficient of mean deviation of weights of the students.

	Weight (kg) x	Deviation $x-\bar{x}$ $\bar{x}=60$	Absolute deviation $\lvert x-\bar{x} \rvert$
	46	–14	14
	58	– 2	2
	60	0	0
	43	– 17	17
	75	15	15
	66	6	6
	51	– 9	9
	81	21	21
Total	**480**	**0**	**84**

$\bar{x} = \frac{\Sigma x}{n} = \frac{480}{8} = 60$ kg

Mean Deviation $MD = \frac{\Sigma|x-\bar{x}|}{n}$

$= \frac{84}{8}$

$= 10.5$

$\therefore$ $MD = 10.5$ kg

Thus, mean deviation of the weights of students is 10.5 kg.

Coefficient of mean Deviation $= \frac{MD}{\bar{x}}$

$= \frac{10.5}{60}$

$= 0.175$

$\therefore$ Coefficient of mean Deviation ≈ 0.18

Thus, coefficient of mean deviation of the students is 0.18.

Illustration 9 : Calculate mean deviation and coefficient of mean deviation of typing time from the following information of time (in minutes) taken to type a report by 32 typists.

Typing time (minutes)	10	11	12	13	14
No. of typists	2	8	12	8	2

Typing time (minutes) x	**No. of typist f**	fx	**Deviation $x-\bar{x}$ $\bar{x}=12$**	**Absolute Deviation $\lvert x-\bar{x}\rvert$**	$f\lvert x-\bar{x}\rvert$
10	2	20	− 2	2	4
11	8	88	− 1	1	8
12	12	144	0	0	0
13	8	104	1	1	8
14	2	28	2	2	4
Total	**32**	**384**	**–**	**–**	**24**

Mean $\bar{x} = \frac{\Sigma fx}{n}$

$= \frac{384}{32}$

$= 12$

$\therefore$ $\bar{x} = 12$ minutes

Mean deviation $MD = \frac{\Sigma f|x - \bar{x}|}{n}$

$= \frac{24}{32}$

$= 0.75$

$\therefore \quad MD = 0.75$ minutes

Thus, mean deviation of time taken to type a report is 0.75 minutes.

Coefficient of mean deviation $= \frac{MD}{\bar{x}}$

$= \frac{0.75}{12}$

$= 0.0625$

$\therefore$ Coefficient of mean deviation ≈ 0.06

Thus, the coefficient of mean deviation of time taken to type a report is 0.06.

Illustration 10 : 20 children are selected for a district level spelling test. The distribution of their marks out of 50 marks is given below :

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49
No. of Children	1	3	8	6	2

Find the mean deviation of the marks obtained by the children from this data.

Marks	**No. of children f**	**Mid-value x**	fx	$x - \bar{x}$ $\bar{x} = 27$	$\lvert x - \bar{x}\rvert$	$f\lvert x - \bar{x}\rvert$
0 - 9	1	4.5	4.5	– 22.5	22.5	22.5
10 - 19	3	14.5	43.5	– 12.5	12.5	37.5
20 - 29	8	24.5	196	– 2.5	2.5	20
30 - 39	6	34.5	207	7.5	7.5	45
40 - 49	2	44.5	89	17.5	17.5	35
Total	**20**	–	**540**	–	–	**160**

Mean $\bar{x} = \frac{\Sigma fx}{n}$

$= \frac{540}{20} = 27$

$\therefore \quad \bar{x} = 27$ Marks

$$\text{Mean } MD = \frac{\Sigma f|x-\bar{x}|}{n}$$

$$= \frac{160}{20}$$

$$= 8$$

$\therefore \quad MD = 8$ marks

Thus, mean deviation of the marks obtained by children is 8 marks.

Illustration 11 : From the following information of fortnight sale of two wheelers by 30 dealers of a city, find the mean deviation of 'number of two wheelers sold'.

No. of two wheelers	12 - 16	17 - 21	22 - 26	27 - 31	32 - 36
No. of dealers	2	3	14	8	3

No. of two wheelers	**No. of dealers** f	**Mid-value** x	fx	$x-\bar{x}$ $\bar{x} = 25.17$	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
12 - 16	2	14	28	– 11.17	11.17	22.34
17 - 21	3	19	57	– 6.17	6.17	18.51
22 - 26	14	24	336	– 1.17	1.17	16.38
27 - 31	8	29	232	3.83	3.83	30.64
32 - 36	3	34	102	8.83	8.83	26.49
Total	**30**	–	**755**	–	–	**114.36**

$$\text{Mean } \bar{x} = \frac{\Sigma fx}{n}$$

$$= \frac{755}{30}$$

$$= 25.1667$$

$\therefore \bar{x} \doteq 25.17$ two wheelers

$$\text{Mean deviation } MD = \frac{\Sigma f|x-\bar{x}|}{n}$$

$$= \frac{114.36}{30}$$

$$= 3.812$$

$\therefore MD \doteq 3.81$ two wheelers

Thus, mean deviation of 'number of two wheelers sold' is 3.81.

Advantages and Disadvantages of Mean Deviation

Advantages :

(1) The mean deviation is a clearly defined measure of dispersion.

(2) It is superior measure to the range and the quartile deviation as all the observations are used in its computation.

(3) Its value is less affected by the extreme values (i.e. unduly the large and the small values) as compared to some other measures of dispersion.

(4) The absolute value of the difference between observation and the mean is used to measure the distance between an observation from the mean, which is an appropriate measure of distance.

Disadvantages :

(1) The computation of mean deviation is complicated as compared to the range and the quartile deviation.

(2) This measure is not suitable for algebraic operations.

(3) This measure is less used in advanced study of statistics as its definition is based on absolute value.

(4) It cannot be computed if the frequency distributions has open-ended classes.

Note : Mean deviation is frequently used to study the problems occurring in social sciences in general. It is also useful in Economics to determine economic inequality, in computing the distribution of personal wealth in the community or country, in forecasting weather and business cycles, etc.

EXERCISE 4.3

1. The measurements of height (in centimeters) of 10 soldiers are given below :

160, 175, 158, 165, 170, 166, 173, 176, 163, 168

Find the mean deviation of the heights of the soldiers.

2. The distribution of number of ball bearings used in machines of a factory is given below. Calculate the mean deviation and coefficient of mean deviation of number of ball bearings per machine.

No. of ball bearings	2	4	6	8	10	12	14	16
No. of machines	2	2	4	5	3	2	1	1

3. Find the mean deviation and the coefficient of mean deviation of the distribution of talk time (in minutes) per call :

Talk time (minutes)	3	5	10	12	15
No. of calls	4	7	6	2	1

4. Find the mean deviation and coefficient of mean deviation of number of TV sets using the following frequency distribution of TV sets sold in last 16 months in a town.

No. of TV sets	10 – 30	30 – 50	50 – 70	70 – 90	90 – 110
No. of months	1	4	6	4	1

5. There are 50 boxes containing different number of units of an item in a factory. Find the mean deviation of number of units per box using the following distribution of the units.

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

*

4.3.4 Standard Deviation

We have seen that the definition of mean deviation is based on absolute values of the deviations of observations of the data from the mean. Since the algebraic signs of the deviations are ignored, mean deviation is less used in advanced study of statistics. This limitation of mean deviation is overcome by an important measure of dispersion known as Standard Deviation. Instead of taking the absolute value of deviation of each observation from the mean, the square of the deviation is taken. If the sum of squares of these deviations is divided by the total number of observations, we get an important measure of dispersion known as Variance. It is denoted by s^2. The positive square root of the variance is called the Standard Deviation. It is denoted by s.

