9

MEASURES OF DISPERSION

The previous lesson provides the measure of central tendency that sum up or describe the data into a single representative value. The measures of central tendency may not be adequate to describe data unless we know the manner in which the individual items scatter around it. In other words, a further description of a series on the scatter or variability known as **dispersion** is necessary, if we are to gauge how representative the average is.

Let us take the following three sets.

Students	Group X	Group Y	Group Z
1	50	45	05
2	50	50	45
3	50	55	100
mean $\overline{X}$	50	50	50

Thus, the three groups have same mean i.e. 50. In fact the median of group X and Y are also equal. Now if one would say that the students from the three groups are of equal capabilities, it is totally a wrong conclusion. Close examination reveals that in group X students have equal marks as the mean, students from group Y are very close to the mean but in the third group Z, the marks are widely scattered. It is thus clear that the measures of the central tendency is alone not sufficient to describe the data. The measure of dispersion helps us to know the degree of variability in the data and provide a better understanding of the data.

OBJECTIVES

After completing this lesson, you will be able to:

- know the meaning and need of measures of dispersion;

Notes

- distinguish between absolute and the relative measures of dispersion;
- apply the various measures of dispersion; and
- calculate and compare the different measures of dispersion.

9.1 MEANING OF DISPERSION

Dispersion is the extent to which values in a distribution differ from the average of the distribution.

In measuring dispersion, it is imperative to know the amount of variation (absolute measure) and the degree of variation (relative measure). In the former case we consider the range, Quartile Deviation, standard deviation etc. In the latter case we consider the coefficient of range, coefficient quartile deviation, the coefficient of variation etc.

9.1.1 Absolute and Relative Measures of Dispersion

The dispersion of a series may be measured either absolutely or relatively. If the dispersion is expressed in terms of the original units of the series, it is called absolute measure of dispersion. The disadvantage of absolute measure of dispersion is that it is not suitable for comparative study of the characteristics of two or more series.

For example if the data is expressed in kilograms then the absolute variation will also be expressed in kilograms but if the same data is expressed in grams then the variation will appear 1000 times more. So for comparison point of view it is necessary to calculate the relative measures of dispersion which are expressed as percentage form (i.e. unitless number). These types of expressions are called coefficients of dispersion. Each absolute measure of dispersion has a relative measure of dispersion.

9.2 MEASURES AND METHODS OF COMPUTING DISPERSION

The following are the important measures of dispersion:

1. Range
2. Quartile deviation or Semi-Inter quartile range.
3. Mean deviation
4. Standard deviation
5. Lorenz Curve

Range and Quartile Deviation measure the dispersion by calculating the spread within which the values lie. Mean Deviation and Standard Deviation calculate the extent to which the values differ from the average.

9.2.1 Range

Range (R) is the difference between the largest (L) and the smallest value (S) in a distribution. Thus

$$\text{Range (R)} = L - S$$

Coefficient of Range: It is a relative measure of the range. It is used in the comparative study of the dispersion

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

In case of continuous series Range is just the difference between the upper limit of the highest class and the lower limit of the lowest class.

Range: Evaluation

Range is very simple to understand and easy to calculate. However, it is not based on all the observations of the distribution and is unduly affected by the extreme values. Any change in the data not related to minimum and maximum values will not affect range. It cannot be calculated for open-ended frequency distribution.

Example 1: The amount spent (in ₹) by the group of 10 students in the school canteen is as follows:

110, 117, 129, 197, 190, 100, 100, 178, 255, 790.

Find the range and the co-efficient of the range.

Solution: R = L - S = 790 - 100 = ₹ 690

$$\text{Co-efficient of Range} = \frac{L-S}{L+S} = \frac{790-100}{790+100} = \frac{690}{890} = 0.78$$

Example 2: Find the range and it’s co-efficient from the following data.

Size	10-20	20-30	30-40	40-50	50-100
Frequency	2	3	5	4	2

Solution: R = L – S = 100 – 10 = 90

$$\text{Co-efficient of range} = \frac{L-S}{L+S} = \frac{100-10}{100+10} = \frac{90}{110} = 0.82$$

INTEXT QUESTIONS 9.1

1. The difference between the largest and the smallest data values is the

   (a) variance (b) inter-quartile range

   (c) range (d) coefficient of variation

2. A researcher has collected the following sample data. The mean of the sample is 5.

   3 5 12 3 2

   The range is

   (a) 1 (b) 2 (c) 10 (d) 12

9.2.2 Quartile Deviation

It is based on the lower quartile Q_1 and the upper quartile Q_3. The difference $Q_3 - Q_1$ is called the inter-quartile range. The difference $Q_3 - Q_1$ divided by 2 is called semi-inter-quartile range or the quartile deviation.

$$\text{Thus Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2}$$

9.2.2.1 Coefficient of Quartile Deviation

A relative measure of dispersion based on the quartile deviation is called the coefficient of quartile deviation. It is defined as

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

It is pure number free of any units of measurement. It can be used for comparing the dispersion in two or more than two sets of data.

