CHAPTER 32

STATISTICS

32.1 INTRODUCTION

In earlier classes, we have learnt about methods of representing data graphically and in tabular form. Such representations exhibit certain characteristics or salient features of the data. We have also studied various methods of finding a representative value of the given data. This value is called the central value for the given data and various methods for finding the central value are known as the measures of central tendency. The measures of central tendency are: mean (arithmetic mean), median and mode. We have learnt that the measures of central tendency give us one single figure that represents the entire data i.e., they give us one single figure around which the observations are concentrated. In other words, measures of central tendency give us a rough idea where observations are centred. But the central values are inadequate to give us a complete idea of the distribution as they do not tell us the extent to which the observations vary from the central value. In order to make better interpretation from the data, we should also have an idea how the observations are scattered or how much they are bunched around a central value. There can be two or more distributions having the same central value but still there can be wide disparities in the formation of the distribution as discussed below.

Consider following three distributions:

(i) 1, 5, 9, 13, 17 (ii) 3, 6, 9, 12, 15 (iii) 7, 8, 9, 10, 11

In all these distributions we have the same number of observations and the same mean and median both equal to 9. Therefore, if we are given that the mean of 5 observations is 9, we are unable to say whether it is the average of first distribution or second distribution or third distribution.

Let us now plot these distributions on a number line as shown below:

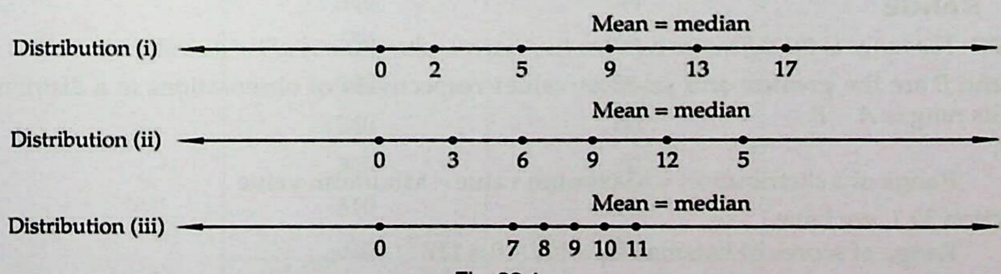

Fig. 32.1

We observe that the dots representing observations in distribution (iii) are more close to each other and are clustering around the mean and median (central value). So, we say that there is more variability in the values of observations in distribution (i) in comparison to distributions (ii) and (iii). We can also say that the distribution (iii) is more consistent than distributions (i) and (ii).

Let us now consider the runs scored by two batsmen B_1 and B_2 in their last ten matches as given below:

Match:	1	2	3	4	5	6	7	8	9	10
Batsman B_1:	30	91	0	64	42	80	30	5	117	71
Batsman B_2:	53	46	48	50	53	53	58	60	57	52

The mean and median of the scores are as under:

	Mean	Median
Batsman B_1	53	53
Batsman B_2	53	53

We observe that the mean and median of the runs scored by both the batsmen B_1 and B_2 are same. On the basis of this a natural question arises : Is the performance of two players same? The answer is of course not in affirmative. Because the variability in the scores of batsman B_1 is more as he has scored runs from 0 (minimum) to 117 (maximum), where as the batsman B_2 has scored runs more consistently as the runs scored by him vary from 46 (minimum) to 60 (maximum). If the scores of batsmen B_1 and B_2 are plotted on a number line, we find that the points representing scores of batsman B_2 cluster around the central value (mean = median) while those corresponding to batsman B_1 are scattered or more spread out.

It follows from the above discussion that the central values (mean, mode, median) are not sufficient to give complete information about a distribution. Variability in the values of the observations of given data gives us better information about the data. So, variability is another factor which is required to be studied in statistics. Like central value, we have a single number to describe variability of a distribution. This single number is called the dispersion of the distribution and various methods of determining or measuring dispersion are called the measures of dispersion. In this chapter, we shall learn some of the important measures of dispersion.

32.2 MEASURES OF DISPERSION

As discussed above that the dispersion is the measure of variations in the values of the variable. It measures the degree of scatteredness of the observations in a distribution around the central value.

Following are commonly used measures of dispersion:

(i) Range (ii) Quartile deviation (iii) Mean deviation (iv) Standard deviation.

In this chapter, we shall study all of these measures of dispersion except the quartile deviation.

32.3 RANGE

RANGE *The range is the difference between two extreme observations of the distribution.*

If A and B are the greatest and smallest values respectively of observations in a distribution, then its rang is $A - B$.

Thus,

Range of a distribution = Maximum value – Minimum value

In section 32.1, we have

Range of scores of batsman $B_1 = 117 - 0 = 117$

Range of scores of batsman $B_2 = 60 - 46 = 14$.

Clearly, the range of scores of batsman B_1 is more than that of B_2. Therefore, the scores of batsman B_1 are more scattered or dispersed while the scores are more close to each other for batsman B_2.

Range is the simplest but a crude measure of dispersion. As it is based upon two extreme observations so it does not measure the dispersion of the data from a central value. Therefore, we require some other measures of variability which depend upon the difference (or deviation) of the values from the central value. Such measures of dispersion are mean deviation and standard deviation. Let us discuss them in detail.

32.4 MEAN DEVIATION

In this section, we will learn how to calculate mean deviation about mean and median for various types of data.

32.4.1 MEAN DEVIATION FOR UNGROUPED DATA OR INDIVIDUAL OBSERVATIONS

If $x_1, x_2, \ldots, x_n$ are n values of a variable X, then the mean deviation from an average A (median or AM) is given by

$$M.D. = \frac{1}{n} \sum_{i=1}^{n} |x_i - A| = \frac{1}{n} \Sigma |d_i|, \text{ where } d_i = x_i - A$$

We may use the following algorithm to find mean deviation of individual observations:

ALGORITHM

STEP I *Compute the central value or average 'A' about which mean deviation is to be calculated.*

STEP II *Take deviations of the observations about the central value 'A' obtained in step I ignoring $\pm$ signs and denote these deviations by $|d_i|$.*

STEP III *Obtain the total of these deviations i.e. $\Sigma |d_i|$.*

STEP IV *Divide the total obtained in step III by the number of observations.*

Following examples illustrate the procedure.

ILLUSTRATIVE EXAMPLES

LEVEL-1

EXAMPLE 1 *Calculate the mean deviation about median from the following data: 340, 150, 210, 240, 300, 310, 320.*

SOLUTION Arranging the observations in ascending order of magnitude, we obtain 150, 210, 240, 300, 310, 320, 340.

Clearly, the middle observation is 300. So, median = 300.

Calculation of Mean Deviation

x_i	$\lvert d_i\rvert = \lvert x_i - 300\rvert$
340	40
150	150
210	90
240	60
300	0
310	10
320	20
Total	$d_i = \Sigma \lvert x_i - 300\rvert = 370$

$$\therefore \quad M.D. = \frac{1}{n} \Sigma |d_i| = \frac{1}{7} \Sigma |x_i - 300| = \frac{370}{7} = 52.8$$

EXAMPLE 2 *The scores of a batsman in ten innings are : 38, 70, 48, 34, 42, 55, 63, 46, 54, 44. Find the mean deviation about the median.*

SOLUTION Arranging the data in ascending order, we obtain

34, 38, 42, 44, 46, 48, 54, 55, 63, 70

Here $n = 10$. So, median is the *A.M.* of 5th and 6th observations.

$\therefore$ $\quad$ Median $= \dfrac{46+48}{2} = 47$

Calculation of Mean Deviation

x_i	$\lvert d_i \rvert = \lvert x_i - 47 \rvert$
38	9
70	23
48	1
34	13
42	5
55	8
63	16
46	1
54	7
44	3
Total	$\Sigma \lvert d_i \rvert = 86$

$\therefore$ $\quad$ $M.D. = \dfrac{1}{n}\Sigma \lvert d_i \rvert = \dfrac{86}{10} = 8.6$

EXAMPLE 3 *Find the mean deviation from the mean for the data:* 6, 7, 10, 12, 13, 4, 8, 20 **[NCERT]**

SOLUTION Let $\overline{X}$ be the mean of the given data. Then,

$$\overline{X} = \frac{6+7+10+12+13+4+8+20}{8} = 10$$

Computation of Mean Deviation

x_i	$\lvert d_i \rvert = \lvert x_i - \overline{X} \rvert = \lvert x_i - 10 \rvert$
6	4
7	3
10	0
12	2
13	3
4	6
8	2
20	10
Total	$\Sigma\, d_i = 30$

Now, $\quad \Sigma \lvert d_i \rvert = 30$ and $n = 8$

$\therefore$ $\quad$ $M.D. = \dfrac{1}{n}\Sigma \lvert d_i \rvert = \dfrac{30}{8} = 3.75$

Thus, the mean deviation from the mean for the given data is 3.75.