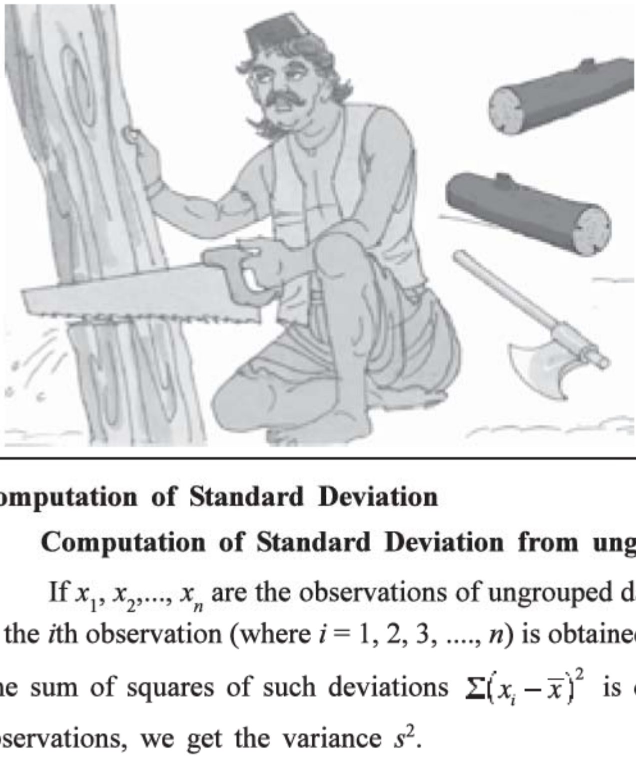
Well known statistician Karl Pearson defined the Standard Deviation as, "Standard Deviation is the positive square root of the mean of the squares of the deviations measured from the mean."

After mean, standard deviation is another very useful measure which gives information about values of the observations of a population.

Note that the standard deviation is an absolute measure of dispersion. If the standard deviation is divided by the mean of the data, we get its relative measure of dispersion. It is called the **coefficient of standard deviation.**

$\therefore$ Coefficient of standard deviation $= \frac{s}{\bar{x}}$

Note : Among all the measures of dispersion, standard deviation is the most important and widely used measure. The variance and the standard deviation are widely used in experimental research in applied fields such as physics, agricultural science and medicine.They are also useful and an important measure in the study of statistical inference, correlation analysis, sampling and other areas of study.



""The two measures mean and standard deviation are to the statistician what the axe and cross cut saw to the woods man — the basic tools for working upon his raw material".

– M. M. Blair

Computation of Standard Deviation

Computation of Standard Deviation from ungrouped data :

If $x_1, x_2, ..., x_n$ are the observations of ungrouped data and $\bar{x}$ is its mean, then first the deviation $x_i - \bar{x}$ of the *i*th observation (where $i = 1, 2, 3, ..., n$) is obtained as discussed in the definition of standard deviation. The sum of squares of such deviations $\Sigma(x_i - \bar{x})^2$ is obtained. Dividing the sum by the total number of observations, we get the variance s^2.

$\therefore$ variance $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n}$

The positive square root of the variance gives the standard deviation and its formula is as follows :

Standard Deviation $s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

Illustration 12 : The runs scored by a batsman in his last 7 matches are given below :

52, 58, 40, 60, 54, 38, 48

Find the variance of the runs of the batsman. Also find the standard deviation.

	Runs x	$x-\bar{x}$ $\bar{x}=50$	$(x-\bar{x})^2$
	52	2	4
	58	8	64
	40	– 10	100
	60	10	100
	54	4	16
	38	– 12	144
	48	– 2	4
Total	**350**	**0**	**432**

Mean $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{350}{7}$

$= 50$ runs

Variance $s^2 = \frac{\Sigma(x-\bar{x})^2}{n}$

$= \frac{432}{7}$

$= 61.7143$

$\therefore$ $s^2 \doteq 61.71$ (runs)^2

Standard deviations $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{61.7143}$

$= 7.8558$

$\therefore$ $s \doteq 7.86$ runs

Thus, the standard deviation of runs scored by batsman is 7.86 runs.

Note : The standard deviation is expressed in the units of the observations. We know that variance is square of the standard deviation, hence the unit of variance is square of the unit of the standard deviation.

e.g. : If the unit of the observations is kg., the unit of its standard deviation is also kg, whereas the unit of its variance is (kg.)^2.

Note : When the value of mean $\bar{x}$ is a fractional number and observations are not numerically large, the computation of *s* can be made simpler by the following formula for the ungrouped data.

$$s = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2}$$

Illustration 13 : The time (in minutes) taken to solve a puzzle by 5 students are 5, 8, 3, 6, 10. Compute the standard deviation of the time taken to solve the puzzle.

	Time (minutes) x	x^2
	5	25
	8	64
	3	9
	6	36
	10	100
Total	**32**	**234**

$$\text{Mean } \bar{x} = \frac{\Sigma x}{n}$$

$$= \frac{32}{5}$$

= 6.4 minutes

Since the value of $\bar{x}$ is fractional, we shall use the following alternative formula to obtain the standard deviation.

$$s = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{234}{5} - (6.4)^2}$$

$$= \sqrt{46.8 - 40.96}$$

$$= \sqrt{5.84}$$

= 2.4166

$\therefore \quad s \doteq 2.42$ minutes

Thus the standard deviation of time taken by students to solve a puzzle is 2.42 minutes.

Short-cut Method :

To make the computation of standard deviation simpler, the following short-cut method can the used.

$$s = \sqrt{\frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2}$$

where $d_i = x_i - A$

A = Assumed mean

n = total number of observations

Illustration 14 : The following are closing prices (in ₹) of 5 shares :

132, 147, 120, 152, 125

Find the standard deviation by short-cut method.

We take assumed mean A = 135

Price (₹) x	$d = x - A$ $A = 135$	d^2
132	– 3	9
147	12	144
120	– 15	225
152	17	289
125	– 10	100
Total	**1**	**767**

Standard Deviation

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

$$= \sqrt{\frac{767}{5} - \left(\frac{1}{5}\right)^2}$$

$$= \sqrt{153.4 - 0.04}$$

$$= \sqrt{153.36}$$

$$= 12.3839$$

$\therefore \quad s \approx$ ₹ 12.38

Thus, the standard deviation of price of share is ₹ 12.38.

Computation of Standard Deviation For Grouped Data

For Discrete Frequency Distribution :

Suppose the values of variable x of a discrete frequency distribution are $x_1, x_2, ..., x_k$ with frequencies $f_1, f_2, ..., f_k$ respectively, then the formula for the standard deviation of the frequency distribution is as follows :

$$s = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{n}}$$

$$= \sqrt{\frac{\Sigma f_i x_i^2}{n} - \bar{x}^2}$$

where f_i = frequency of the i-th value x_i

$$\bar{x} = \frac{\Sigma f_i x_i}{n}$$

$x_i - \bar{x}$ = deviation of x_i from the mean $\bar{x}$

$n = \Sigma f_i$ = total number of observations

Short-cut Method :

The standard deviation for the discrete frequency distribution is obtained by short cut method using the following formula :

$$s = \sqrt{\frac{\Sigma f_i d_i^2}{n} - \left(\frac{\Sigma f_i d_i}{n}\right)^2}$$

where f_i = frequency of the i value x_i of the variable

A = assumed mean

$d_i = x_i - A$ = deviation of x_i from assumed mean A

$n = \Sigma f_i$ = total number of observations

Note : The value of assumed mean A can be taken as any one of the observations $x_1, x_2, ..., x_k$ or any other suitable value.

Illustration 15 : The distribution of number of absent days of 15 students of a class in the month of January is given below. Find the standard deviation and coefficient of standard deviation of their number of absent days.

No. of absent days	0	1	2	3	4
No. of Students	1	3	7	3	1

x	f	fx	$x - \bar{x}$	$(x - \bar{x})^2$	$f(x - \bar{x})^2$	fx^2
0	1	0	– 2	4	4	0
1	3	3	– 1	1	3	3
2	7	14	0	0	0	28
3	3	9	1	1	3	27
4	1	4	2	4	4	16
Total	$n = 15$	**30**	**0**	**–**	**14**	**74**

$$\bar{x} = \frac{\Sigma fx}{n} = \frac{30}{15} = 2 \text{ days}$$

$$s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n}}$$

$$s = \sqrt{\frac{14}{15}}$$

$$s = \sqrt{0.9333}$$

$$= 0.9661$$

$$\therefore \quad s \approx 0.97 \text{ days}$$

The value of standard deviation can be computed by the alternative formula as follows :

$$s = \sqrt{\frac{\Sigma fx^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{74}{15} - (2)^2}$$

$$= \sqrt{4.9333 - 4}$$

$$= \sqrt{0.9333}$$

$$= 0.9661$$

$\therefore \quad s \risingdotseq 0.97$ days

Thus, the standard deviation of number of 'absent days' is 0.97 day.