9.2.2.2 Computation of Quartile Deviation of Ungrouped Data

Example 3: Find out the quartile deviation of daily wages (in ₹) of 7 persons is given below:120,70,150,100,190,170,250

Solution:

Arranging the data in an ascending order we get

70, 100, 120, 150, 170, 190, 250

Here n = 7,

$$Q_1 = \text{Size of } \frac{(N+1)}{4} \text{th item}$$

$$= \text{Size of } \frac{(7+1)}{4} \text{th item} = 2^{nd} \text{item} = 100 \text{ rupees}$$

$$Q_3 = \text{Size of } \frac{3(N+1)}{4} \text{th item}$$

$$= \text{Size of } \frac{3(7+1)}{4} \text{th item} = 6^{th} \text{item} = 190 \text{ rupees}$$

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{190-100}{2} = 45 \text{ rupees}$$

INTEXT QUESTION 9.2

1. If the first quartile is 104 and quartile deviation is 18. Find the third quartile.

Example 4: The wheat production (in Kg) of 20 acres is given as: 1120, 1240, 1320, 1040, 1080, 1200, 1440, 1360, 1680, 1730, 1785, 1342, 1960, 1880, 1755, 1720, 1600, 1470, 1750, and 1885. Find the quartile deviation and coefficient of quartile deviation.

Solution:

After arranging the observations in ascending order, we get 1040, 1080, 1120, 1200, 1240, 1320, 1342, 1360, 1440, 1470, 1600, 1680, 1720, 1730, 1750, 1755, 1785, 1880, 1885, 1960.

$$Q_1 = \text{value of} \left(\frac{N+1}{4}\right) \text{th item}$$

$$= \text{Value of} \left(\frac{20+1}{4}\right) \text{th item}$$

Notes

$= $ Value of (5.25)th item

$= $ 5th item + 0.25(6th item – 5th item)

$= 1240 + 0.25(1320 - 1240)$

$Q_1 = 1240 + 20 = 1260$

$Q_1 = 1240 + 20 = 1260$ kg

$$Q_3 = \text{Value of } \frac{3(N+1)}{4} \text{th item}$$

$$= \text{Value of } \frac{3(20+1)}{4} \text{th item}$$

$= $ Value of (15.75)th item

$= $ 15th item + 0.75(16th item – 15th item)

$= 1750 + 0.75\ (1755 - 1750)$

$Q_3 = 1750{+}3.75 = 1753.75$ kg

Quartile Deviation (Q.D)

$$= \frac{Q_3 - Q_1}{2} = \frac{1753.75 - 1260}{2} = \frac{492.75}{2} = 246.875$$

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{1753.75 - 1260}{2}$$

$$= \frac{492.75}{2} = 246.875\text{kg.}$$

Coefficient of Quartile Deviation

$$= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1753.75 - 1260}{1753.75 + 1260} = 0.164$$

Computation of Q.D. for a frequency distribution

9.2.2.2 Computation in case of Discrete Series:

Example 5: The Tax authority collected the following amount of tax from different firms in a particular market.

Notes

Amount of Taxes (in 000 ₹)	10	11	12	13	14
No. of Firms	3	12	18	12	3

Calculate the quartile deviation and the coefficient of quartile deviation.

Solution:

Table 9.1: Calculation of Quartile deviation

Amount of Taxes (in ’000 ₹)	No. of Firms (f)	Cummulative Frequency (C.F.)
10	3	3
11	12	15
12	18	33
13	12	45
14	3	48
	Σf = 48	

Here N = 48,

$$Q_1 = \text{Size of } \frac{(N+1)}{4} \text{th item}$$

$$= \text{Size of } \frac{(48+1)}{4} \text{th item}$$

$$= \text{Size of } 12.25^{\text{th}} \text{ item} = 11 \text{ (in '000 rupees)}$$

$$Q_3 = \text{Size of } \frac{3(N+1)}{4} \text{th item}$$

$$= \text{Size of } \frac{3(48+1)}{4} \text{th item}$$

$$= \text{Size of } 36.75^{\text{th}} \text{ item} = 13 \text{(in '000 rupees)}$$

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{13-11}{2} = 1 \text{ (in '000 rupees)}$$

$$\text{Coeff of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{13-11}{13+11} = 0.083$$

9.2.2.3 Computation of Q.D. for a Continuous Series

Example 6: Calculate quartile deviation and coefficient of quartile deviation from the following distribution:

Weekly Wages(in '000 ₹)	5-7	8-10	11-13	14-16	17-19
No. of Workers	14	24	38	20	04

Solution:

Table 9.2: Calculation of Quartile deviation and coefficient of quartile deviation

Weekly Wages (in '000 ₹)	No. of Workers (f)	Cummulative Frequency (C.F.)
4.5-7.5	14	14
7.5-10.5	24	38
10.5-13.5	38	76
13.5-16.5	20	96
16.5-19.5	4	100
	$\Sigma f = 100$	

$$Q_1 = l_1 + \frac{l_2 - l_1}{f}\left(\frac{N}{4} - cf\right)$$

$\frac{N}{4} = 25$. Q_1 lies in the class of 7.5-10.5

$$Q_1 = l_1 + \frac{l_2 - l_1}{f}\left(\frac{N}{4} - cf\right) = 7.5 + \frac{25-14}{24} \times 3 = 8.875 \text{ (in ₹000)}$$

$$Q_3 = l_1 + \frac{l_2 - l_1}{f}\left(\frac{3N}{4} - cf\right) = 10.5 + \frac{75-38}{38} \times 3 = 13.42 \text{ (in ₹000)}$$