LEVEL-2

EXAMPLE 4 *Calculate the mean deviation about the mean of the set of first n natural numbers when n is odd natural number.* **[NCERT, NCERT EXEMPLAR]**

SOLUTION Since n is an odd natural number. Therefore, $n = 2m+1$ for some natural number m. Let $\overline{X}$ be the mean of first n natural numbers. Then,

$$\overline{X} = \frac{1+2+3+\ldots+(n-1)+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\Rightarrow \quad \overline{X} = \frac{2m+1+1}{2} = m+1$$

The mean deviation (M.D.) about mean is given by

$$M.D. = \frac{1}{n} \sum_{r=1}^{n} |r - \overline{X}|$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \sum_{r=1}^{2m+1} |r-(m+1)| \qquad [\because\ n = 2m+1]$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \left\{ \sum_{r=1}^{m} |r-(m+1)| + \sum_{r=m+1}^{2m+1} |r-(m+1)| \right\}$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \left\{ \sum_{r=1}^{m} -\{r-(m+1)\} + \sum_{r=m+1}^{2m+1} \{r-(m+1)\} \right\}$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \left\{ -\sum_{r=1}^{m} r + (m+1) \sum_{r=1}^{m} 1 + \sum_{r=m+1}^{2m+1} r - (m+1) \sum_{r=m+1}^{2m+1} 1 \right\}$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \left\{ -\frac{m(m+1)}{2} + m(m+1) + \left(\frac{m+1}{2}\right) \{(m+1)+(2m+1)\} - (m+1)(m+1) \right\}$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \left\{ -\frac{m(m+1)}{2} + m(m+1) + \frac{1}{2}(m+1)(3m+2) - (m+1)^2 \right\}$$

$$\Rightarrow \quad M.D. = \frac{1}{2m+1} \left\{ \frac{m(m+1)}{2} + \frac{1}{2}(m+1)(3m+2) - (m+1)^2 \right\}$$

$$\Rightarrow \quad M.D. = \frac{m+1}{2(2m+1)} \{m + (3m+2) - 2(m+1)\}$$

$$\Rightarrow \quad M.D. = \frac{m+1}{2(2m+1)} (2m) = \frac{m(m+1)}{2m+1} = \frac{\left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}+1\right)}{n} \qquad [\because\ n = 2m+1]$$

$$\Rightarrow \quad M.D. = \frac{1}{n}\left(\frac{n-1}{2}\right)\left(\frac{n+1}{2}\right) = \frac{n^2-1}{4n}$$

EXAMPLE 5 *Calculate the mean deviation about the mean of the set of first n natural numbers when n is even natural number.* **[NCERT EXEMPLAR]**

SOLUTION Since n is an even natural number. Therefore, $n = 2m$ for some natural number m. Let $\overline{X}$ be the mean of first n natural numbers. Then,

$$\overline{X} = \frac{1+2+3+\ldots+n}{n} = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\Rightarrow \quad \overline{X} = \frac{2m+1}{2} = m + \frac{1}{2} \qquad [\because\ n = 2m]$$

The mean deviation (M.D.) about mean is given by

$$M.D. = \frac{1}{n} \sum_{r=1}^{n} |r - \overline{X}|$$

$$\Rightarrow \quad M.D. = \frac{1}{2m} \sum_{r=1}^{2m} \left| r - \left(m + \frac{1}{2}\right) \right| \qquad [\because\ n = 2m+1]$$

$$\Rightarrow \quad M.D. = \frac{1}{2m} \left[\sum_{r=1}^{m} \left| r - \left(m+\frac{1}{2}\right) \right| + \sum_{r=m+1}^{2m} \left| r - \left(m+\frac{1}{2}\right) \right| \right]$$

$$\Rightarrow \quad M.D. = \frac{1}{2m}\left[\sum_{r=1}^{m} -\left\{r-\left(m+\frac{1}{2}\right)\right\} + \sum_{r=m+1}^{2m}\left\{r-\left(m+\frac{1}{2}\right)\right\}\right]$$

$$\Rightarrow \quad M.D. = \frac{1}{2m}\left\{-\sum_{r=1}^{m} r + \sum_{r=1}^{m}\left(m+\frac{1}{2}\right) + \sum_{r=m+1}^{2m} r - \sum_{r=m+1}^{2m}\left(m+\frac{1}{2}\right)\right\}$$

$$\Rightarrow \quad M.D = \frac{1}{2m}\left\{-\frac{m(m+1)}{2} + m\left(m+\frac{1}{2}\right) + \frac{m}{2}\{(m+1)+2m\} - \left(m+\frac{1}{2}\right)m\right\}$$

$$\Rightarrow \quad M.D = \frac{1}{2m}\left\{-\frac{m(m+1)}{2} + \frac{m(2m+1)}{2} + \frac{m(3m+1)}{2} - \frac{m(2m+1)}{2}\right\}$$

$$\Rightarrow \quad M.D = \frac{1}{2m}\left\{\frac{-m(m+1)}{2} + \frac{m(3m+1)}{2}\right\} = \frac{m}{4m}(-m-1+3m+1) = \frac{m}{2} = \frac{n}{4} \qquad [\because n = 2m]$$

EXERCISE 32.1

LEVEL-1

1. Calculate the mean deviation about the median of the following observations:
   (i) 3011, 2780, 3020, 2354, 3541, 4150, 5000
   (ii) 38, 70, 48, 34, 42, 55, 63, 46, 54, 44
   (iii) 34, 66, 30, 38, 44, 50, 40, 60, 42, 51
   (iv) 22, 24, 30, 27, 29, 31, 25, 28, 41, 42
   (v) 38, 70, 48, 34, 63, 42, 55, 44, 53, 47 [NCERT]
2. Calculate the mean deviation from the mean for the following data:
   (i) 4, 7, 8, 9, 10, 12, 13, 17 [NCERT]
   (ii) 13, 17, 16, 14, 11, 13, 10, 16, 11, 18, 12, 17 [NCERT]
   (iii) 38, 70, 48, 40, 42, 55, 63, 46, 54, 44 [NCERT]
   (iv) 36, 72, 46, 42, 60, 45, 53, 46, 51, 49 [NCERT]
   (v) 57, 64, 43, 67, 49, 59, 44, 47, 61, 59
3. Calculate the mean deviation of the following income groups of five and seven members from their medians:

I *Income in ₹*	*II* *Income in ₹*
4000	3800
4200	4000
4400	4200
4600	4400
4800	4600
	4800
	5800

4. The lengths (in cm) of 10 rods in a shop are given below:
   40.0, 52.3, 55.2, 72.9, 52.8, 79.0, 32.5, 15.2, 27.9, 30.2
   (i) Find mean deviation from median
   (ii) Find mean deviation from the mean also.
5. In question 1 (iii), (iv), (v) find the number of observations lying between $\overline{X}$ – M.D. and $\overline{X}$ + M.D, where M.D. is the mean deviation from the mean.

ANSWERS

1.	(i) 649.4	(ii) 8.6	(iii) 8.7	(iv) 4.7	(v) 8.4
2.	(i) 3	(ii) 2.33	(iii) 8.4	(iv) 7.2	(v) 74

3. 320, 457.14 4. (i) 16.4 (ii) 16.44 5. 6, 5 and 6

HINTS TO NCERT & SELECTED PROBLEMS

1. (v) Arranging the observations in ascending order of magnitudes, we obtain

34, 38, 42, 44, 47, 48, 53, 55, 63, 70

These are 10 in number. Therefore,

$$\text{Median} = \text{AM of 5th and 6th observation} = \frac{47+48}{2} = 47.5$$

Calculation of Mean deviation about median

x_i	$\lvert d_i \rvert = \lvert x_i - 47.5 \rvert$
34	13.5
38	9.5
42	5.5
44	3.5
47	0.5
48	0.5
53	5.5
55	7.5
63	15.5
70	22.5
Total	$\Sigma \lvert d_i \rvert = 84$

Clearly, $n = 10$ and $\Sigma|d_i| = 84$.

$\therefore$ $$\text{Mean Deviation} = \frac{1}{n}\Sigma|d_i| = \frac{84}{10} = 8.4$$

2. (i) Let $\overline{X}$ be the mean of the given observations. Then,

$$\overline{X} = \frac{4+7+8+9+10+12+13+17}{8} = 10$$

Computation of Mean deviation about mean

x_i	$\lvert d_i \rvert = \lvert x_i - \overline{X} \rvert = \lvert x_i - 10 \rvert$
4	6
7	3
8	2
9	1
10	0
12	2
13	3
17	7
Total	$\Sigma \lvert d_i \rvert = 24$

Clearly, $\Sigma |d_i| = 24$ and $n = 8$.

$\therefore$ Mean deviation $= \frac{1}{n} \Sigma |d_i| = \frac{24}{8} = 3$

(ii) Let $\overline{X}$ be the mean of the given data. Then,

$$\overline{X} = \frac{13 + 17 + 16 + 14 + 11 + 13 + 10 + 16 + 11 + 18 + 12 + 17}{12} = \frac{168}{12} = 14$$

Computation of Mean deviation about mean

x_i	$\lvert d_i \rvert = \lvert x_i - 14 \rvert$
13	1
17	3
16	2
14	0
11	3
13	1
10	4
16	2
11	3
18	4
12	2
17	3
Total	$\Sigma \lvert d_i \rvert = 28$

Thus, we have

$n = 12$ and $\Sigma |d_i| = 28$

$\therefore$ Mean deviation $= \frac{1}{n} \Sigma |d_i| = \frac{28}{12} = 2.33$

(iii) Let $\overline{X}$ be the mean of given observations. Then,

$$\overline{X} = \frac{38 + 70 + 48 + 40 + 42 + 55 + 63 + 46 + 54 + 44}{10} = \frac{500}{10} = 50$$

Computation of Mean deviation about mean

x_i	$\lvert d_i \rvert = \lvert x_i - 50 \rvert$
38	12
70	20
48	2
40	10
42	8
55	5
63	13
46	4
54	4
44	6
Total	$\Sigma \lvert d_i \rvert = 84$

Thus, we have

$$n=10 \text{ and } \Sigma|d_i|=104$$

$$\therefore \quad \text{Mean deviation} = \frac{1}{n}\Sigma|d_i| = \frac{84}{10} = 8.4$$

(iv) Let $\overline{X}$ be the mean of given data. Then,

$$\overline{X} = \frac{36+72+46+42+60+45+53+46+51+49}{10} = \frac{500}{10} = 50$$

Computation of Mean deviation about mean

x_i	$\lvert d_i \rvert = \lvert x_i - 50 \rvert$
36	14
72	22
46	4
42	8
60	10
45	5
53	3
46	4
51	1
49	1
Total	$\Sigma\lvert d_i \rvert = 72$

Thus, we have, $n=10$ and $\Sigma|d_i|=72$.

$$\therefore \quad \text{Mean deviation} = \frac{1}{n}\ \Sigma\, d_i = \frac{72}{10} = 7.2$$

32.4.2 MEAN DEVIATION OF A DISCRETE FREQUENCY DISTRIBUTION

If x_i/f_i, $i=1, 2, \ldots, n$ is the frequency distribution, then mean deviation from an average A (median or AM) is given by

$$M.D. = \frac{1}{N}\ \Sigma f_i\,|x_i - A|, \text{ where } \sum_{i=1}^{n} f_i = N$$

We may use the following algorithm to find the mean deviation of a discrete frequency distribution.