Co-efficient of standard deviation $\frac{s}{\bar{x}} = \frac{0.97}{2}$

$= 0.485$

$\risingdotseq 0.49$

Thus, coefficent of standard deviation of 'absent days' is 0.49

Illustration 16 : The information of number of mobiles phones sold in last 35 days in a mobile shop is given below. Find the co-efficient of standard deviation of number of mobile phones sold. (Use short-cut method).

No. of mobile phones sold	5	6	7	8	9	10
No. of Days	2	5	8	12	7	1

Assumed mean A = 8

x	f	$d = x - A$ $A = 8$	fd	fd^2
5	2	– 3	– 6	18
6	5	– 2	– 10	20
7	8	– 1	– 8	8
8	12	0	0	0
9	7	1	7	7
10	1	2	2	4
Total	$n = 35$	–	**– 15**	**57**

$$\bar{x} = A + \frac{\Sigma fd}{n}$$

$$= 8 + \frac{(-15)}{35}$$

$$= 8 - 0.4286$$

$$= 7.5714$$

$\therefore \quad \bar{x} \risingdotseq 7.57$ mobile phones

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

$$= \sqrt{\frac{57}{35} - \left(\frac{-15}{35}\right)^2}$$

$$= \sqrt{1.6286 - 0.1837}$$

$$= \sqrt{1.4449}$$

$$= 1.2020$$

$\therefore \quad s \doteq 1.20$ mobiles

Thus, the standard deviation of number of mobile phones sold is 1.20.

Coefficient of standard deviation $= \frac{s}{\bar{x}}$

$$= \frac{1.20}{7.57}$$

$$= 0.1585$$

$\therefore$ Coefficient of standard deviation $\doteq 0.16$

Thus, coefficient of standard deviation of 'number of mobile phones sold' is 0.16.

For Continuous Frequency Distribution :

Suppose the mid-values of k classes of a continuous frequency distribution are $x_1\ x_2, ..., x_k$ and the respective frequencies of k classes are $f_1, f_2, ..., f_k$. Then the formula for computing the standard deviation of frequencies distribution is given by

$$s = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{n}}$$

$$s = \sqrt{\frac{\Sigma f_i x_i^2}{n} - \bar{x}^2}$$

Where f_i = frequency of the i th class

x_i = mid-value of the i th class

$\bar{x} = \frac{\Sigma f_i x_i}{n}$

$x_i - \bar{x}$ = deviation of the mid-value x_i from the mean $\bar{x}$

$n = \Sigma f_i$ = total number of observations

Short-cut Method :

When a continuous frequency distribution with same class length is given, the following formula is used to compute the standard deviation.

$$s = \sqrt{\frac{\Sigma f_i d_i^2}{n} - \left(\frac{\Sigma f_i d_i}{n}\right)^2} \times c$$

Where, x_i = mid-value of the ith class

A = assumed mean

f_i = frequency of the i-th class

c = class length

$$d_i = \frac{x_i - A}{c}$$

$n = \Sigma f_i$ = total number of observations.

Note :
- The value of the assumed mean A can be taken as any of the mid-values or any other suitable value.
- The value of the standard deviation remains same, if it is obtained by any suitable form of the formula.

Illustration 17 : Calculate the standard deviation from the following frequency distribution of marks obtained by 200 students of a school in an examination :

Marks	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of students	5	12	30	45	50	37	21

We do not need the value of the mean $\bar{x}$ as only the standard deviation is to be obtained. In such a situation, generally, short-cut method is preferred.

Here assumed mean $A = 35$ and class length $c = 10$

Marks	**No. of students f**	**Mid-value x**	$d = \frac{x-A}{c}$ **$A = 35, c = 10$**	fd	fd^2
0 - 10	5	5	– 3	– 15	45
10 - 20	12	15	– 2	– 24	48
20 - 30	30	25	– 1	– 30	30
30 - 40	45	35	0	0	0
40 - 50	50	45	1	50	50
50 - 60	37	55	2	74	148
60 - 70	21	65	3	63	189
Total	$n = 200$	–	–	**118**	**510**

Standard Deviation

$$s = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} \times c$$

$$= \sqrt{\frac{510}{200} - \left(\frac{118}{200}\right)^2} \times 10$$

$$= \sqrt{2.55 - 0.3481} \times 10$$

$$= \sqrt{2.2019} \times 10$$

$$= 14.8388$$

$\therefore \quad s \approx 14.84$ Marks

Thus, standard deviation of the marks of the students is 14.84 marks.

Illustration 18 : Find the standard deviation of the daily wages from following information of wages (in ₹) of workers of a factory.

Daily wages (₹)	More than 130	More than 150	More than 170	More than 190	More than 210	More than 230
No. of persons	150	142	116	57	14	0

Here, the 'more than' cumulative frequency distribution is given. Converting it into frequency distribution, we get the following frequency distribution.

Daily wages (₹)	130 – 150	150 – 170	170 - 190	190 – 210	210 - 230
No. of persons	150 – 142 = 8	142 – 116 = 26	116 – 57 = 59	57 – 14 = 43	14 – 0 = 14

Daily wages(₹)	**No. of persons f**	**Mid-value x**	$d = \frac{x - A}{c}$ **$A = 180$, $c = 20$**	fd	fd^2
130 – 150	8	140	– 2	– 16	32
150 – 170	26	160	– 1	– 26	26
170 – 190	59	180	0	0	0
190 – 210	43	200	1	43	43
210 – 230	14	220	2	28	56
Total	$n = 150$	–	–	**29**	**157**

Standard Deviation

$$s = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} \times c$$

$$= \sqrt{\frac{157}{150} - \left(\frac{29}{150}\right)^2} \times 20$$

$$= \sqrt{1.0467 - (0.1933)^2} \times 20$$

$$= \sqrt{1.0467 - 0.0374} \times 20$$

$$= \sqrt{1.0093} \times 20$$

$$= 20.0928$$

$\therefore \quad s \approx 20.09$ ₹

Thus standard deviation of the daily wages of workers of the factory is 20.09 ₹

Note : If all observations under the study are same i.e. $x_1 = x_2 = x_3 = ... x_n = k$; where k = some constant value, then the value of any measure of dispersion is zero.

Exercise 4.4

1. The marks obtained by 9 students in a test of 100 marks in Mathematics are given below :

64, 63, 72, 65, 68, 69, 66, 67, 69

Find the standard deviation of marks obtained by the students.

2. The numbers of cars coming for service in five service stations of a company on a particular day are 7, 3, 11, 8, 9. Calculate the standard deviation of number of cars coming at the service station.

3. The following frequency distribution represents the amounts of deposits and the number of depositors in a bank. Find the coefficient of standard deviation of the deposits.

Deposits (thousand ₹)	5	10	15	20	25	30	35
No. of depositors	2	7	11	15	10	4	1

4. The information of profits (in lakhs ₹) of 50 firms in the last year is given below. Find the standard deviation of the profit of the firms.

Profit (lakh ₹)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50
No. of firms	7	6	15	12	10

5. Find the standard deviation of age of the persons from the following distribution of 125 persons living in a society. Also find the coefficient of standard deviation.

Age (years)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80
No. of persons	15	15	23	22	25	10	5	10

*

Coefficient of Variation :

We have seen that the standard deviation is an absolute measure and it is expressed in terms of unit of the given observations of the data. Therefore, for the comparison of variability of two or more groups, their absolute measures cannot be used. For such a comparison, its relative measure, coefficient of standard deviation $\left(\frac{s}{\bar{x}}\right)$, should be used. Often, the value of coefficient of standard deviation $\left(\frac{s}{\bar{x}}\right)$ comes in fractional form, so Karl Pearson has suggested "Coefficient of Variation" as a relative measure which can be easily understood by common people. The coefficient of variation is obtained by multiplying coefficient of standard deviation by 100.