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{13.42 - 8.875}{2} = 2.273 \text{ (in ₹000)}$$

$$\text{Coeff. of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{13.42 - 8.875}{13.42 + 8.875} = 0.21$$

Quartile Deviation: An Evaluation

Since Quartile deviation is based only on Q_1 and Q_3, it means this measure is based on middle 50% of the data of the series. Thus unlike range, Quartile deviation is not affected' by extreme items as it ignores 25% of the data from the beginning of the dataset and 25% of the data from the end (data arranged in ascending order). It can be calculated in case of open-ended distribution. However, it is not based on all the observations in the data.

INTEXT QUESTIONS 9.3

1. Which of the following is a measure of dispersion?
   - (a) percentiles
   - (b) quartiles
   - (c) inter-quartile range
   - (d) all of the above are measures of dispersion
2. The inter-quartile range is
   - (a) the 50th percentile
   - (b) another name for the standard deviation
   - (c) the difference between the largest and smallest values
   - (d) the difference between the third quartile and the first quartile
3. Which of the following limitation of the range is overcome by the inter-quartile range?
   - (a) the sum of the range variances is zero
   - (b) the range is difficult to compute
   - (c) the range is influenced too much by extreme values
   - (d) the range is negative
4. A researcher has collected the following sample data. The mean of the sample is 5.

   3 5 12 3 2

   The inter-quartile range is:

   (a) 1 (b) 2 (c) 10 (d) 12

9.2.3 Mean Deviation

Mean deviation (MD) of a series is the arithmetic average of the deviation of various items from a measure of central tendency (mean, median and mode)

Mean deviation is based on the items of the distribution and is calculated as an average, on the basis of deviation obtained from either mean, median or mode but generally from the median.

First we compute deviations of all the items from either mean or median ignoring plus (+) and (–) signs. They are called absolute values of deviations where the two parallel bars (ii) indicate that the absolute value is taken. This is also called modulas value. Then the aggregate of these deviations are divided by the number of observations this is called mean deviation.

9.2.3.1 Calculation of mean deviation

(i) Arrange the data in ascending order (for calculating of median)

(ii) Calculate median/mean/mode

(iii) Take deviations of items from median/mean ignoring ± signs and denote the column as $|D|$

(iv) Calculate the sum of these deviation in case of discrete and continuous series $|D|$ is multiplied by respective frequency of the item to get $\Sigma f\,|D|$

(v) Divide the total obtained by number of items to get mean deviation

$$\text{M.D.} = \frac{\Sigma f\,|D|}{N}$$

(vi) apply the formula to get coefficient of mean deviation

$$\text{Coefficient of M.D.} = \frac{\text{M.D}}{\text{Median / Mean / Mode}}$$

Example 6: Calculate mean deviaiton and coefficient of mean deviation from both mean and median for the following data on the monthly income (in ₹) of households

Income (₹)	8520	6350	7920	8360	7500

Table 9.3: Calculation of mean deviaiton from median

Monthly Income (₹)	Deviation from mean (7730) ignoring ± signs |D|	Deviations from median (7920) ignoring ± signs |D|
6350	1380	1570
7500	230	420
7920	190	0
8360	630	440
8520	790	600
$\Sigma X = 38650$	$\Sigma\|D\| = 3220$	$\Sigma\|D\| = 3030$

$$\text{Mean} = \frac{\Sigma X}{N}$$

$$= \frac{38650}{5} = 7730$$

$$\text{M.D.} = \frac{\Sigma|D|}{N}$$

$$= \frac{3220}{5} = \text{Rs } 644$$

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

$$= \frac{644}{7730} = 0.083$$

$$\text{and Median} = \text{Size of} \left(\frac{N+1}{2}\right)^{th} \text{item}$$

$$= \text{Size of} \left(\frac{5+1}{2}\right)^{th} \text{item}$$

$$= \text{Size of 3rd item}$$

$$= ₹\ 7920$$

$$\text{Mean Deviation} = \frac{\Sigma|D|}{N}$$

$$= \frac{3030}{5} = \text{Rs } 606$$

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

$$= \frac{606}{7920} = 0.076$$

9.2.3.2 Calculation of mean deviaiton in discrete series

Example 7: Calculate (a) median (b) mean deviation and (c) Coefficient of mean deviation

Size of item (X)	6	12	18	24	30	36	42
Frequency (f)	4	7	9	18	15	10	5

Solution.

Table 9.4: Calculation of mean deviaiton from Median

X	f	cf	|D|	f|D|
6	4	4	18	72
12	7	11	12	84
18	9	20	6	54
24	18	38	0	0
30	15	53	6	90
36	10	63	12	120
42	5	68	18	90
	$\Sigma f = 68$		$\Sigma\|D\| = 72$	$\Sigma f\|D\| = 510$

$$\text{Median} = \text{Size of} \left(\frac{N+1}{2}\right)^{th} \text{item}$$

$$= \text{Size of} \left(\frac{68+1}{2}\right)^{th} \text{item}$$

$$= 34.5\text{th item}$$

$$\text{M.D.} = \frac{\Sigma f|D|}{N} = \frac{510}{68} = 7.5$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Median}} = \frac{7.5}{24} = 0.312$$

Calculation of mean deviaiton in continuous series

Example 8: Calculate (i) mean, (ii) mean deviation from mean and (iii) co-efficient of mean deviation.