ALGORITHM

STEP I *Calculate the central value or average 'A' of the given frequency distribution about which mean deviation is to be calculated.*

STEP II *Take deviations of the observations from the central value in step I ignoring signs and denote them by $|d_i|$.*

STEP III *Multiply these deviations by respective frequencies and obtain the total $\Sigma f_i\,|d_i|$.*

STEP IV *Divide the total obtained in step III by the number of observations i.e. $N = \Sigma f_i$ to obtain the mean deviation.*

Following examples illustrate the above algorithm.

ILLUSTRATIVE EXAMPLES

LEVEL-1

EXAMPLE 1 *Calculate mean deviation about mean from the following data:*

x_i :	3	9	17	23	27
f_i :	8	10	12	9	5

SOLUTION Calculation of mean deviation about mean.

x_i	f_i	$f_i x_i$	$\lvert x_i - 15 \rvert$	$f_i \lvert x_i - 15 \rvert$
3	8	24	12	96
9	10	90	6	60
17	12	204	2	24
23	9	207	8	72
27	5	135	12	60
	$N = \Sigma f_i = 44$	$\Sigma f_i x_i = 660$		$\Sigma f_i \lvert x_i - 15 \rvert = 312$

$$\text{Mean} = \overline{X} = \frac{1}{N}(\Sigma f_i x_i) = \frac{660}{44} = 15$$

$$\text{Mean deviation} = M.D. = \frac{1}{N} \Sigma f_i \lvert x_i - 15 \rvert = \frac{312}{44} = 7.09.$$

EXAMPLE 2 *Calculate the mean deviation from the median for the following distribution:*

x_i	10	15	20	25	30	35	40	45
f_i	7	3	8	5	6	8	4	9

SOLUTION We have to calculate mean deviation about median. So, first we calculate median.

x_i	f_i	*Cumulative frequency*	$\lvert d_i \rvert = \lvert x_i - 30 \rvert$	$f_i \lvert d_i \rvert$
10	7	7	20	140
15	3	10	15	45
20	8	18	10	80
25	5	23	5	25
30	6	29	0	0
35	8	37	5	40
40	4	41	10	40
45	9	50	15	135
	$N = \Sigma f_i = 50$			$\Sigma f_i \lvert d_i \rvert = 505$

Clearly, $N = 50 \Rightarrow N/2 = 25$.

The cumulative frequency just greater than $N/2$ is 29 and the corresponding value of x is 30.

Therefore, median = 30.

Clearly, $\Sigma f_i \lvert x_i - 30 \rvert = \Sigma f_i d_i = 505$ and $N = 50$.

$\therefore$ $$\text{Mean deviation} = \frac{1}{N} \Sigma f_i \lvert d_i \rvert = \frac{505}{50} = 10.1$$

EXERCISE 32.2

LEVEL-1

1. Calculate the mean deviation from the median of the following frequency distribution:

Heights in inches	58	59	60	61	62	63	64	65	66
No. of students	15	20	32	35	35	22	20	10	8

2. The number of telephone calls received at an exchange in 245 successive one-minute intervals are shown in the following frequency distribution:

Number of calls	0	1	2	3	4	5	6	7
Frequency	14	21	25	43	51	40	39	12

Compute the mean deviation about median.

3. Calculate the mean deviation about the median of the following frequency distribution:

x_i	5	7	9	11	13	15	17
f_i	2	4	6	8	10	12	8

4. Find the mean deviation from the mean for the following data:

(i)

x_i	5	7	9	10	12	15
f_i	8	6	2	2	2	6

[NCERT]

(ii)

x_i	5	10	15	20	25
f_i	7	4	6	3	5

[NCERT]

(iii)

x_i	10	30	50	70	90
f_i	4	24	28	16	8

[NCERT]

(iv)

Size:	20	21	22	23	24
Frequency:	6	4	5	1	4

[NCERT EXEMPLAR]

(v)

Size:	1	3	5	7	9	11	13	15
Frequency:	3	3	4	14	7	4	3	4

[NCERT EXEMPLAR]

5. Find the mean deviation from the median for the following data:

(i)

x_i	15	21	27	30
f_i	3	5	6	7

[NCERT]

(ii)

x_i	74	89	42	54	91	94	35
f_i	20	12	2	4	5	3	4

(iii)

Mark obtained	10	11	12	14	15
No. of students	2	3	8	3	4

[NCERT EXEMPLAR]

ANSWERS

1. 1.703	2. 1.49	3. 2.72		
4. (i) 3.38	(ii) 6.32	(iii) 15.3	(iv) 0.32	(v) 2.95
5. (i) 5.93	(ii) 12.5	(iii) 1.25		

HINTS TO NCERT & SELECTED PROBLEMS

4. (i) Computation of mean deviation about mean

x_i	f_i	$f_i\, x_i$	$\lvert x_i - 9\rvert$	$\Sigma f_i \lvert x_i - 9\rvert$
5	8	40	4	32
7	6	42	2	12
9	2	18	0	0
10	2	20	1	2
12	2	24	3	6
15	6	90	6	36
	$N = \Sigma f_i = 26$	$\Sigma f_i\, x_i = 234$		$\Sigma f_i \lvert x_i - 9\rvert = 88$

We have, $N = \Sigma f_i = 26$, and $\Sigma f_i\, x_i = 234$

$$\therefore \quad \text{Mean} = \overline{X} = \frac{1}{N} \Sigma f_i\, x_i = \frac{234}{26} = 9$$

$$\text{Mean deviation} = \frac{1}{N} \Sigma f_i \lvert x_i - 9\rvert = \frac{88}{26} = 3.38$$

4. (ii) Computation of Mean deviation about mean

x_i	f_i	$f_i\, x_i$	$\lvert x_i - 14\rvert$	$f_i \lvert x_i - 14\rvert$
5	7	35	9	63
10	4	40	4	16
15	6	90	1	6
20	3	60	6	18
25	5	125	11	55
	$N = \Sigma f_i = 25$	$\Sigma f_i\, x_i = 350$		$\Sigma f_i \lvert x_i - 14\rvert = 158$

$$\overline{X} = \frac{1}{N} \Sigma f_i\, x_i = \frac{350}{25} = 14$$

$$\text{Mean deviation} = \frac{1}{N} \Sigma f_i \lvert x_i - 14\rvert = \frac{158}{25} = 6.32$$

(iii) Computation of mean deviation about mean

x_i	f_i	$f_i\, x_i$	$\lvert x_i - 49\rvert$	$f_i \lvert x_i - 49\rvert$
10	4	40	39	156
30	24	720	19	456
50	28	1400	1	28
70	16	1120	21	336
80	8	640	31	248
	$N = \Sigma f_i = 80$	$\Sigma f_i\, x_i = 3920$		$\Sigma f_i \lvert x_i - 49\rvert = 1224$

$\therefore \quad \overline{X} = \frac{1}{N} \Sigma f_i x_i = \frac{3920}{80} = 49$

and, $\quad$ Mean deviation $= \frac{1}{N} \Sigma f_i |x_i - 49| = \frac{1224}{80} = 15.3$

5. (i) Computation of mean deviation from median

x_i	f_i	$c.f.$	$\|x_i - 30\|$	$f_i \|x_i - 30\|$
15	3	3	15	45
21	5	8	9	45
27	6	14	7	42
30	7	21	0	0
35	8	29	5	40
		$N = \Sigma f_i = 29$		$\Sigma f_i \|x_i - 30\| = 172$

We have, $N = 29 \Rightarrow \frac{N}{2} = 14.5$

The cumulative frequency just greater than $\frac{N}{2}$ i.e. 14.5 is 21. The corresponding value of the variable is 30. So, Median = 30.

Mean deviation $= \frac{1}{N} \Sigma f_i |x_i - 30| = \frac{172}{29} = 5.93$

32.4.3 MEAN DEVIATION OF A GROUPED OR CONTINUOUS FREQUENCY DISTRIBUTION

For calculating mean deviation of a continuous frequency distribution the procedure is same as for a discrete frequency distribution. The only difference is that here we have to obtain the mid-points of the various classes and take the deviations of the these mid-points from the given central value (median or mean).