$\therefore$ Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

The coefficient of variation is measured in terms of percentage. i.e. coefficient of variation is percentage measure of standard deviation with respect to mean.

It is a very useful measures for comparing the dispersion of two or more data sets. A group of observations which has smaller value of coefficient of variation is said to be more stable and having less dispersion. Such a sequence is also said to be consistent from the point of view of variability. A sequence of observations which has larger value of coefficient of variation is said to be less stable and having more dispersion.

Illustration 19 : From the following data of runs scored by two batsmen A and B in last 10 innings, decide who is more consistent.

Runs by batsman A	25	50	45	30	70	42	36	48	34	60
Runs by batsman B	10	70	50	20	95	55	42	60	48	80

In order to know who is a more consistent batsman in terms of runs, we shall obtain the coefficient of variation of runs for both A and B.

Batsman A

	Runs x	$x-\bar{x}$ $\bar{x}=44$	$(x-\bar{x})^2$
	25	– 19	361
	50	6	36
	45	1	1
	30	– 14	196
	70	26	676
	42	– 2	4
	36	– 8	64
	48	4	16
	34	– 10	100
	60	16	256
Total	**440**	**0**	**1710**

Batsman B

	Runs x	$x-\bar{x}$ $\bar{x}=53$	$(x-\bar{x})^2$
	10	– 43	1849
	70	17	289
	50	– 3	9
	20	– 33	1089
	95	42	1764
	55	2	4
	42	– 11	121
	60	7	49
	48	– 5	25
	80	27	729
Total	**530**	**0**	**5928**

For Batsman A

Mean $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{440}{10} = 44$

$\therefore \quad \bar{x} = 44$ runs

Standard deviation $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{1710}{10}}$

$= \sqrt{171}$

$= 13.0767$

$\therefore \quad s \fallingdotseq 13.08$ runs

Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

$= \frac{13.08}{44} \times 100$

$= \frac{1308}{44}$

$= 29.7272$ %

$\therefore$ Coefficient of Variation $\fallingdotseq 29.73$ %

For Batsman B

Mean $\bar{x} = \frac{\Sigma x}{n}$

$= \frac{530}{10} = 53$

$\therefore \quad \bar{x} = 53$ runs

Standard deviation $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$

$= \sqrt{\frac{5928}{10}}$

$= \sqrt{592.8}$

$= 24.3475$

$\therefore \quad s \fallingdotseq 24.35$ runs

Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

$= \frac{24.35}{53} \times 100$

$= \frac{2435}{53}$

$= 45.9434$ %

$\therefore$ Coefficient of Variation $\fallingdotseq 45.94$ %

Since the Coefficient of variation for batsman A is less, batsman A is more consistent.

Additional information for understanding

When the means of runs by batsmen are same or approximately same, then a consistent batsman is a better batsman. But the same cannot be said if the means of two batsmen are different.

Illustration 20 : The following information is available for two workers of a factory :

	Workers A	**Workers B**
Mean time of completing job (minutes)	30	25
Standard Deviation (minutes)	6	4

Which worker has more relative variation or fluctuation in the time taken to complete the job ?

For the decision, coefficients of variation of both workers are to be compared.

worker A

$\bar{x}$ = 30 minutes, s = 6 minute

Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

$= \frac{6}{30} \times 100$

$= 20\ \%$

worker B

$\bar{x}$ = 25 minutes, s = 4 minutes

Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

$= \frac{4}{25} \times 100$

$= 16\ \%$

Since the coefficient of variation of worker A is more, there is more variation in the time taken by worker A to complete the job.

Illustration 21 : The means and standard deviations of heights and weights of 50 students of a class are as follows :

	Weight	Height
Mean	56.2 kg	62.5 inch
Standard Deviation	4.8 kg	9.3 inch

Where do you find more variation, among heights or weights?

The units of weights and heights are different here. So, a relative measure of dispersion must be used for comparison and considering the given data, coefficient of variation is an appropriate measure.

Weight

$\bar{x}$ = 56.2 kg s = 4.8 kg

Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

$= \frac{4.8}{56.2} \times 100$

$= 8.54\ \%$

Height

$\bar{x}$ = 62.5 kg, s = 9.3 inch

Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$

$= \frac{9.3}{62.5} \times 100$

$= 14.88\ \%$

Since the coefficient of variation is more for the heights, we say that height shows more variation.

Advantages and Disadvantages of Standard Deviation

Advantages :

(1) Its definition is clear and precise.

(2) All the observations are used in its computation.

(3) Standard deviation is the most efficient measure of dispersion among all the measures of dispersion.

(4) Standard deviation is a suitable measure for algebraic calculations. For example, if the means and standard deviations of two data sets are given, the combined standard deviation of a new data set formed by combining the observations of two given data sets can be obtained. It is not possible to obtain a combined measure in case of other measures of dispersion by such an algebraic manipulation.

(5) Standard deviation is the most widely used measure of dispersion among all the measures of dispersion.

Disadvantages :

(1) The computation of standard deviation is more complicated as compared to computation of other measures of dispersion.

(2) The extreme observations get undue importance in the value this measure.

(3) It cannot be obtained if the frequency distributions have open-ended classes.

EXERCISE 4.5

1. Price fluctuations of two shares A and B are given below, which type of share has more relative variation in its price ?

Price (₹) share A	321	322	325	322	324	320	323	316	319	318
Price (₹) share B	141	146	130	146	142	145	132	134	132	152

2. The daily salary of administrative staff of two companies yielded the following results :

	Company A	**Company B**
Mean Salary (₹)	600	2100
Standard Deviation (₹)	30	84

Which company has more stable salary ?

3. The Coefficients of variation of two series are 30% and 25% and their standard deviations are 15 and 9 respectively. Find their means.

*

4.4 Combined Standard Deviation

Suppose we have two groups G_1 and G_2 of the data obtained from population and the following information is obtained for the two groups.

Details	**For Group G_1**	**For Group G_2**
No. of observations	n_1	n_2
Mean	$\bar{x}_1$	$\bar{x}_2$
Standard Deviation	s_1	s_2

A new group G is obtained by combining the observations of two groups G_1 and G_2. Then the mean and standard deviation of this combined group are known as combined mean $\bar{x}_c$ and combined standard deviation s_c respectively and their formulae are as follows :

Combined Mean $\bar{x}_c = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Combined Standard Deviation $s_c = \sqrt{\dfrac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}}$

Where, n_1 = No. of observations in group G_1

n_2 = No. of observations in group G_2

s_1 = Standard deviation of group G_1

s_2 = Standard deviation of group G_2

$d_1 = \bar{x}_1 - \bar{x}_c$

$d_2 = \bar{x}_2 - \bar{x}_c$

Illustration 22 : Five observations in each of two groups G_1 and G_2 are given below :

Group G_1 : 1, 3, 5, 7, 9

Group G_2 : 2, 4, 6, 8, 10

Find the mean and variance of both the groups. Hence obtain the combined standard deviation from it.