Marks	0-10	10-20	20-30	30-40	40-50
No. of students	5	8	15	16	6

Solution: Calculation of mean deviation from mean

Table 9.5: Calculation of mean deviaiton from Mean

Marks X	No of students f	Mid-point m	$\frac{m-25}{10}$	fd′	|D| = m – 27	f|D|
0-10	5	5	–2	–10	22	110
10-20	8	15	–1	–8	12	96
20-30	15	25	0	0	2	30
30-40	16	35	+1	+16	8	128
40-50	6	45	+2	+12	18	108
	Σf = 50			Σfd′ = 10		Σf|D|= 472

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times c$$

$$= 25 + \frac{10}{50} \times 10 = 27 \text{ Marks}$$

$$\text{M.D.} = \frac{\Sigma f|D|}{N} = \frac{472}{50} = 9.44 \text{ Marks}$$

$$\text{Coefficient of M.D.} = \frac{\text{M.D.}}{\text{Mean}} = \frac{9.44}{27} = 0.349$$

Mean deviaiton: An evaluation

Mean deviation ignores the ± signs of the deviation which is mathematically unsound and illogical. Therefore, this method is non-algebraic. Moreover, it can not be computed for distribution for open end classes.

INTEXT QUESTIONS 9.4

(i) Calculated mean deviation and co-efficient of mean deviation from median

No of tomatoes per plant	0	1	2	3	4	5	6	7	8	9	10
No of plants	2	5	7	11	18	24	12	8	6	4	3

(ii) Calculate mean deviation from mean

Class:	3-4	4-5	5-6	6-7	7-8	8-9	9-10
Frequency	3	7	22	60	85	32	8

9.2.4 Standard Deviation (S. D.)

Standard deviation is the most important and commonly used measure of dispersion. It measures the absolute dispersion or variability of a distribution. Standard deviation is the positive square root of the mean of the squared deviations of observations from their mean. It is denoted by S.D. or σ_x.

9.2.4.1 Computation of Standard deviation in case of Individual Series

The following four methods are used to calculate the standard deviation:

1. Actual Mean Method

Let X variable takes on N values i.e. $X_1, X_2, \ldots X_N$. The standard deviation of these N observations using actual mean method can be computed as follows:

1. Obtain the arithmetic mean $(\bar{X})$ of the given data.
2. Obtain the deviation of each ith observation from $\bar{X}$ i.e. $(X_i - \bar{X})$. (Note that $\Sigma(X_i - \bar{X}) = 0$)
3. Square each deviation i.e. $(X_i - \bar{X})^2$
4. Obtain the sum in step 3 i.e. $\sum_{i=1}^{N} (X_i - \bar{X})^2$
5. Obtain the square root of the mean of these squared deviations as follows:

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

N = Total No. of observation

2. Assumed Mean Method

This method is applied to calculate the standard deviation when the mean of the data is in fraction. In that case it is difficult and tedious to find the deviation of all observations from the actual mean by the above method. Thus the deviations (d) are taken from the Assumed mean (A) and standard deviation is estimated by using the following formula:

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

where d = (X – A) i.e. deviation taken from the assumed mean (A)

3. Direct Method

The relevance of this method is particularly useful when the items are very small. To obtain standard deviations, we apply the following formula:

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\sum X^2}{N} - (\bar{X})^2}$$

where $\bar{X}$ = arithemetic mean

(Note: The direct method basically implies taking deviations from zero)

4. Step Deviation Method

In this method we divide the deviations by a common class interval (c) and use the following formula for computing standard deviation:

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\sum d'^2}{N} - \left(\frac{\sum d'}{N}\right)^2} \times c$$

where $d' = \left(\frac{X-A}{c}\right)$ i.e. deviation taken from the assumed mean and divide by class interval (c)

Example 9: The wholesale price of a commodity for 6 days in a month of February 2014 is as follows:

Days	1	2	3	4	5	6
Commodity Price(₹ Per Quintal)	5	15	25	35	45	55

Compute the standard deviation using:

(i) Actual Mean Method

(ii) Assumed Mean Method

(iii) Direct Method and

(iv) Step-Deviation Method

Solution:

Table 9.6: Calculation of standard deviation

Days	Price (₹ Per Quintal)	$(X - \bar{X})$ = (X – 30)	$(X - \bar{X})^2$ = $(X - 30)^2$	d = (X – 40)	d^2 = $(X - 40)^2$	X^2	$d' = \frac{X - 40}{5}$	d'^2
1	5	-25	625	–35	1225	25	–7	49
2	15	-15	225	–25	625	225	–5	25
3	25	-5	25	–15	225	625	–3	9
4	35	5	25	–5	25	1225	–1	1
5	45	15	225	5	25	2025	1	1
6	55	25	625	15	225	3025	3	9
		$\sum(X - \bar{X})$ = 0	$\sum(X - \bar{X})^2$ = 1750	$\sum d$ = –60	$\sum d^2$ = 2350	$\sum X^2$ = 7150	$\sum d'$ = –12	$\sum d'^2$ = 94