Following examples will illustrate the procedure.

ILLUSTRATIVE EXAMPLES

LEVEL-1

EXAMPLE 1 *Find the mean deviation about the median of the following frequency distribution:*

Class:	0-6	6-12	12-18	18-24	24-30
Frequency:	8	10	12	9	5

[NCERT EXEMPLAR]

SOLUTION *Calculation of Mean Deviation about the Median*

Class	Mid-Values (x_i)	Frequency (f_i)	Cumulative Frequency (c.f.)	$\|x_i - 14\|$	$f_i \|x_i - 14\|$
0-6	3	8	8	11	88
6-12	9	10	18	5	50
12-18	15	12	30	1	12
18-24	21	9	39	7	63
24-30	27	5	44	13	65
		$N = \Sigma f_i = 44$			$\Sigma f_i \|x_i - 14\| = 278$

Here $N = 44$, so $\frac{N}{2} = 22$ and the cumulative frequency just greater than $\frac{N}{2}$ is 30. Thus 12-18 is the median class.

$\therefore$ Median $= l + \frac{N/2 - F}{f} \times h$, where $l = 12, h = 6, f = 12, F = 18$.

$\Rightarrow$ Median $= 12 + \frac{22 - 18}{12} \times 6 = 12 + \frac{4 \times 6}{12} = 14$.

Clearly, $\Sigma f_i |x_i - 14| = 278$

$\therefore$ Mean deviation about median $= \frac{1}{N} \Sigma f_i |x_i - 14| = \frac{278}{44} = 6.318$

EXAMPLE 2 *Calculate the mean deviation from the median of the following data:*

Wages per week (in ₹)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of workers	4	6	10	20	10	6	4

SOLUTION *Calculation of Mean Deviation from Median*

Wages per week (in ₹)	*Mid-Values* (x_i)	*Frequency* f_i	*Cumulative frequency*	$\|d_i\| = \|x_i - 45\|$	$f_i \|d_i\|$
10-20	15	4	4	30	120
20-30	25	6	10	20	120
30-40	35	10	20	10	100
40-50	45	20	40	0	0
50-60	55	10	50	10	100
60-70	65	6	56	20	120
70-80	75	4	60	30	120
		$N = \Sigma f_i = 60$			$\Sigma f_i \|d_i\| = 680$

Here, $N = 60$. So, $\frac{N}{2} = 30$. The cumulative frequency just greater than $\frac{N}{2} = 30$ is 40 and the corresponding class is 40-50. So, 40-50 is the median class.

$\therefore$ $l = 40, \quad f = 20, \quad h = 10, \quad F = 20$.

So, Median $= l + \frac{N/2 - F}{f} \times h = 40 + \frac{30 - 20}{20} \times 10 = 45$.

Thus, we have

$$\Sigma f_i |x_i - 45| = \Sigma f_i |d_i| = 680 \text{ and } N = 60.$$

$\therefore$ Mean deviation from median $= \frac{\Sigma f_i |d_i|}{N} = \frac{680}{60} = 11.33$.

EXAMPLE 3 *Find the mean deviation from the mean for the following data:*

Classes :	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequencies :	2	3	8	14	8	3	2

SOLUTION *Computation of Mean Deviation from Mean*

Classes	Mid-values x_i	frequencies f_i	$f_i\,x_i$	$\lvert x_i - \overline{X}\rvert = \lvert x_i - 45\rvert$	$f_i\lvert x_i - \overline{X}\rvert$
10-20	15	2	30	30	60
20-30	25	3	75	20	60
30-40	35	8	280	10	80
40-50	45	14	630	0	0
50-60	55	8	440	10	80
60-70	65	3	195	20	60
70-80	75	2	150	30	60
		$N = \Sigma f_i = 40$	$\Sigma f_i\,x_i = 1800$		$\Sigma f_i\lvert x_i - \overline{X}\rvert = 400$

Clearly, $N = 40$ and $\Sigma f_i\,x_i = 1800$

$$\therefore \quad \overline{X} = \frac{\Sigma f_i\,x_i}{N} = \frac{1800}{40} = 45$$

From the above table, we get

$$\Sigma f_i \lvert x_i - \overline{X}\rvert = 400 \text{ and } N = \Sigma f_i = 40$$

$$\therefore \quad M.D. = \frac{1}{N}\Sigma f_i \lvert x_i - \overline{X}\rvert = \frac{400}{40} = 10.$$

EXAMPLE 4 *Find the mean deviation about the mean for the following data:*

Marks obtained:	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Number of students:	2	3	8	14	8	3	2

[NCERT]

SOLUTION In order to avoid the tedious calculations of computing mean ($\overline{X}$), let us compute $\overline{X}$ by step-deviation method. The formula for the same is

$$\overline{X} = a + h\left(\frac{1}{N}\sum_{i=1}^{n} f_i\,d_i\right), \text{ where } d_i = \frac{x_i - a}{h}, \; a = \text{assumed mean and, } h = \text{common factor.}$$

Let us take the assumed mean $a = 45$ and $h = 10$ and form the following table:

Marks obtained	*Number of students* f_i	*Mid-points* x_i	$d_i = \frac{x_i - 45}{10}$	$f_i\,d_i$	$\lvert x_i - \overline{X}\rvert = \lvert x_i - 45\rvert$	$f_i\lvert x_i - \overline{X}\rvert$
10-20	2	15	-3	-6	30	60
20-30	3	25	-2	-6	20	60
30-40	8	35	-1	-8	10	80
40-50	14	45	0	0	0	0
50-60	8	55	1	8	10	80
60-70	3	65	2	6	20	60
70-80	2	75	3	6	30	60
	$N = 40$			$\Sigma f_i\,d_i = 0$		$\Sigma f_i\lvert x_i - \overline{X}\rvert = 400$

Clearly, $N = 40$, $\Sigma f_i\,d_i = 0$.

$$\therefore \quad \overline{X} = a + h\left(\frac{1}{N}\Sigma f_i d_i\right) = 45 + 10 \times \frac{0}{40} = 45.$$

It is evident from the table that $\Sigma f_i |x_i - \overline{X}| = 400$

$$\therefore \quad \text{M.D.} = \frac{1}{N} \Sigma f_i |x_i - \overline{X}| = \frac{400}{40} = 10.$$

EXERCISE 32.3

LEVEL-1

1. Compute the mean deviation from the median of the following distribution:

Class	0-10	10-20	20-30	30-40	40-50
Frequency	5	10	20	5	10

2. Find the mean deviation from the mean for the following data :

(i)

Classes	0-100	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Frequencies	4	8	9	10	7	5	4	3

[NCERT]

(ii)

Classes	95-105	105-115	115-125	125-135	135-145	145-155
Frequencies	9	13	16	26	30	12

[NCERT]

(iii)

Classes	0-10	10-20	20-30	30-40	40-50	50-60
Frequencies	6	8	14	16	4	2

[NCERT]

3. Compute mean deviation from mean of the following distribution:

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of students	8	10	15	25	20	18	9	5

4. The age distribution of 100 life-insuance policy holders is as follows:

Age (on nearest birth day)	17-19.5	20-25.5	26-35.5	36-40.5	41-50.5	51-55.5	56-60.5	61-70.5
No. of persons	5	16	12	26	14	12	6	5

Calculate the mean deviation from the median age.

5. Find the mean deviation from the mean and from median of the following distribution:

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

6. Calculate mean deviation about median age for the age distribution of 100 persons given below:

Age:	16-20	21-25	26-30	31-35	36-40	41-45	46-50	51-55
Number of persons	5	6	12	14	26	12	16	9

7. Calculate the mean deviation about the mean for the following frequency distribution:

Class interval:	0-4	4-8	8-12	12-16	16-20
Frequency:	4	6	8	5	2

8. Calculate mean deviation from the median of the following data: [NCERT EXEMPLAR]

Class interval:	0-6	6-12	12-18	18-24	24-30
Frequency:	4	5	3	6	2

ANSWERS

1. 9 2. (i) 157.92 (ii) 12.005 (iii) 10.576 3. 14.218
4. 10.605 5. 9.44, 9.56 7. 0.99 8. 7.08

HINTS TO NCERT & SELECTED PROBLEMS

2. (i) Computation of mean deviation from the mean

Classes	f_i	Mid-points x_i	$d_i = \frac{x_i - 450}{100}$	$f_i\, d_i$	$\lvert x_i - \overline{X} \rvert = \lvert x_i - 358 \rvert$	$f_i \lvert x_i - \overline{X} \rvert$
0 – 100	4	50	– 4	– 16	308	1232
100 – 200	8	150	– 3	– 24	208	1664
200 – 300	9	250	– 2	– 18	108	972
300 – 400	10	350	– 1	– 10	8	80
400 – 500	7	450	0	0	92	644
500 – 600	5	550	1	5	192	960
600 – 700	4	650	2	8	292	1168
700 – 800	3	750	3	9	392	1176
	$\Sigma f_i = 50$			$\Sigma f_i d_i = -46$		$\Sigma f_i \lvert x_i - \overline{X} \rvert = 7896$

Thus, we have $a = 450$, $h = 100$, $N = 50$, $\Sigma f_i\, d_i = -46$ and $\Sigma f_i \lvert x_i - \overline{X} \rvert = 7896$

$$\therefore \quad \overline{X} = a + h\left(\frac{1}{N}\Sigma f_i\, d_i\right) = 450 + 100 \times \left(-\frac{46}{50}\right) = 358$$

and, $$\text{Mean deviation} = \frac{1}{N}\Sigma f_i \lvert x_i - \overline{X} \rvert = \frac{7896}{50} = 157.92$$