For Group G_1

	x	$(x - \bar{x})$ $\bar{x} = 5$	$(x - \bar{x})^2$
	1	− 4	16
	3	− 2	4
	5	0	0
	7	2	4
	9	4	16
Total	**25**	**0**	**40**

For Group G_2

	x	$(x - \bar{x})$ $\bar{x} = 6$	$(x - \bar{x})^2$
	2	− 4	16
	4	− 2	4
	6	0	0
	8	2	4
	10	4	16
Total	**30**	**0**	**40**

Mean of Group G_1

$$\bar{x}_1 = \frac{\Sigma x}{n_1}$$

$$= \frac{25}{5}$$

$$= 5$$

$$\therefore \quad \bar{x}_1 = 5$$

Mean of Group G_2

$$\bar{x}_2 = \frac{\Sigma x}{n_2}$$

$$= \frac{30}{5}$$

$$= 6$$

$$\therefore \quad \bar{x}_2 = 6$$

Variance of Group G_1

$$s_1^{\,2} = \frac{\Sigma(x-\bar{x}_1)^2}{n_1}$$

$$= \frac{40}{5}$$

$$= 8$$

$$\therefore \quad s_1^{\,2} = 8$$

Variance of Group G_2

$$s_2^{\,2} = \frac{\Sigma(x-\bar{x}_2)^2}{n_2}$$

$$= \frac{40}{5}$$

$$= 8$$

$$\therefore \quad s_2^{\,2} = 8$$

Combined Mean $\bar{x}_c = \dfrac{n_1\bar{x}_1 - n_2\bar{x}_2}{n_1 - n_2}$

$$= \frac{5\,(5)+5\,(6)}{5+5}$$

$$= \frac{25-30}{10}$$

$$= \frac{55}{10}$$

$$= 5.5$$

$$\therefore \quad \bar{x}_c = 5.5$$

$d_1 = \bar{x}_1 - \bar{x}_c = 5 - 5.5 = -0.5$

$d_2 = \bar{x}_2 - \bar{x}_c = 6 - 5.5 = 0.5$

Combined Standard Deviation

$$s_c = \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{5\,(8 + (-0.5)^2) + 5\,(8 + (0.5)^2)}{5 + 5}}$$

$$= \sqrt{\frac{5(8.25) + 5(8.25)}{10}}$$

$$= \sqrt{\frac{41.25 + 41.25}{10}}$$

$$= \sqrt{\frac{82.5}{10}}$$

$$= \sqrt{8.25}$$

$$= 2.8723$$

$\therefore \quad s_c \approx 2.87$

Activity

Combine the observations of group G_1 and group G_2 from example 22. So, you will have 10 observations 1, 3, 5, 7, 9, 2, 4, 6, 8, 10. Now find mean and standard deviation of these 10 observations and you will see that the mean and the standard deviation will be same as $\bar{x}_c$ and s_c of the example 22.

Illustration 23 : A factory manufactures certain items in two shifts. The information regarding the time taken by workers to manufacture the items is given below. Using the following information, find the combined standard deviation :

	Shift I	**Shift II**
No. of workers	60	40
Mean manufacturing time (minutes)	25	20
Standard deviation (minutes)	5	3

We shall first find the combined mean $\bar{x}_c$

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

$$= \frac{60(25) + 40(20)}{60 + 40}$$

$$= \frac{1500 + 800}{100}$$

$$= \frac{2300}{100}$$

$= 23$ minutes

Thus, it can be said that the mean time taken by all workers of the factory is 23 minutes.

$d_1 = \bar{x}_1 - \bar{x}_c = 25 - 23 = 2$

$d_2 = \bar{x}_2 - \bar{x}_c = 20 - 23 = -3$

Combined Standard deviation

$$s_c = \sqrt{\frac{n_1(s_1^{\ 2} + d_1^{\ 2}) + n_2(s_2^{\ 2} + d_2^{\ 2})}{n_1 + n_2}}$$

$$= \sqrt{\frac{60(5^2 + 2^2) + 40(3^2 + (-3)^2)}{60 + 40}}$$

$$= \sqrt{\frac{60(25 + 4) + 40(9 + 9)}{100}}$$

$$= \sqrt{\frac{60(29) + 40(18)}{100}}$$

$$= \sqrt{\frac{1740 + 720}{100}}$$

$$= \sqrt{\frac{2460}{100}}$$

$$= \sqrt{24.6}$$

$$= 4.9598$$

$\therefore \quad s_c \approx 4.96$ minutes

Thus the standard deviation of the time taken by all workers of both the shifts is 4.96 minutes.

Additional information for understanding

Mean, median and mode are known as "first order averages" where as measures of dispersion are known as "second order averages."

EXERCISE 4.6

1. The information regarding marks of the students of two classes of a school is given below. Find the combined standard deviation of the marks obtained by the students.

	Division A	Division B
No. of students	50	60
Mean marks	60	48
Standard deviation	10	12

2. The following information is available for two sections of a factory. Obtain the combined standard deviation of the production time.

	Section A	Section B
No. of workers	10	40
Mean production time per unit (minute)	25	20
Variance	16	25

*

Illustration 24 : The information of number of accidents in 10 days on a particular road is given below. Find the mean and standard deviation of number of accidents per day. Find the percentage of days having the number of accidents lying between the limits $\bar{x} \pm s$.

No. of accidents	1	2	3	4	5
No. of days	2	3	3	1	1

No. of accidents x	No. of days f	fx	fx^2
1	2	2	2
2	3	6	12
3	3	9	27
4	1	4	16
5	1	5	25
Total	n **= 10**	**26**	**82**

Mean

$$\bar{x} = \frac{\Sigma fx}{n}$$

$$= \frac{26}{10}$$

$$= 2.6$$

$$\therefore \bar{x} = 2.6 \text{ accidents}$$

Standard Deviation

$$s = \sqrt{\frac{\Sigma fx^2}{n} - \bar{x}^2}$$

$$= \sqrt{\frac{82}{10} - (2.6)^2}$$

$$= \sqrt{8.2 - 6.76}$$

$$= \sqrt{1.44}$$

$$= 1.2$$

$$\therefore s = 1.2 \text{ accidents}$$

Now, $\bar{x} - s = 2.6 - 1.2 = 1.4$ accidents

$\bar{x} + s = 2.6 + 1.2 = 3.8$ accidents

It can be seen from the frequency distribution that the number of accidents lying within limits 1.4 and 3.8 are 2 and 3 and the number of days having 2 and 3 accidents are 3 and 3 respectively. Hence the number of days having number of accidents within the limits $\bar{x} - s = 1.4$ and $\bar{x} + s = 3.8$ are $3 + 3 = 6$. Since the total number of days is 10, the percentage of days having accidents within the limits is $\frac{6}{10} \times 100 = 60$

Illustration 25 : The mean number of units produced by 100 workers of a factory per day is 60 units and its standard deviation is 10 units. Later on, it was noticed that two workers have actually produced 30 and 20 units respectively but it was registered as 5 and 45 units respectively. By considering this, find corrected mean and corrected standard deviation of the number of units produced by the workers.

We are given $n = 100$, $\bar{x} = 60$, $s = 10$

$$\bar{x} = \frac{\Sigma x}{n} \qquad s = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$60 = \frac{\Sigma x}{100} \qquad 10 = \sqrt{\frac{\Sigma x^2}{100} - (60)^2}$$

$$\therefore \quad \Sigma x = 6000 \qquad 100 = \frac{\Sigma x^2}{100} - 3600$$

$$3700 = \frac{\Sigma x^2}{100}$$

$$\therefore \quad \Sigma x^2 = 3,70,000$$

But, these values of Σx and Σx^2 are not correct. Now, replacing wrong number of units by the correct number of units produced by the workers, we get,

Corrected $\Sigma x = 6000 - 5 - 45 + 30 + 20 = 6000$

Corrected $\Sigma x^2 = 3,70,000 - (5)^2 - (45)^2 + (30)^2 + (20)^2$

$= 3,70,000 - 25 - 2025 + 900 + 400$

$= 3,69,250$

$$\therefore \text{ Corrected Mean } \bar{x} = \frac{\text{Corrected } \Sigma x}{n}$$

$$= \frac{6000}{100}$$

$= 60$ units

$$\text{Corrected Standard Deviation } s = \sqrt{\frac{\text{corrected } \Sigma x^2}{n} - (\text{corrected } \bar{x})^2}$$

$$= \sqrt{\frac{369250}{100} - (60)^2}$$

$$= \sqrt{3692.5 - 3600}$$

$$= \sqrt{92.5}$$

$= 9.6177$

≈ 9.62 units

Illustration 26 : The following results are obtained on the basis of daily wages (in ₹) paid to workers of two firms A and B :

	Firm A	**Firm B**
No. of workers	20	30
Mean daily wages (₹)	250	400
Standard deviation of wages (₹)	10	12

Answer the following questions using the above information.