Applying Actual Mean Method

$$\bar{X} = \frac{\sum X}{N} = \frac{180}{6} = 30 \text{ (in Rupees)}$$

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum(X - \bar{X})^2}{N}} = \sqrt{\frac{1750}{6}} = 17.078 \text{ (in rupees)}$$

Applying Assumed Mean Method:

Here $\quad \bar{X} = A + \frac{\sum d}{N} = 40 + \frac{-60}{6} = 30$

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\sum d^2}{N} - \left(\frac{\sum d}{N}\right)^2}$$

$$\sum d^2 = 2350, \ \sum d = -60, N = 6$$

$$\therefore \quad (\sigma_X) = \sqrt{\frac{2350}{6} - \left(\frac{-60}{6}\right)^2} = \text{Rs.}17.078$$

Applying Direct Method

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\sum X^2}{N} - (\bar{X})^2}$$

where $\bar{X} = 30,\ \sum X^2 = 7150,\ N = 6$

$$\therefore \qquad (\sigma_X) = \sqrt{\frac{7150}{6} - (30)^2} = \text{Rs.}17.078$$

Applying Step Deviation Method

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\sum d'^2}{N} - \left(\frac{\sum d'}{N}\right)^2} \times c$$

where $c = 5, \sum d'^2 = 94, N = 6, \sum d' = -12$

$$\therefore \qquad (\sigma_X) = \sqrt{\frac{94}{6} - \left(\frac{-12}{6}\right)^2} \times 5 = \text{Rs.}17.078$$

Note: The sum of deviations taken from mean is Zero. But the sum of deviations from a value other than actual mean is not equal to zero

9.2.4.2 Computation of Standard Deviation in case of Continuous Series

In continuous series, the class-interval and frequencies are given. The following methods are used to compute standard deviation in this case:

1. Actual Mean Method

In this method the following steps are involved:

- Calculate the mean of the distribution.
- Estimate deviations of mid-values from the actual mean i.e. $x = m - \bar{X}$.
- Multiply the deviations with their corresponding frequencies to get 'fx'. [Note that $\Sigma fx = 0$].
- Calculate fx^2 values by multiplying 'fx' values with 'x' values and sum up these to get $\sum fx^2$.
- Apply the following formula to obtain standard deviation:

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\Sigma fx^2}{\Sigma f}}$$

where x = $(m - \bar{X})$ i.e. deviation taken from the arithemetic mean $(\bar{X})$

2. Assumed Mean Method

The steps involved in the calculation of standard deviation are as follows:

- Calculate mid-points (i.e. m) of classes.
- Estimate the deviations of mid-points from the assumed mean (A) i.e. d = m – A.
- Multiply values of'd' with corresponding frequencies to get 'fd' values (note that the total of this column is not zero since deviations have been taken from assumed mean).
- Apply the following formula to calculate standard deviation:

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$$

where d = (m – A) i.e. deviation taken from the assumed mean (i.e. A)

3. Step Deviation Method

The steps involved in the calculation of standard deviation are as follows:

- Calculate class mid-points (m) and deviations (d) from an arbitrarily chosen value just like in the assumed mean method. i.e. d= m-A.
- Divide the deviations by a common factor 'C' denoted by $d' = \left(\frac{m - A}{c}\right)$.
- Multiply d' values with corresponding f values to obtain fd' values.
- Multiply fd' values with d' values to get fd'^2 values.
- Obtain $\Sigma fd'$ and $\Sigma fd'^2$ values.
- Apply the following formula.

$$\text{Standard Deviation } (\sigma_x) = \sqrt{\frac{\Sigma fd'^2}{\Sigma f} - \left(\frac{\Sigma fd'}{\Sigma f}\right)^2} \times c$$

where $d' = \left(\frac{m - A}{c}\right)$ i.e. deviation taken from the assumed mean and divide by class interval (c) (or the common factor in case the class intervals are unequal), m is the mid value of the interval.

Standard Deviation: Interesting Properties

1. The value of Standard Deviation remains same if each observation in a series is increased or decreased by a constant value i.e. Standard deviation is independent of change of origin.
2. The value of Standard Deviation changes if each of observation in a series is multiplied or divided by a constant value i.e. Standard deviation is not independent of change of scale.

Example 10: A study of 1000 companies gives the following information

Profit (in ₹ crores)	0-10	10-20	20-30	30-40	40-50	50-60
No. of Companies	10	20	30	50	40	30

Calculate the standard deviation of the profit earned.