(ii) Computation of mean deviation about mean

Classes	f_i	Mid-values x_i	$d_i = \frac{x_i - 130}{10}$	$f_i d_i$	$\lvert x_i - \overline{X} \rvert$	$f_i \lvert x_i - \overline{X} \rvert$
95 – 105	9	100	– 3	– 27	28.58	257.22
105 – 115	13	110	– 2	– 26	18.58	241.54
115 – 125	16	120	– 1	– 16	8.58	137.28
125 – 135	26	130	0	0	1.42	36.92
135 – 145	30	140	1	30	11.42	342.6
145 – 155	12	150	2	24	21.42	257.04
	$N = \Sigma f_i = 106$			$\Sigma f_i d_i = -15$		$\Sigma f_i \lvert x_i - \overline{X} \rvert = 1272.60$

Clearly, $N = \Sigma f_i = 106,\ a = 130,\ h = 10$ and Σf_i

$$\therefore \quad \overline{X} = a + h\left(\frac{1}{N} \Sigma f_i d_i\right) = 130 + 10 \times \frac{-15}{106} = 128.58$$

Also, $\Sigma f_i \lvert x_i - \overline{X} \rvert = 1272.60$ and $N = 106$

$$\therefore \quad \text{Mean deviation} = \frac{1}{N} \Sigma f_i \lvert x_i - \overline{X} \rvert = \frac{1272.60}{106} = 12.005$$

(iii) Computation of mean deviation about mean

Classes	f_i	Mid-values x_i	$d_i = \frac{x_i - 25}{10}$	$f_i d_i$	$\lvert x_i - \overline{X} \rvert = \lvert x_i - 29.8 \rvert$	$f_i \lvert x_i - \overline{X} \rvert$
0 – 10	6	5	– 2	– 12	24.8	148.8
10 – 20	8	15	– 1	– 8	14.8	118.4
20 – 30	14	25	0	14	4.8	67.2
30 – 40	16	35	1	16	5.2	83.2
40 – 50	4	45	2	8	15.2	60.8
50 – 60	2	55	3	6	25.2	50.4
	$N = 50$			$\Sigma f_i d_i = 24$		$\Sigma f_i \lvert x_i - \overline{X} \rvert = 528.8$

Clearly, $N = 50, a = 25, h = 10$ and $\Sigma f_i d_i = 24$.

$$\therefore \quad \overline{X} = a + h\left(\frac{\Sigma f_i d_i}{N}\right) = 25 + \frac{24}{50} \times 10 = 29.8$$

Also, $\Sigma f_i \lvert x_i - \overline{X} \rvert = 528.8$ and $N = 50$

$$\therefore \quad \text{Mean deviation} = \frac{1}{N} \sum f_i \lvert x_i - \overline{X} \rvert = \frac{528.8}{50} = 10.576$$

6. Convert the given data into continous frequency distribution by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit of each class interval as given below:

Age	15.5-20.5	20.5-25.5	25.5-30.5	30.5-35.5	35.5-40.5	40.5-45.5	45.5-50.5	50.5-55.5
Number of persons:	5	6	12	14	26	12	16	9

32.4.4 LIMITATIONS OF MEAN DEVIATION

Following are some limitations or demerits of mean deviation.

(i) In a frequency distribution the sum of absolute values of deviations from the mean is always more than the sum of the deviations from median. Therefore, mean deviation about mean is not very scientific. Thus, in many cases, mean deviation may give unsatisfactory results.

(ii) In a distribution, where the degree of variability is very high, the median is not a representative central value. Thus, the mean deviation about median calculated for such series can not be fully relied.

(iii) In the computation of mean deviation we use absolute values of deviations. Therefore, it cannot be subjected to further algebraic treatment.

32.5 VARIANCE AND STANDARD DEVIATON

VARIANCE *The variance of a variate X is the arithmetic mean of the squares of all deviations of X from the arithmetic mean of the observations and is denoted by Var(X) or σ^2.*

The positive square root of the variance of a variate X is known as its standard deviation and is denoted by σ.

Thus, Standard deviation $= +\sqrt{\text{Var}(X)}$

Similar to the mean deviation, we shall discuss the calculation of variance and standard deviation in the following three cases:

(i) Individual observations (ii) Discrete frequency distribution

(iii) Continuous or grouped frequency distribution.

32.5.1 VARIANCE OF INDIVIDUAL OBSERVATIONS

If $x_1, x_2, \dots, x_n$ are n values of a variable X, then

$$\text{Var}(X) = \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \overline{X})^2\right\} \text{ or, } \sigma^2 = \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \overline{X})^2\right\} \quad \text{...(i)}$$

Now,
$$\text{Var}(X) = \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i - \overline{X})^2\right\} = \frac{1}{n}\left\{\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{X} + \overline{X}^2)\right\}$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \frac{1}{n}\sum_{i=1}^{n}2x_i\overline{X} + \frac{1}{n}\sum_{i=1}^{n}\overline{X}^2$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\overline{X}\left\{\frac{1}{n}\sum_{i=1}^{n}x_i\right\} + \frac{n\overline{X}^2}{n}$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2\overline{X}^2 + \overline{X}^2 \qquad \left[\because \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{X}\right]$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \overline{X}^2$$

$$\therefore \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \left\{\frac{1}{n}\sum_{i=1}^{n}x_i\right\}^2 \quad \text{...(ii)}$$

If the values of variable X are large, the calculation of variance from the above formulae is quite tedious and time consuming. In that case we take deviations from an arbitrary point A (say). If $d_i = x_i - A_i$ $i = 1, 2, \dots, n$, then

$$\sum_{i=1}^{n}d_i = \sum_{i=1}^{n}(x_i - A) = \sum_{i=1}^{n}x_i - nA$$

$$\Rightarrow \quad \frac{1}{n}\left\{\sum_{i=1}^{n} d_i\right\} = \frac{1}{n}\sum_{i=1}^{n} x_i - A$$

$$\Rightarrow \quad \frac{1}{n}\left\{\sum_{i=1}^{n} d_i\right\} = \overline{X} - A$$

$$\Rightarrow \quad \overline{d} = \overline{X} - A, \text{ where } \overline{d} = \frac{1}{N}\sum_{i=1}^{n} d_i$$

$$\therefore \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{X})^2$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - A + A - \overline{X})^2$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(d_i - \overline{d}\,)^2$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(d_i^2 - 2\,d_i\,\overline{d} + \overline{d}^2) = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \frac{1}{n}\sum_{i=1}^{n} 2d_i\,\overline{d} + \frac{1}{n}\sum_{i=1}^{n}\overline{d}^2$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - 2\left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)\overline{d} + \frac{n\,\overline{d}^2}{n}$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - 2\,\overline{d}^2 + \overline{d}^2$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \overline{d}^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^2$$

$$\text{Thus,} \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} di^2 - \left(\frac{1}{n}\sum_{i=1}^{n} di\right)^2 \qquad \text{...(iii)}$$

It follows from the above discussion that in case of individual observations, variance and standard deviation may be computed by applying any of the above three formulas. Following algorithm is useful for finding the variance when deviations are taken from the actual mean.

ALGORITHM

STEP I *Compute the mean $\overline{X}$ of the given observations $x_1, x_2, \dots, x_n$.*

STEP II *Take the deviations of the observations from the mean i.e. find $x_i - \overline{X}$; $i = 1, 2, \dots, n$.*

STEP III *Square the deviations obtained in step II and obtain the sum $\sum_{i=1}^{n}(x_i - \overline{X})^2$.*

STEP IV *Divide the sum $\sum_{i=1}^{n}(x_i - \overline{X})^2$ obtained in step III by n. This gives the value of variance of X.*

Following examples will illustrate the above algorithm.

ILLUSTRATIVE EXAMPLES

LEVEL-1

Type I **ON FINDING VARIANCE AND STANDARD DEVIATION OF INDIVIDUAL OBSERVATIONS**

EXAMPLE 1 *Compute the variance and standard deviation of the following observations of marks of 5 students of a tutorial group:*

Marks out of 25 : 8, 12, 13, 15, 22

SOLUTION Clearly,

$$\overline{X} = \frac{8+12+13+15+22}{5} = 14$$

Calculation of variance

x_i	$x_i - \overline{X}$	$(x_i - \overline{X})^2$
8	−6	36
12	−2	4
13	−1	1
15	1	1
22	8	64
		$\Sigma(x_i - \overline{X})^2 = 106$

Here, $n = 5$ and $\Sigma(x_i - \overline{X})^2 = 106$

$\therefore$ $\quad \text{Var}(X) = \frac{1}{n}\Sigma(x_i - \overline{X})^2 = \frac{106}{5} = 21.2$

$\Rightarrow$ $\quad \text{S.D.} = \sqrt{\text{Var}(X)} = \sqrt{21.2} = 4.604.$

EXAMPLE 2 *Find the variance and standard deviation for the following data:*

65, 68, 58, 44, 48, 45, 60, 62, 60, 50

SOLUTION Let $\overline{X}$ be the mean of the given set of observations. Then,

$$\overline{X} = \frac{65+68+58+44+48+45+60+62+60+50}{10} = \frac{560}{10} = 56$$