1. **Which firm is paying more total daily wages to their workers ?**
2. **Determine which firm shows more relative variation in wages paid to its workers.**
3. **Find the combined mean and combined standard deviation for the two firms.**

(1) Firm A

(n_1 = 20, $\bar{x}_1$ = 250)

Total daily wages = $n_1\bar{x}_1$

= 20 (250)

= 5000

Firm B

(n_2 = 30, $\bar{x}_2$ = 400)

Total daily wages = $n_2\bar{x}_2$

= 30 (400)

= 12000

Hence, Firm B pays more daily wages.

(2) Firm A

Coefficient of variation $= \frac{s_1}{\bar{x}_1} \times 100$

$= \frac{10}{250} \times 100$

= 4 %

Firm B

Coefficient of variation $= \frac{s_2}{\bar{x}_2} \times 100$

$= \frac{12}{400} \times 100$

= 3 %

The Coefficient of variation for firm A is more. Hence, there is more variation in the daily wages of firm A.

(3) Combined Mean $\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$$= \frac{20(250) + 30(400)}{20 + 30}$$

$$= \frac{5000 + 12000}{50}$$

$$\therefore \quad \bar{x}_c = \frac{17000}{50} = ₹\ 340$$

Thus, after combining the workers of both the firms, their mean daily wages becomes ₹ 340.

$d_1 = \bar{x}_1 - \bar{x}_c = 250 - 340 = -90$

$d_2 = \bar{x}_2 - \bar{x}_c = 400 - 340 = 60$

Combined Standard Deviation

$$s_c = \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}}$$

$$= \sqrt{\frac{20\left((10)^2 + (-90)^2\right) + \left(30(12)^2 + (60)^2\right)}{20 + 30}}$$

$$= \sqrt{\frac{20(100 + 8100) + 30(144 + 3600)}{50}}$$

$$= \sqrt{\frac{164000 + 112320}{50}}$$

$$= \sqrt{\frac{276320}{50}}$$

$$= \sqrt{55264}$$

$$= 74.3398$$

$\therefore \quad s_c \approx ₹\ 74.34$

Thus, after combining the workers of both the firms their standard deviation is ₹ 74.34.

Illustration 27 : The information of number of roses on 30 rose plants in a nursery is given below. Find range, coefficient of range, quartile deviation, coefficient of quartile deviation, mean deviation and coefficient of mean deviation of the number of roses from it.

No. of roses	1	3	5	6 – 8	8 – 12	12 – 16	16 – 22
No. of plants	1	2	5	10	8	3	1

No. of roses	No. of plants f	cf	Mid-value x	fx	$\lvert x - \bar{x} \rvert$ $\bar{x} = 8.1$	$f \lvert x - \bar{x} \rvert$
1	1	1	1	1	7.1	7.1
3	2	3	3	6	5.1	10.2
5	5	8	5	25	3.1	15.5
6–8	10	18	7	70	1.1	11
8–12	8	26	10	80	1.9	15.2
12–16	3	29	14	42	5.9	17.7
16–22	1	30	19	19	10.9	10.9
Total	$n = 30$	–	–	**243**	**35.1**	**87.6**

Mean $\bar{x} = \frac{\Sigma fx}{n}$

$= \frac{243}{30}$

$= 8.1$ roses

Here, $x_H = 22$ and $x_L = 1$

$\therefore$ Range $= x_H - x_L$

$= 22 - 1$

$= 21$ roses

Coefficient of range $= \frac{x_H - x_L}{x_H + x_L}$

$= \frac{21}{22+1}$

$= \frac{21}{23}$

$= 0.9130$

$\therefore$ Coefficient of range ≈ 0.91

Q_1 = Value of the $\left(\frac{n}{4}\right)$th observation

= Value of the $\left(\frac{30}{4}\right)$th observation

= Value of the 7.5th observation

Referring to the column of cumulative frequency (cf), we see that the value of the 7.5th observation is 5.

$\therefore$ $Q_1 = 5$ roses

Q_3 = value of the $3\left(\frac{n}{4}\right)$th observation

= value of the 3(7.5) th observation

= value of the 22.5 th observation

Referring to the column of cumulative frequency (*cf*), we see that the value of the 22.5th observation lies in the class 8-12. Hence Q_3 class is 8-12.

Now, $Q_3 = L + \frac{3\left(\frac{n}{4}\right) - cf}{f} \times c$

Here, $L = 8$, $3\left(\frac{n}{4}\right) = 22.5$, $cf = 18$, $f = 8$, $c = 4$

$\therefore \quad Q_3 = 8 + \frac{22.5 - 18}{8} \times 4$

$= 8 + \frac{4.5}{2}$

$= 8 + 2.25$

$= 10.25$ roses

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

$= \frac{10.25 - 5}{2}$

$= \frac{5.25}{2}$

$= 2.625$

$\therefore \quad Q_d \approx 2.63$ roses

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{5.25}{10.25 + 5}$

$= \frac{5.25}{15.25}$

$= 0.3443$

$\therefore$ Coefficient of quartile deviation = 0.34

Now, Mean Deviation $MD = \frac{\Sigma f \left| x - \bar{x} \right|}{n}$

$= \frac{87.6}{30}$

$= 2.92$

$\therefore$ Mean Deviation MD = 2.92 roses

Coefficient of mean deviation $= \frac{MD}{\bar{x}}$

$= \frac{2.92}{8.1}$

$= 0.3605$

$\therefore$ Coefficient of mean deviation ≈ 0.36

Some Useful Results

Suppose the range, quartile deviation, mean deviation and standard deviation for observations $x_1, x_2, ..., x_n$ are R_x, Q_{dx}, MD_x and s_x respectively. Now, if each observation x_i (where $i = 1, 2, ..., n$) is multiplied by a non-zero constant "b" and a constant "a" is added to it, a new set of observations $y_1, y_2, ..., y_n$ is obtained. i.e. $y_i = bx_i + a$

Then the range, quartile deviation, mean deviation, standard deviation and variance of y can be obtained from the respective measures of x as follows :

Measures	**For x**	**For y**
Range	R_x	$R_y = \|b\|.R_x$
Quartile Deviation	Q_{dx}	$Q_{dy} = \|b\|.Q_{dx}$
Mean Deviation	MD_x	$MD_y = \|b\|.MD_x$
Standard Deviation	s_x	$s_y = \|b\|.s_x$
Variance	s^2_x	$s^2_y = b^2.s^2_x$

Note : $|b| = b \text{ if } b \geq 0$

$|b| = -b \text{ if } b < 0$

Illustration 28 : The range, quartile deviation mean deviation and standard deviation for a variable x are 10, 2, 3 and 5 respectively. If $y = 5x + 3$ find the range, quartile deviation, mean deviation and standard deviation for the variable y.

Here, For variable x, Range $R_x = 10$, Quartile deviation $Q_{dx} = 2$, Mean deviation $MD_x = 3$, standard deviation $s_x = 5$.

Now, $y = 5x + 3$. Using the results discussed earlier, the measures of dispersion for the variable y are obtained as follows.

Range $R_y = |5| \cdot R_x = 5(10) = 50$

Quartile deviation $Q_{dy} = |5| \cdot Q_{dx} = 5(2) = 10$

Mean deviation $MD_y = |5| \cdot MD_x = 5(3) = 15$

standard deviation $s_y = |5| \cdot s_x = 5(5) = 25$

Illustration 29 : The demand function of a commodity is $d = 15 - 2p$, where p = price (in ₹) per unit and d = demand (units). From the closing price of each month of the last year, it is known that for the price, range is ₹ 5, mean deviation is ₹ 2 and variance is 9 (₹)^2. Find range, mean deviation and variance of the demand from it.

Here, for price, Range $R_p = 5$ ₹, Mean deviation $MD_p = 2$ ₹ and variance $s_p^2 = 9$ (₹)^2. Now, the demand function is $d = 15 - 2p$. Using the results discussed earlier, we get the following measures for the demand of the commodity :

Range $R_d = |-2| \cdot R_p = 2(5) = 10$ units

Mean deviation $MD_d = |-2| \cdot MD_p = 2(2) = 4$ units

variance $s_d^2 = (-2)^2 \cdot s_p^2 = 4(9) = 36$ (units)^2

Illustration 30 : The range and standard deviation of marks obtained out of 100 in the first test by the students of a school are 80 marks and 20 marks respectively. These marks are divided by 4 for the calculation of the internal marks. Find the range and standard deviation of the marks so obtained.