(i) Actual Mean Method

(ii) Assumed Mean Method

(iii) Step-Deviation Method

Solution:

Table 9.7: Calculation of standard deviation

Profit (in ₹ crores)	**No. of Companies**	**m**	**fm**	**d = (m – 40)**	**$d' = \frac{m-45}{10}$**	**fd**	**fd^2**	**fd′**	**fd'^2**
0-10	10	5	50	–35	–4	–350	12250	–40	160
10-20	20	15	300	–25	–3	–500	12500	–60	180
20-30	30	25	750	–15	–2	–450	6750	–60	120
30-40	50	35	1750	–5	–1	–250	1250	–50	50
40-50	40	45	1800	5	0	200	1000	0	0
50-60	30	55	1650	15	1	450	6750	30	30
			6300	**Σd = –60**	**Σd′ = –9**	**–900**	**40500**	**–180**	**540**

Applying Actual Mean Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fx^2}{\sum f}}$$

$\sum fx^2 = 36000, \sum f = 180$

$$\therefore (\sigma_X) = \sqrt{\frac{36000}{180}} = 14.142 \text{(in rupees crores)}$$

Applying Assumed Mean Method:

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

$\sum fd^2 = 40500, \sum fd = -900, \sum f = 180, d = X - 40$

$$\therefore (\sigma_X) = \sqrt{\frac{40500}{180} - \left(\frac{-900}{180}\right)^2} = 14.142 \text{(in rupees crores)}$$

Applying Step Deviation Method:

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times c$$

$d' = \frac{m - 45}{10}, \sum fd'^2 = 540, \sum fd' = -180, \sum f = 180, c = 10$

$$\therefore (\sigma_X) = \sqrt{\frac{540}{180} - \left(\frac{-180}{180}\right)^2} \times 10 = 14.142 \text{(in rupees crores)}$$

Example 11: The following table shows the daily wages of a random sample of construction workers. Calculate its mean deviation and standard deviation.

Daily Wages (₹)	Number of Workers
200 - 399	5
400 - 599	15
600 - 799	25
800 - 999	30
1000 - 1199	18
1200 - 1399	7
Total	100

Solution:

Table 9.8: Computation of mean deviation

Daily Wages (₹) X	Number of Workers (f)	Class Mark (m)	fm	$f_i \lvert m - \bar{X} \rvert$ $= f_i \lvert m - 823.5 \rvert$
200 – 399	5	299.5	1497.50	2,620
400 – 599	15	499.5	7492.50	4,860
600 – 799	25	699.5	17487.50	3,100
800 – 999	30	899.5	26985.00	2,280
1000 – 1199	18	1,099.5	19791.00	4,968
1200 – 1399	7	1,299.5	9096.50	3,332
Total	100		82350.00	21,160

$$\text{Mean deviation } = \frac{\Sigma f_i \lvert m - \bar{X} \rvert}{\Sigma f_i} = \frac{21,160}{100} = 211.60 \text{ (₹)}$$

Table 9.9: Computation of Standard deviation

Daily Wages (₹)	Number of Workers	Class Mark (M.V.)	$f_i (m - \bar{X})^2$
200 – 399	5	299.5	1, 372,880
400 – 599	15	499.5	1,574,640
600 – 799	25	699.5	384,400
800 – 999	30	899.5	173,280
1000 – 1199	18	1,099.5	1,371,168
1200 – 1399	7	1,299.5	1,586,032
Total	100		6,462,400

$$\text{Standard deviation } = \sqrt{\frac{6462400}{100}} = 254.21 \text{ (Rupees)}$$

INTEXT QUESTION 9.5

1. Sona, Karina, Omar, Mustafa and Amie obtained marks of 6, 7, 3, 7, 2 on a standardized test respectively. Find the standard deviation of their scores.

9.2.4.3 Comparison of the variation of two series using standard deviation

The values of the standard deviations cannot be used as the basis of the comparison mainly because units of measurements of the two distributions may be different. The correct measure that should be used for comparison purposes is the **Coefficient of Variation (C.V.)** given by **Karl Pearson**:

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

$$\sigma_X = \text{S.D. of variable } X, \quad \overline{X} = \text{mean of variable } X$$

Example 12: The following table shows the summary statistics for the daily wages of two types of workers.

Worker's Type	Daily Wages	
	Mean	Standard deviation
I	₹ 100	₹ 20
II	₹ 150	₹ 24

Compare these two daily wages distributions.

Solution:

Table 9.10: Calculation of coefficient of variations

In comparison	Distribution	Reason
Average magnitude	II > I	$\overline{X}_{II} = 150 > \overline{X}_I = 100$
Variation	I > II	$CV_I = \frac{20}{100} \times 100 = 20\% > CV_{II}$
		$= \frac{24}{150} \times 100 = 16\%$

INTEXT QUESTIONS 9.6

1. The hourly wages of a sample of 130 system analysts are given below:

   mean = 60 range = 20

   mode = 73 variance = 324

   median = 74

   The coefficient of variation equals

   (a) 0.30% (b). 30% (c) 5.4% (d) 54%

2. The measure of dispersion that is influenced most by extreme values is:

   (a) the standard deviation

   (b) the range

   (c) the inter-quartile range

3. The descriptive measure of dispersion that is based on the concept of a deviation about the mean is

   (a) the range (b) the inter-quartile range

   (c) the absolute value of the range (d) the standard deviation

4. The numerical value of the standard deviation can never be

   (a) zero (b) negative (c) one

5. A researcher has collected the following sample data. The mean of the sample is 5.

   3 5 12 3 2

   The Standard deviation is

   a. 8.944 b. 4.062 c. 13.2 d. 16.5

   Answer: b

   And

   The Coefficient of Variation is

   a. 72.66% b. 81.24% c. 264% d. 330%

   Answer: b

9.2.5 Lorenz Curve

Lorenz curve is the graphical method of studying dispersion. Lorenz curve is the cumulative frequency curve showing the distribution of a variable such as population against any independent variable such as income or area settled. If the distribution of the dependent variable is equal, the plot will show as a straight, 45° line. Unequal distributions will yield a curve. The gap between this curve and the 45° line is the inequality gap. The farther the curve from this 45° line, the greater is the variability present in the distribution. Lorenz curve is used to see the degree of concentration of income or health. For example, it may show top 25% of population accounts for 70% of income or bottom 25% of population has only 5% of income (see figure 9.1).