Computation of Variance

x_i	$x_i - \overline{X} = x_i - 56$	$(x_i - \overline{X})^2$
65	9	81
58	2	4
68	12	144
44	−12	144
48	−8	64
45	−11	121
60	4	16
62	6	36
60	4	16
50	−6	36
		$\Sigma(x_i - \overline{X})^2 = 662$

Clearly, $n = 10$ and $\Sigma(x_i - \overline{X})^2 = 662.$

$\therefore$ $\quad \text{Variance} = \frac{1}{n}\Sigma(x_i - \overline{X})^2 = \frac{662}{10} = 66.2$

Hence, Standard deviation $(\sigma) = \sqrt{\text{Variance}} = \sqrt{66.2} = 8.13$

Type II ON PROVING RESULTS ON VARIANCE

EXAMPLE 3 *Let $x_1, x_2, x_3, \dots, x_n$ be n values of a variable X. If these values are changed to $x_1 + a, x_2 + a, \dots, x_n + a$, where $a \in R$, show that the variance remains unchanged.* [NCERT]

SOLUTION Let $u_i = x_i + a$, $i = 1, 2, \dots, n$ be the n values of variable U. Then,

$$\overline{U} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n}(x_i + a) = \frac{1}{n}\left\{\sum_{i=1}^{n} x_i + na\right\} = \frac{1}{n}\sum_{i=1}^{n} x_i + a = \overline{X} + a$$

$\therefore \quad u_i - \overline{U} = (x_i + a) - (\overline{X} + a) = x_i - \overline{X},\ i = 1, 2, ..., n.$

$\Rightarrow \quad \sum_{i=1}^{n} (u_i - \overline{U})^2 = \sum_{i=1}^{n} (x_i - \overline{X})^2$

$\Rightarrow \quad \frac{1}{n} \sum_{i=1}^{n} (u_i - \overline{U})^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})^2$

$\Rightarrow \quad \text{Var}(U) = \text{Var}(X).$

EXAMPLE 4 *Let $x_1, x_2, ..., x_n$ values of a variable X and let 'a' be a non-zero real number. Then, prove that the variance of the observations $ax_1, ax_2, ..., ax_n$ is a^2 Var (X). Also, find their standard deviation.* **[NCERT]**

SOLUTION Let $u_1, u_2, ..., u_n$ be the n values of variable U such that $u_i = ax_i$, $i = 1, 2, ..., n$. Then,

$$\overline{U} = \frac{1}{n} \sum_{i=1}^{n} u_i = \frac{1}{n} \sum_{i=1}^{n} (ax_i) = a \left\{ \frac{1}{n} \sum_{i=1}^{n} x_i \right\} = a\overline{X}$$

$\therefore \quad u_i - \overline{U} = ax_i - a\overline{X}$ for all $i = 1, 2, ..., n$

$\Rightarrow \quad u_i - \overline{U} = a(x_i - \overline{X})$

$\Rightarrow \quad (u_i - \overline{U})^2 = a^2 (x_i - \overline{X})^2$

$\Rightarrow \quad \sum_{i=1}^{n} (u_i - \overline{U})^2 = a^2 \sum_{i=1}^{n} (x_i - \overline{X})^2$

$\Rightarrow \quad \frac{1}{n} \sum_{i=1}^{n} (u_i - \overline{U})^2 = a^2 \left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{X})^2 \right\}$

$\Rightarrow \quad \text{Var}(U) = a^2\ \text{Var}(X)$

$\therefore \quad \sigma_U = \sqrt{\text{Var}(U)} = \sqrt{a^2\ \text{Var}(X)} = |a| \sqrt{\text{Var}(X)} = |a|\ \sigma_X.$

<u>REMARK</u> *The variance of 20 observations is 5. If each observation is multiplied by 2, then from the above example,*
New variance of the resulting observations $= 2^2 \times 5 = 20$

EXAMPLE 5 *Let $x_1, x_2, x_3, ..., x_n$ be n values of a variable X, and let $x_i = a + hu_i$, $i = 1, 2, ..., n$, where $u_1, u_2, ..., u_n$ are the values of variable U. Then, prove that* Var (X) $= h^2$ Var (U), $h \neq 0$.

SOLUTION We have,

$x_i = a + h u_i,\ i = 1, 2, ..., n$

$\Rightarrow \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} (a + h u_i)$

$\Rightarrow \quad \sum_{i=1}^{n} x_i = na + h \sum_{i=1}^{n} u_i$

$\Rightarrow \quad \frac{1}{n} \sum_{i=1}^{n} x_i = a + h \left(\frac{1}{n} \sum_{i=1}^{n} u_i \right)$

$\Rightarrow \quad \overline{X} = a + h\overline{u} \qquad \left[\because \overline{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \overline{U} = \frac{1}{n} \sum_{i=1}^{n} u_i \right]$

$\therefore \quad x_i - \overline{X} = (a + h u_i) - (a + h\overline{U}),\ i = 1, 2, ..., n$

$\Rightarrow \quad x_i - \overline{X} = h(u_i - \overline{U}),\ i = 1, 2, ..., n$

$$\Rightarrow \quad (x_i - \overline{X})^2 = h^2 (u_i - \overline{U})^2, i = 1, 2, \ldots, n$$

$$\Rightarrow \quad \sum_{i=1}^{n} (x_i - \overline{X})^2 = h^2 \sum_{i=1}^{n} (u_i - \overline{U})^2$$

$$\Rightarrow \quad \frac{1}{n} \Sigma (x_i - \overline{X})^2 = h^2 \left\{ \frac{1}{n} \sum_{i=1}^{n} (u_i - \overline{U})^2 \right\} \qquad \text{[Dividing both sides by } n\text{]}$$

$$\Rightarrow \quad \text{Var}(X) = h^2 \text{ Var}(U).$$

Type III **ON FINDING THE DESIRED VALUES BY USING THE FORMULAS FOR MEAN AND VARIANCE OF INDIVIDUAL OBSERVATIONS**

EXAMPLE 6 *If the mean and standard deviation of 100 observations are 50 and 4 respectively. Find the sum of all the observations and the sum of their squares.* **[NCERT EXEMPLAR]**

SOLUTION Let $x_1, x_2, \ldots, x_{100}$ be 100 observations and their mean and standard deviation be $\overline{X}$ and σ respectively. Then,

$$\overline{X} = \frac{1}{100} \sum_{i=1}^{100} x_i \text{ and, } \sigma^2 = \frac{1}{100} \sum_{i=1}^{100} x_i^2 - \overline{X}^2$$

$$\Rightarrow \quad 50 = \frac{1}{100} \sum_{i=1}^{100} x_i \text{ and, } 16 = \frac{1}{100} \sum_{i=1}^{100} x_i^2 - (50)^2 \qquad [\because \overline{X} = 50 \text{ and } \sigma = 4]$$

$$\Rightarrow \quad 5000 = \sum_{i=1}^{100} x_i \text{ and, } 1600 = \sum_{i=1}^{100} x_i^2 - 250000$$

$$\Rightarrow \quad \sum_{i=1}^{100} x_i = 5000 \text{ and, } \sum_{i=1}^{100} x_i^2 = 251600$$

EXAMPLE 7 *If for a distribution of 18 observations $\Sigma(x_i - 5) = 3$ and $\Sigma(x_i - 5)^2 = 43$, find the mean and standard deviation.* **[NCERT EXEMPLAR]**

SOLUTION We have

$$\sum_{i=1}^{18} (x_i - 5) = 3 \text{ and, } \sum_{i=1}^{18} (x_i - 5)^2 = 43$$

$$\Rightarrow \quad \sum_{i=1}^{18} x_i - \sum_{i=1}^{18} 5 = 3 \text{ and, } \sum_{i=1}^{18} x_i^2 - 10 \sum_{i=1}^{18} x_i + \sum_{i=1}^{18} 25 = 43$$

$$\Rightarrow \quad \sum_{i=1}^{18} x_i - 18 \times 5 = 3 \text{ and, } \sum_{i=1}^{18} x_i^2 - 10 \sum_{i=1}^{18} x_i + 18 \times 25 = 43$$

$$\Rightarrow \quad \sum_{i=1}^{18} x_i = 93 \text{ and, } \sum_{i=1}^{18} x_i^2 - 10 \times 93 + 18 \times 25 = 43$$

$$\Rightarrow \quad \sum_{i=1}^{18} x_i = 93 \text{ and, } \sum_{i=1}^{18} x_i^2 = 523$$

$$\therefore \quad \text{Mean} = \frac{1}{18} \sum_{i=1}^{18} x_i = \frac{93}{18} = 5.17$$

$$\text{S.D.} = \sqrt{\frac{1}{18} \sum_{i=1}^{18} x_i^2 - \left(\frac{1}{18} \sum_{i=1}^{18} x_i \right)^2} = \sqrt{\frac{523}{18} - \left(\frac{93}{18} \right)^2} = \sqrt{\frac{9414 - 8649}{324}} = \frac{\sqrt{765}}{18} = \frac{27.6586}{18} = 1.536$$

Type IV **ON FINDING CORRECTED MEAN AND CORRECTED VARIANCE OR S.D.**

EXAMPLE 8 *For a group of 200 candidates the mean and S.D. were found to be 40 and 15 respectively. Later on it was found that the score 43 was misread as 34. Find the correct mean and correct S.D.*

SOLUTION We have, $n = 200$, $\overline{X} = 40$, $\sigma = 15$.