Here, if marks obtained out of 100 are denoted by x then Range R_x = 80 marks and standard deviation s_x = 20 marks. Now for the calculation of the internal marks these marks are divided by 4. Let us denote the marks so obtained by y . Thus $y = \frac{x}{4} = \frac{1}{4}\ x$

So by using the results discussed earlier range and standard deviation of y are obtained as follows :

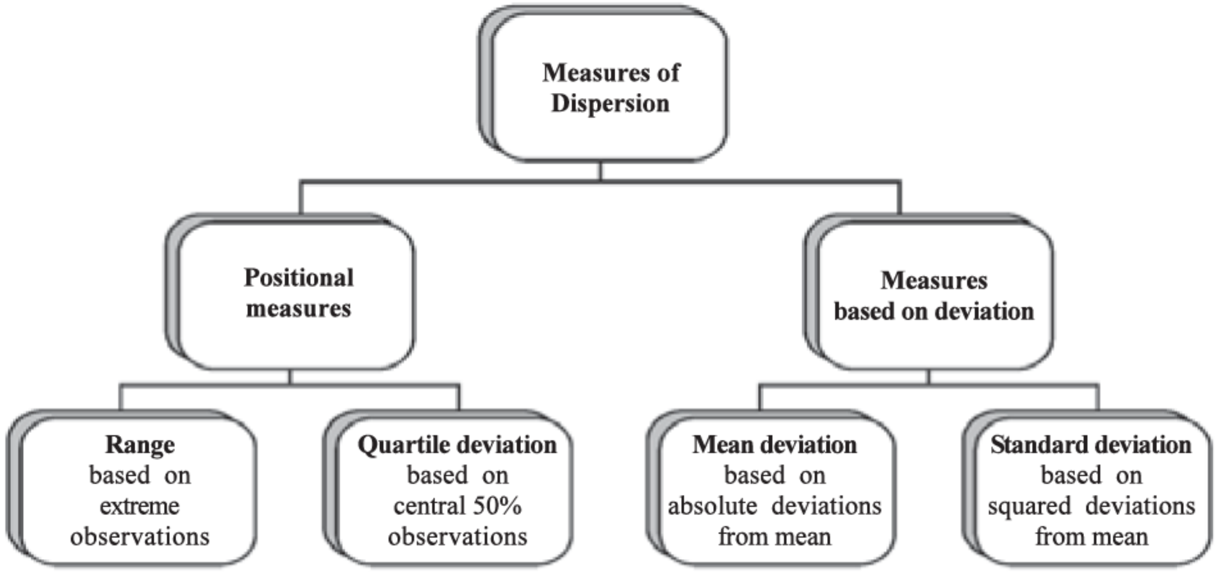
Range $\quad R_y = \left|\frac{1}{4}\right| R_x = \frac{1}{4}\ (80) = 20$ Marks

Standard deviation $s_y = \left|\frac{1}{4}\right| s_x = \frac{1}{4}\ (20) = 5$ Marks

Summary

- **Dispersion or Variation :** It is a measure which shows scatter or spread the observations of a data.
- **Range :** It is a position measure of dispersion obtained by taking the difference between the largest and the smallest value of the data.
- **Quartile Deviation :** It is also a position measure of dispersion. It considers only middle 50% observation. It is also called semi-inter quartile range.
- **Mean Deviation :** It is the mean of the absolute deviations of the observation from its mean.
- **Variance :** The mean of the squares of deviation of all observation from its mean.
- **Standard Deviation :** It is the best measure of dispersion. The standard deviation can be obtained by taking positive square root of the variance s^2.
- **Relative Measures of dispersion :** A measure free from the unit of the variable under study of dispersion is called relative measure. All relative measures of dispersion used to compare two or more groups in terms of their variability.
- **Coefficient of variation :** It is a percentage relative measure of dispersion based on the standard deviation. The lower value of the coefficient of variation suggests consistency of data.

CHAPTER AT A GLANCE



List of Formulae :

	Measure of Dispersion	Absolute Measure	Relative Measure
1.	**Range**	$R = x_H - x_L$	Coefficient of Range $= \frac{x_H - x_L}{x_H + x_L}$
2.	**Quartile Deviation**	$Q_d = \frac{Q_3 - Q_1}{2}$	Coefficient of Quartile Deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$
3.	**Mean Deviation**	$MD = \frac{\Sigma\lvert x_i - \bar{x}\rvert}{n}$ (For Ungroped Data) $MD = \frac{\Sigma f\lvert x - \bar{x}\rvert}{n}$ (For Groped Data)	Coefficient of Mean Deviation $= \frac{MD}{\bar{x}}$
4.	**Standard Deviation**	**For Ungrouped Data :** $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$ OR $\sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$ **Short-cut Method :** $s = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}$ **For Grouped Data :** $s = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{n}}$ OR $\sqrt{\frac{\Sigma f x^2}{n} - \bar{x}^2}$ **Short-cut Method :** When, $\boldsymbol{d = x - A}$ $s = \sqrt{\frac{\Sigma f d^2}{n} - \left(\frac{\Sigma f d}{n}\right)^2}$ When, $\boldsymbol{d = \frac{x - A}{c}}$ $s = \sqrt{\frac{\Sigma f d^2}{n} - \left(\frac{\Sigma f d}{n}\right)^2} \times c$	Coefficient of standard Deviation $= \frac{s}{\bar{x}}$ Coefficient of Variation $= \frac{s}{\bar{x}} \times 100$
5.	**Combined Standard Deviation** $s_c = \sqrt{\frac{n_1(s_1^2 + d_1^2) + n_2(s_2^2 + d_2^2)}{n_1 + n_2}}$		

EXERCISE 4

section A

For the following multiple choice questions choose the correct option :

1. From the following, which is the formula for coefficient of range ?

(a) $x_H - x_L$ (b) $\frac{x_H - x_L}{x_H + x_L}$ (c) $\frac{x_L - x_H}{x_L - x_H}$ (d) $x_L - x_H$

2. In which measure of dispersion, the absolute difference of the observation and its mean is considered ?

(a) Mean Deviation (b) Standard Deviation (c) Range (d) Quartile Deviation

3. Which of the following measures is a unit free measure ?

(a) Mean Deviation (b) Quartile Deviation (c) Range (d) Coefficient of Variation

4. Which measure of dispersion is least affected by the extreme values of the observations?

(a) Range (b) Standard Deviation (c) Quartile Deviation (d) Mean Deviation

5. The coefficient of variation of group A is less than the coefficient of variation of group B. Which group is more consistent with respect to variability ?

(a) A (b) B (c) Both (d) None of these

6. The weight (in kg.) for 10 students are 53, 47, 60, 55, 71, 65, 61, 68, 63, 70.What is the range of the data ?

(a) 17 kg. (b) 23 kg. (c) 24 kg. (d) 18 kg.

7. If the first quartile and the third quartile of a data are 30 and 50 respectively then what is the value of coefficient of quartile deviation ?

(a) 0.25 (b) 50 (c) 4 (d) 20

8. What is the value of any measure of dispersion for the observations 5, 5, 5, 5, 5 ?

(a) 1 (b) 5 (c) 0 (d) 25

9. If mean of a variable is 10 and the coefficient of variation is 60%. What is the variance of the variable ?

(a) 6 (b) 36 (c) 60 (d) 50

10. Suppose the standard deviation of the series k_1, k_2, k_3, ... k_n is 5. What will be the standard deviation of the following series ?

(i) $k_1 + 2$, $k_2 + 2$, $k_3 + 2$, ..., $k_n + 2$

(ii) $3k_1$, $3k_2$, $3k_3$, ..., $3k_n$

(a) (i) 7 (ii) 3 (b) (i) 5 (ii) 3 (c) (i) 5 (ii) 15 (d) (i) 7 (ii) 15

11. The mean and standard deviation for a variable x are 5 and 2 respectively. Now, if $y = 3x + 4$ then what are the mean and the standard deviation of y ?