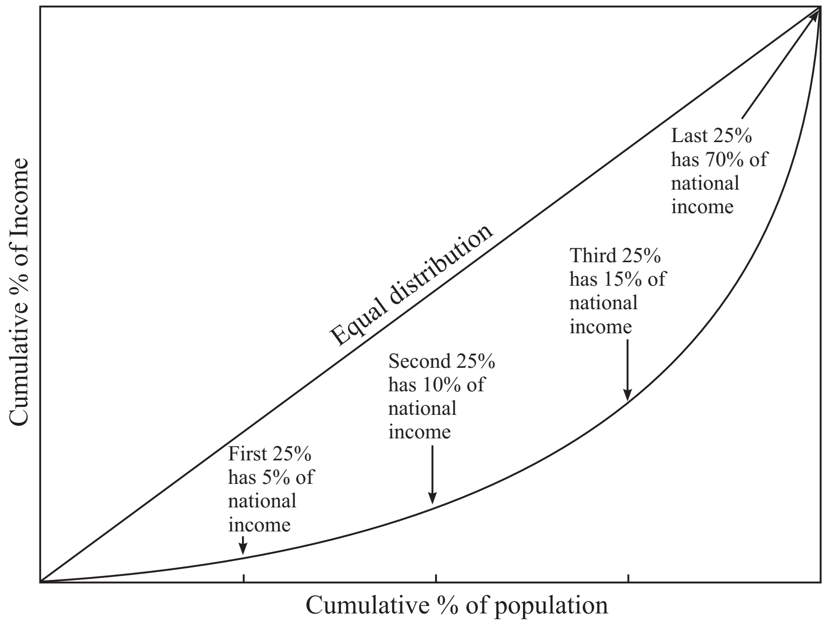

Fig. 9.1

Steps involved in drawing Lorenz Curve:

The drawing of Lorenz curve requires following steps:

1. Find cumulative totals of variables. In case of continuous variable calculate the cumulative totals of mid-points.
2. Estimate cumulative frequencies.
3. Express the cumulative mid-points and frequencies into percentages by taking each of the sum total as 100.
4. Take the cumulative percentages of the variable on Y axis and cumulative percentages of frequencies on X-axis. Each axis will have values from '0' to '100'.
5. Draw a line joining Co-ordinate (0, 0) with (100,100). This is called the line of equal distribution.
6. Plot the cumulative percentages of the variable with corresponding cumulative percentages of frequency. Join these points to get the Lorenz Curve.

WHAT YOU HAVE LEARNT

- The important measures of dispersion are:
  i) Range
  ii) Quartile deviation or Semi-Inter quartile range.
  iii) Standard deviation
  iv) Lorenz Curve
- Range (R) is the difference between the largest (L) and the smallest value (S) in a distribution i.e. Range (R) = L – S
- The Coefficient of Range is the relative measure of the range and is given by:
  $$\frac{L-S}{L+S}$$
- Quartile Deviation (Q.D) is given by $\text{Q.D.} = \frac{Q_3 - Q_1}{2}$
- The Coefficient of quartile deviation is given by $\text{Coeffof Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$
- Standard deviation is the most important and commonly used measure of dispersion It is denoted by S.D. or σ_x.
- Standard deviation in case of Individual Series is given by four methods:

(i) Actual Mean Method

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

N = Total No. of observation

(ii) Assumed Mean Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

where d = (X – A) i.e. deviation taken from the assumed mean (A)

Notes

(iii) Direct Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum X^2}{N} - (\bar{X})^2}$$

where $\bar{X}$ = arithemetic mean

(iv) Step Deviation Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum d'^2}{N} - \left(\frac{\sum d'}{N}\right)^2} \times c$$

where $d' = \left(\frac{X - A}{c}\right)$ i.e. deviation taken from the assumed mean and divide by class interval (c)

- Standard Deviation in case of Continuous Series is given by

(i) Actual Mean Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fx^2}{\sum f}}$$

where $x = (m - \bar{X})$ i.e. deviation taken from the arithemetic mean $(\bar{X})$

(ii) Assumed Mean Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fd^2}{\sum f} - \left(\frac{\sum fd}{\sum f}\right)^2}$$

where d = (m – A) i.e. deviation taken from the assumed mean (i.e. A)

(iii) Step Deviation Method

$$\text{Standard deviation } (\sigma_x) = \sqrt{\frac{\sum fd'^2}{\sum f} - \left(\frac{\sum fd'}{\sum f}\right)^2} \times c$$

where $d' = \left(\frac{m - A}{c}\right)$ i.e. deviation taken from the assumed mean and divide by class interval (c) (or the common factor in case the class intervals are unequal, m is the mid value of the interval.