$\therefore \quad \overline{X} = \frac{1}{n} \Sigma x_i \Rightarrow \Sigma x_i = n\overline{X} = 200 \times 40 = 8000.$

Now, Corrected Σx_i = Incorrect Σx_i – (Sum of incorrect values) + (Sum of correct values)
$= 8000 - 34 + 43 = 8009.$

$\therefore$ Corrected mean $= \dfrac{\text{Corrected } \Sigma x_i}{n} = \dfrac{8009}{200} = 40.045$

and, $\sigma = 15$

$\Rightarrow \quad 15^2 = \text{Variance}$

$\Rightarrow \quad 15^2 = \frac{1}{200}(\Sigma x_i^2) - \left(\frac{1}{200}\Sigma x_i\right)^2$

$\Rightarrow \quad 225 = \frac{1}{200}(\Sigma x_i^2) - \left(\frac{8000}{200}\right)^2$

$\Rightarrow \quad 225 = \frac{1}{200}(\Sigma x_i^2) - 1600$

$\Rightarrow \quad \Sigma x_i^2 = 200 \times 1825 = 365000$

$\Rightarrow$ Incorrect $\Sigma x_i^2 = 365000$.

$\therefore$ Corrected Σx_i^2 = (Incorrect Σx_i^2) – (Sum of squares of incorrect values)
+ (Sum of squares of correct values)

$\Rightarrow$ Corrected $\Sigma x_i^2 = 365000 - (34)^2 + (43)^2 = 365693$

So, $\text{Corrected } \sigma = \sqrt{\frac{1}{n}\text{Corrected } \Sigma x_i^2 - \left(\frac{1}{n}\text{Corrected } \Sigma x_i\right)^2} = \sqrt{\frac{365693}{200} - \left(\frac{8009}{200}\right)^2}$

$= \sqrt{1828.465 - 1603.602} = 14.995.$

EXAMPLE 9 *The mean and standard deviation of 20 observations are found to be 10 and 2 respectively. On rechecking, it was found that an observation 8 was incorrect. Calculate the correct mean and standard deviation in each of the following cases:* (i) *If the wrong item is omitted.* *(ii) If it is replaced by 12.*

SOLUTION We have, $n = 20$, $\overline{X} = 10$ and $\sigma = 2$.

$\therefore \quad \overline{X} = \frac{1}{n}\Sigma x_i$

$\Rightarrow \quad \Sigma x_i = n\overline{X} = 20 \times 10 = 200$

$\Rightarrow$ Incorrected $\Sigma x_i = 200$.

and, $\sigma = 2$

$\Rightarrow \quad \sigma^2 = 4$

$\Rightarrow \quad \frac{1}{n}\Sigma x_i^2 - (\text{Mean})^2 = 4$

$\Rightarrow \quad \frac{1}{20}\Sigma x_i^2 - 100 = 4$ [$\because$ Mean = 10]

$\Rightarrow \quad \Sigma x_i^2 = 104 \times 20$

$\Rightarrow$ Incorrected $\Sigma x_i^2 = 2080$.

(i) *When 8 is omitted from the data:* If 8 is omitted from the data, then 19 observations are left.

Now, Incorrected $\Sigma x_i = 200$

$\Rightarrow$ Corrected $\Sigma x_i + 8 = 200$

$\Rightarrow$ Corrected $\Sigma x_i = 192$

and,

Incorrected $\Sigma x_i^2 = 2080$

$\Rightarrow$ Corrected $\Sigma x_i^2 + 8^2 = 2080$

$\Rightarrow$ Corrected $\Sigma x_i^2 = 2080 - 64$

$\Rightarrow$ Corrected $\Sigma x_i^2 = 2016$

$\therefore$ Corrected mean $= \dfrac{\text{Corrected } \Sigma x_i}{19} = \dfrac{192}{19} = 10.10$

$\Rightarrow$ Corrected variance $= \dfrac{1}{19}$ (Corrected Σx_i^2) $-$ (Corrected mean)^2

$\Rightarrow$ Corrected variance $= \dfrac{2016}{19} - \left(\dfrac{192}{19}\right)^2 = \dfrac{38304 - 36864}{361} = \dfrac{1440}{361}$

$\therefore$ Corrected standard deviation $= \sqrt{\dfrac{1440}{361}} = \dfrac{12\sqrt{10}}{19} = 1.997.$

(ii) *When the incorrect observation 8 is replaced by 12:*

Now, Incorrected $\Sigma x_i = 200$

$\Rightarrow$ Corrected $\Sigma x_i - 12 + 8 = 200$

$\Rightarrow$ Corrected $\Sigma x_i = 200 - 8 + 12 = 204$

and, Incorrected $\Sigma x_i^2 = 2080$

$\therefore$ Corrected $\Sigma x_i^2 = 2080 - 8^2 + 12^2 = 2160.$

Now, Corrected mean $= \dfrac{204}{20} = 10.2$

Corrected Variance $= \dfrac{1}{20}$ (Corrected Σx_i^2) $-$ (Corrected mean)^2

$\Rightarrow$ Corrected Variance $= \dfrac{2160}{20} - \left(\dfrac{204}{20}\right)^2$

$\Rightarrow$ Corrected Variance $= \dfrac{2160 \times 20 - (204)^2}{(20)^2} = \dfrac{43200 - 41616}{400} = \dfrac{1584}{400}$

$\therefore$ Corrected standard deviation $= \sqrt{\dfrac{1584}{400}} = \dfrac{\sqrt{396}}{10} = \dfrac{19.899}{10} = 1.9899$

EXAMPLE 10 *The mean and variance of 7 observations are 8 and 16 respectively. If 5 of the observations are 2, 4, 10, 12, 14, find the remaining two observations.*

SOLUTION Let x and y be the remaining two observations. Then,

$$\text{Mean} = 8$$

$$\Rightarrow \quad \frac{2 + 4 + 10 + 12 + 14 + x + y}{7} = 8$$

$$\Rightarrow \quad 42 + x + y = 56$$

$$\Rightarrow \quad x + y = 14 \quad \text{...(i)}$$

and,

$$\text{Variance} = 16$$

$$\Rightarrow \quad \frac{1}{7}(2^2 + 4^2 + 10^2 + 12^2 + 14^2 + x^2 + y^2) - (\text{Mean})^2 = 16$$

$\Rightarrow \quad \frac{1}{7}(4 + 16 + 100 + 144 + 196 + x^2 + y^2) - 64 = 16$

$\Rightarrow \quad 460 + x^2 + y^2 = 7 \times 80$

$\Rightarrow \quad x^2 + y^2 = 100$ ...(ii)

Now, $(x + y)^2 + (x - y)^2 = 2(x^2 + y^2)$

$\Rightarrow \quad 196 + (x - y)^2 = 2 \times 100$

$\Rightarrow \quad (x - y)^2 = 4$

$\Rightarrow \quad x - y = \pm 2$

If $x - y = 2$, then $x + y = 14$ and $x - y = 2$ give $x = 8, y = 6$

If $x - y = -2$, then $x + y = 14$ and $x - y = -2$ give $x = 6, y = 8$.

Hence, the remaining two observations are 6 and 8.

Type V **ON FINDING THE VARIANCE WHEN DEVIATIONS ARE TAKEN FROM AN ASSUMED MEAN**

Following algorithm is helpful for finding the variances when deviations are taken from an assumed mean.

ALGORITHM

STEP I *Choose an assumed mean A (say).*

STEP II *Take the deviations d_i of the observations from an assumed mean i.e. obtain $d_i = x_i - A,\ i = 1, 2, \ldots, n$. Take the total of these deviations i.e. $\sum_{i=1}^{n} d_i$.*

STEP III *Square the deviations obtained in step II and obtain the total $\sum_{i=1}^{n} d_i^2$.*

STEP IV *Substitute the values of $\sum_{i=1}^{n} d_i^2$, $\sum_{i=1}^{n} d_i$ and n in the formula*

$$\text{Var}(X) = \frac{1}{n}\left(\sum_{i=1}^{n} d_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^2$$

EXAMPLE 11 *The scores of a batsman in 10 matches were as follows:* 38, 70, 48, 34, 42, 55, 63, 46, 54, 44 *Compute the variance and standard deviation.*

SOLUTION Let the assumed mean be $A = 48$.

Calculation of Variance

x_i	$d_i = x_i - A$	d_i^2
38	–10	100
70	22	484
48	0	0
34	–14	196
42	–6	36
55	7	49
63	15	225
46	–2	4
54	6	36
44	–4	16
	$\Sigma d_i = 14$	$\Sigma d_i^2 = 1146$

Here, $n = 10$, $\Sigma d_i = 14$ and $\Sigma d_i^2 = 1146$

$$\therefore \quad \text{Var}(X) = \frac{1}{n}\left(\Sigma d_i^2\right) - \left(\frac{1}{n}\Sigma d_i\right)^2 = \frac{1146}{10} - \left(\frac{14}{10}\right)^2 = 112.64$$

Hence, $S.D. = \sqrt{\text{Var}(X)} = \sqrt{112.64} = 10.61$

LEVEL-2

Type VI ON FINDING VARIANCE AND STANDARD DEVIATION OF INDIVIDUAL OBSERVATIONS

EXAMPLE 12 *Calculate the mean and standard deviation of first n natural numbers.* **[NCERT]**

SOLUTION Here $x_i = i$; $i = 1, 2, \ldots, n$.