(a) 19 and 6 (b) 15 and 49 (c) 19 and 10 (d)15 and 10

12. The mean and standard deviation of a set of observations are 45 and 5 respectively. If a constant 5 is added to each observation, what is the coefficient of variation of the new set of the observations ?

(a) 10 % (b) 50 % (c) 11.11 % (d) 900 %

Section B

Give answer in one sentence for the following questions :

1. Define the range.
2. Define the quartile deviation.
3. Which types of measures of dispersion are used for comparing two or more groups in terms of their variability ?
4. Which is the best measure of dispersion ?
5. If the data of heights of 10 students are given in centimeter, what is the unit of its variance ?
6. For a company making pipes, the following information of diameter (in cm) of pipes is obtained. Find the range of the diameter of the pipes.

Diameter (cm)	20 - 40	40 - 60	60 - 80	80 - 100	100 - 120
No. of pipes	15	40	75	20	11

7. The 25th and 75th percentiles of a frequency distribution are 72.18 and 103.99 respectively. Find the quartile deviation.
8. Seven students of a group get 20, 20, 20, 20, 20, 20, 20 marks in a test of 25 marks. What is the standard deviation of their marks.
9. Find the mean deviation for the observations – 1, 0, 4.

Section C

Give answer for the following questions :

1. Define the following :
   (i) Mean Deviation (ii) Standard Deviation (iii) Coefficient of Variation.
2. What is meant by the absolute and relative measures of dispersion ?
3. Write names of the absolute measures of dispersion.
4. Which measures of dispersion are based on the deviations of observation from their mean ?
5. Find the range and the coefficient of range for the observation 6, 11, – 3, 0, 8
6. Find the coefficient of quartile deviation for the observations :
   8, 15, 2, 11, 20, 3, 5
7. Find the mean deviation for the observations :
   3, 8, 1, 7, 6
8. If $\bar{x}$ = 25 and the coefficient of variation is 20 %, find the variance.
9. Find the standard deviation for the observations 1, 2, 3, 4, 5.
10. Which of the following factories is more stable with respect to daily production ?

	Factory A	**Factory B**
Average Daily Production(units)	50	48
Standard deviation (units)	10	12

11. The 25th and the 75th percentiles of a data set are 20 and 36 respectively. Find the coefficient of quartile deviation of the data set.

Section D

Give answer for the following questions :

1. Explain the meaning of dispersion and state different measures of dispersion.
2. State the desirable characteristics of dispersion.
3. Write advantages and disadvantages of the range.
4. Write advantages and disadvantages of the quartile deviation.
5. Write advantages and disadvantages of the mean deviation.
6. Write advantages and disadvantages of the standard deviation.
7. What is standard deviation ? Why is it considered as the best measure of dispersion ?
8. Write a brief note on coefficient of variation.
9. The information of number of flowers on 100 plants of a nursery is given below. Find the quartile deviation of the number of flowers from it.

No. of flowers	11	13	15	17	19	21	23	25
No. of plants	5	8	13	20	22	18	10	4

10. The information of number of goals in 16 matches of hockey tournament is given. Find the mean deviation of number of goals for it.

No. of goals	1	2	3	4	5
No. of matches	1	4	6	4	1

11. In usual notations, $\Sigma d = 25$, $\Sigma d^2 = 272$, $n = 100$ and assumed mean = 4. Find the coefficient of variation.
12. Find the combined standard deviation using the following information :

	Data set A	**Data set B**
No. of observations	50	60
Mean	113	120
Standard deviation	6	7

13. The sum of 10 observations is 80 and the sum of their squares is 800. Find the coefficient of variation of the observations.

Section E

Solve the following :

1. In a language spelling test of 50 marks, the frequency distribution of marks secured by 30 students is given below. Find the mean deviation of the frequency distribution.

Marks	12 - 16	17 - 21	22 - 26	27 - 31	32 - 36
No. of students	2	3	14	8	3

2. Find the quartile deviation of advertisement expenditure using following frequency distribution of advertising expenditure of 50 companies.

Advertisement cost (thousand ₹)	0 - 5	5 - 15	15 - 30	30 - 40	40 - 60	60 - 100	Total
No. of companies	3	8	15	10	8	6	50

3. The information of runs scored by a batsman in his 100 matches is given below. Find the standard deviation of runs scored by him from it.

Runs	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
No. of matches	10	15	25	25	10	10	5

4. The information of marks obtained by 220 students of a college is given below. Find the quartile deviation of the marks obtained by the students.

Marks	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 or more
No. of students	30	50	64	42	29	5

5. Goals scored by two teams in a Football session are as follows.Which team is more consistent in its game ?

No. of goals scored in a football match	**No. of football matches played**	
	Team A	**Team B**
0	15	20
1	10	10
2	7	5
3	5	4
4	3	2
5	2	1

6. For a sequence of 100 observations, the mean and standard deviation are 40 and 10 respectively. In calculating these measures, two observations were taken as 30 and 70 instead of 3 and 27 by mistake. Find the corrected mean and corrected standard deviation.

7. The total cost function for a factory is $y = 10 + 3x$ where x = No. of units produced and y = total cost of producing x units. The range, the quartile deviation, the mean deviation and standard deviation of daily production of the factory are 50, 5, 8 and 10 units respectively. Find the range quartile deviation mean deviation and standard deviation for total cost y from it.

Section F

Solve the following :

1. Find range, coefficient of range, quartile deviation, coefficient of quartile deviation, mean deviation and coefficient of mean deviation from the following data of number of emergency visits of 80 doctors to their patients in a town.

No. of visits	3	5	8	12	17	20	24	30	35
No. of doctors	1	3	7	15	20	13	10	7	4

2. Find the percentage of observations lying within the limits $\bar{x} \pm s$ using the following distribution of credit days taken by the merchants.

Credit days	12	13	14	15	16	17	18	19
No. of merchants	5	10	25	65	45	35	8	7

3. Find an appropriate measure of dispersion from the following data. Also find its relative measure :

Marks	Less than 10	10 -20	20 - 30	30 - 40	Above 40
No. of students	2	4	10	3	1

4. The information of the salary of 200 employees of company is given below. Find the standard deviation of the salary of the employees.

Salary (thousand ₹)	Less than 10	Less than 20	Less than 30	Less than 40	Less than 50	Less than 60	Less than 70
No. of persons	5	17	47	92	142	179	200

5. The following is a distribution of closing prices (in ₹) of shares of 100 different small scale industries on a certain day. Find the mean deviation of the closing prices of shares.

Price (₹)	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 - 90
No. of industries	3	8	15	20	25	10	9	6	4

6. The information of daily wages (in ₹) of 230 workers of a factory is given below. Calculate the coefficient of variation from the following data for the daily wages of workers.

Daily wages (₹)	No. of workers
Less than 100	12
Less than 200	30
Less than 300	65
Less than 400	107
Less than 500	157
Less than 600	202
Less than 700	222
Less than 800	230

7. The marks obtained by two students A and B in 10 sets of examinations are given below :

Set	1	2	3	4	5	6	7	8	9	10
Marks of A	44	80	76	48	52	72	68	56	60	64
Marks of B	48	75	54	60	63	69	72	51	57	56

Which student is more consistent in his study ?

8. The following are the distributions of weights (in kg) for the students of two groups A and B. Find the coefficient of variation of each group. Which group has greater relative variation ?

Weight (kg)	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Group A	7	10	20	18	7
Group B	5	9	21	15	6

Karl Pearson
(1857 - 1936)

Karl Pearson was a major contributor to the early development of statistics. His most famous contribution is the Pearson's chi-square test.

In 1911, he founded the world's first university statistics department at University College, London. He applied statastics to biological problems of heredity and evolution. These papers contain contributions to regression analysis, the correlation coefficeint and include the chi-square test of statistical significane (1900). He coined the term 'standard deviation' in 1893. His work was influenced by the work of Edgeworth and in turn influenced the work of Yule. He was a co-founder of the statistical journal Biometrika.

●