Notes

- Standard deviation is independent of change of origin but not independent of change of scale.
- The coefficient of variation (C.V.) is the relative measure of dispersion which is used for the comparison of variability of two or more distributions. It is given by:

$$C.V. = \frac{\sigma_X}{\bar{X}} \times 100$$

σ_X = S.D. of variable X, $\bar{X}$ = mean of variable X

- Lorenz curve is the graphical method of estimating dispersion.

TERMINAL EXERCISES

Range

1. The following are the prices of shares of A B Co. Ltd. from Monday to Saturday:

Days	Price (in ₹)	Days	Price (in ₹)
Monday	200	Thursday	160
Tuesday	210	Friday	220
Wednesday	208	Saturday	250

Calculate range and its coefficient

2. Find the range of given data

108, 107, 105, 106, 107, 104, 103, 101, 104

3. Find the value of range of frequency distribution

Age in years:	14	15	16	17	18	19	20
No. of students :	1	2	2	2	6	4	0

4. Calculate the range for the distribution given below

Height in cms	150	151	152	154	159	160	165	166
No. of Boys	2	2	9	15	18	10	4	1

5. Find the range of the following data

Profit (in '000 ₹):	0-10	10-20	20-30	30-40	40-50
No. of firms	0	6	0	7	15

Notes

6. Find the range of the following distribution

Class Interval	10-20	20-30	30-40	40-50	50-60
Frequency	8	10	15	18	19

Quartile Deviation

7. Calculate the QD for a group of data,

241, 521, 421, 250, 300, 365, 840, 958

8. From the following figures find the quartile deviation and its coefficient:

Height (cms.):	150	151	152	153	154	155	156	157	158
No. of Students:	15	20	32	35	33	22	20	12	10

9. Using quartile deviation, state which of the two variables – A and B is more variable:

A		**B**	
Mid-Point	**Frequency**	**Mid-Point**	**Frequency**
15	15	100	340
20	33	150	492
25	56	200	890
30	103	250	1420
35	40	300	620
40	32	350	360
45	10	400	187
		450	140

10. Find the quartile deviation from the following table:

Size:	4-8	8-12	12-16	16-20	20-24	24-28	28-32	32-36	36-40
Frequency:	6	10	18	30	15	12	10	6	2

11. Calculate the coefficient of quartile deviation from the following data:

Class Interval	**Frequency**
10–15	4
15–20	12
20–25	16
25–30	22
30–40	10
40–50	8
50–60	6
60–70	4
	8

Standard deviation

12. Determine the standard deviation of the following student test results percentages.

92% 66% 99% 75% 69% 51% 89% 75% 54% 45% 69%

13. Calculate the coefficient of variation for the following data set.

The price (in ₹), of a stock over five trading days was 52, 58, 55, 57, 59.

14. The frequency table of the monthly salaries of 20 people is shown below.

Salary (in ₹)	Frequency
3500	5
4000	8
4200	5
4300	2

(a) Calculate the mean of the salaries of the 20 people.

(b) Calculate the standard deviation of the salaries of the 20 people.

15. The following table shows the grouped data, in classes, for the heights of 50 people.

Height (in cm) – classes	Frequency
$120 \leq 130$	2
$130 \leq 140$	5
$140 \leq 150$	25
$150 \leq 160$	10
$160 \leq 170$	8

a) Calculate the mean of the salaries of the 50 people.

b) Calculate the standard deviation of the salaries of the 50 people.

16. The following is the frequency distribution for the speeds of a sample of automobiles traveling on an interstate highway.

Speed Miles per Hour	Frequency
50 – 54	2
55 – 59	4
60 – 64	5
65 – 69	10
70 – 74	9
75 – 79	5
	35

Calculate the mean, and the standard deviation of speed.

17. In 2012, the average age of workers in a company was 22 with a standard deviation of 3.96. In 2013, the average age was 24 with a standard deviation of 4.08. In which year do the ages show a more dispersed distribution? Show your complete work and support your answer.

Therefore the year 2012 shows a more dispersed distribution.

18. The following is a frequency distribution for the ages of a sample of employees at a local company.

Age (in years)	Frequency
30 – 39	2
40 – 49	3
50 – 59	7
60 – 69	5
70 – 79	1

(a) Determine the average age for the sample.

(b) Compute the standard deviation.

(c) Compute the coefficient of variation.

19. The population change between 1990 and 2000 for several small cities are shown below.

City	Population Change (number of residents)
A	3083
B	1466
C	-461
D	1113
E	-11
F	395
G	3290
H	437

For the above **sample**, determine the following measures.

(a) The mean

(c) The standard deviation

(d) The median

ANSWERS TO INTEXT QUESTIONS

9.1

1. (c)
2. (c)

9.2

1. 140

9.3

1. (c)
2. (d)
3. (c)
4. (b)

9.4

1. Median = 5, M.D. = 1.68
2. M.D. 0.915, Coefficient of M.D. = 0.336

9.5

1. 2.1 marks

9.6

1. (b)
2. (b)
3 (d)
4. (b)
5. Standard deviation (b) C.V