Let $\overline{X}$ be the mean and σ be the S.D. Then,

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}(1+2+3+\ldots+n) = \frac{n(n+1)}{2n} = \frac{n+1}{2}$$

$$\text{Now,} \quad \sigma^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2\right) - (\overline{X})^2$$

$$\Rightarrow \quad \sigma^2 = \frac{1}{n}(1^2 + 2^2 + \ldots + n^2) - \left(\frac{n+1}{2}\right)^2$$

$$\Rightarrow \quad \sigma^2 = \frac{n(n+1)(2n+1)}{6n} - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}$$

$$\therefore \quad \text{Mean} = \frac{n+1}{2} \text{ and S.D.} = \sqrt{\frac{n^2-1}{12}}.$$

EXAMPLE 13 *Find the mean and standard deviation of first n terms of an A.P. whose first term is a and common difference is d.* **[NCERT EXEMPLAR]**

SOLUTION The terms of the A.P. are: $a, a+d, a+2d, a+3d, \ldots, a+(r-1)d, \ldots, a+(n-1)d$.
Let $\overline{X}$ be the mean of these terms. Then,

$$\overline{X} = \frac{1}{n}\{a+(a+d)+(a+2d)+\ldots+(a+(n-1)d)\} = \frac{1}{n}\left[\frac{n}{2}\left\{2a+(n-1)d\right\}\right] = a+(n-1)\frac{d}{2}$$

Let σ be the standard deviation of n terms of the A.P. then,

$$\sigma^2 = \frac{1}{n}\sum_{r=1}^{n}\left[\left\{a+(r-1)d\right\} - \overline{X}\right]^2 \qquad \left[\text{Using: } \sigma^2 = \frac{1}{n}\sum_{r=1}^{n}(x_i - \overline{X})^2\right]$$

$$\Rightarrow \quad \sigma^2 = \frac{1}{n}\sum_{r=1}^{n}\left[\left\{a+(r-1)d\right\} - \left\{a+(n-1)\frac{d}{2}\right\}\right]^2$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{4n}\left[\sum_{r=1}^{n}(2r-2-n+1)^2\right]$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{4n}\left[\sum_{r=1}^{n}\{2r-(n+1)\}^2\right]$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{4n}\left[\sum_{r=1}^{n}\left\{4r^2 - 4(n+1)r + (n+1)^2\right\}\right]$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{4n}\left[4\left(\sum_{r=1}^{n} r^2\right) - 4(n+1)\left(\sum_{r=1}^{n} r\right) + \sum_{r=1}^{n}(n+1)^2\right]$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{4n}\left\{\frac{4n(n+1)(2n+1)}{6} - \frac{4(n+1)n(n+1)}{2} + n(n+1)^2\right\}$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{4n}\left\{\frac{2n(n+1)(2n+1)}{3} - n(n+1)^2\right\}$$

$$\Rightarrow \quad \sigma^2 = \frac{d^2}{12n}\, n(n+1)\{2(2n+1) - 3(n+1)\} = \frac{(n^2-1)\, d^2}{12}$$

$$\Rightarrow \quad \sigma = d\sqrt{\frac{n^2-1}{12}}.$$

EXERCISE 32.4

LEVEL-1

1. Find the mean, variance and standard deviation for the following data:
   (i) 2, 4, 5, 6, 8, 17. (ii) 6, 7, 10, 12, 13, 4, 8, 12 [NCERT]
   (iii) 227, 235, 255, 269, 292, 299, 312, 321, 333, 348. (iv) 15, 22, 27, 11, 9, 21, 14, 9.
2. The variance of 20 observations is 5. If each observation is multiplied by 2, find the variance of the resulting observations. [NCERT]
3. The variance of 15 observations is 4. If each observation is increased by 9, find the variance of the resulting observations.
4. The mean of 5 observations is 4.4 and their variance is 8.24. If three of the observations are 1, 2 and 6, find the other two observations. [NCERT]
5. The mean and standard deviation of 6 observations are 8 and 4 respectively. If each observation is multiplied by 3, find the new mean and new standard deviation of the resulting observations. [NCERT]
6. The mean and variance of 8 observations are 9 and 9.25 respectively. If six of the observations are 6, 7, 10, 12, 12 and 13, find the remaining two observations. [NCERT]
7. For a group of 200 candidates, the mean and standard deviations of scores were found to be 40 and 15 respectively. Later on it was discovered that the scores of 43 and 35 were misread as 34 and 53 respectively. Find the correct mean and standard deviation.
8. The mean and standard deviation of 100 observations were calculated as 40 and 5.1 respectively by a student who took by mistake 50 instead of 40 for one observation. What are the correct mean and standard deviation? [NCERT]
9. The mean and standard deviation of 20 observations are found to be 10 and 2 respectively. On rechecking it was found that an observation 8 was incorrect. Calculate the correct mean and standard deviation in each of the following cases:
   (i) If wrong item is omitted [NCERT] (ii) if it is replaced by 12.
10. The mean and standard deviation of a group of 100 observations were found to be 20 and 3 respectively. Later on it was found that three observations were incorrect, which were recorded as 21, 21 and 18. Find the mean and standard deviation if the incorrect observations were omitted. [NCERT]

LEVEL-2

11. Show that the two formulae for the standard deviation of ungrouped data

$$\sigma = \sqrt{\frac{1}{n}\sum (x_i - \overline{X})^2} \text{ and } \sigma' = \sqrt{\frac{1}{n}\sum x_i^2 - \overline{X}^2} \text{ are equivalent, where } \overline{X} = \frac{1}{n}\sum x_i.$$

ANSWERS

1. (i) 7, 23.33, 4.83 (ii) 9, 9.25, 3.04 (iii) 289.10, 1539.77, 39.24 (iv) 16, 38.68, 6.22
2. 20 3. 4 4. 9, 4 5. 18, 12 6. 4, 8 7. 39.955, 14.9
8. Mean = 39.9 S.D. = 5 9. (i) 1.997 (ii) 1.98 10. 20, 3.035

HINTS TO NCERT & SELECTED PROBLEMS

1. (iii) Let the assumed mean be $A = 9$

Calculation of Variance

x_i	$d_i = (x_i - 9)$	d_i^2
6	-3	9
7	-2	4
10	1	1
12	3	9
13	4	16
4	-5	25
8	-1	1
12	3	9
	$\Sigma d_i = 0$	$\Sigma d_i^2 = 74$

Hence, $n = 8, a = 9, \Sigma d_i = 0$ and $\Sigma d_i^2 = 74$

$$\therefore \quad \overline{X} = a + \frac{\Sigma d_i}{n} = 9 + 0 = 9$$

and, $$\text{Var}(X) = \frac{1}{n}\left(\Sigma d_i^2\right) - \left(\frac{1}{n}\Sigma d_i\right)^2 = \frac{74}{8} - \left(\frac{0}{8}\right)^2 = 9.25$$

So, $$\text{S.D.} = \sqrt{\text{Var}(X)} = \sqrt{9.25} = 3.041$$

2. We know that if the observations $x_1, x_2, x_3, ..., x_n$ has variance Var (X). Then, observations $ax_1, ax_2, ..., ax_n$ has variance a^2 Var (X).

Thus, if variance of 20 observations is 5 and each observation is multiplied by 2, then variance of resulting observations is $2^2(5) = 20$.

4. Let the other two observations be x and y. Then,

$$\text{Mean} = 4.4 \Rightarrow \frac{1+2+6+x+y}{5} = 4.4 \Rightarrow x + y = 13 \quad ...(i)$$

Variance = 8.24

$$\Rightarrow \quad \frac{1}{5}\left(1^2 + 2^2 + 6^2 + x^2 + y^2\right) - (4.4)^2 = 8.24$$

$$\Rightarrow \quad \frac{41 + x^2 + y^2}{5} - 19.36 = 8.24$$

$$\Rightarrow \quad x^2 + y^2 + 41 = 138 \Rightarrow x^2 + y^2 = 97 \quad ...(ii)$$

Now, $(x-y)^2 + (x+y)^2$

$$\Rightarrow \quad (x-y)^2 + 169 = 2 \times 97 \quad \text{[Using (i) and (ii)]}$$

$$\Rightarrow \quad (x-y)^2 = 25$$

$$\Rightarrow \quad x - y = 5 \quad ...(iii)$$

Solving (i) and (iii), we get $x = 9$ and $y = 4$.

5. If the mean and standard deviation of observations $x_1, x_2 \ldots, x_n$ are $\overline{X}$ and σ respectively, then the mean and standard deviations of $ax_1, ax_2 \ldots, ax_n$ are $a\overline{X}$ and $|a|\sigma$ respectively.

$\therefore$ New mean $= 8 \times 3 = 24$ and, New standard deviation $= 3 \times 4 = 12$.

6. Let the remaining two observations be x and y. Then,

$$\text{Mean} = 9 \Rightarrow \frac{6+7+10+12+12+13+x+y}{8} = 9 \Rightarrow x+y = 12 \quad \text{...(i)}$$

Variance $= 9.25$

$$\therefore \quad \text{Variance} = \frac{1}{n}\Sigma x_i^2 - (\text{Mean})^2$$

$$\Rightarrow \quad \frac{1}{8}\left(36+49+100+144+144+169+x^2+y^2\right) - 9^2 = 9.25$$

$$\Rightarrow \quad 642 + x^2 + y^2 = 722$$

$$\Rightarrow \quad x^2 + y^2 = 80 \quad \text{...(ii)}$$

Now, $(x-y)^2 + (x+y)^2 = 2(x^2+y^2)$

$$\Rightarrow \quad (x-y)^2 + 144 = 2 \times 80$$

$$\Rightarrow \quad (x-y)^2 = 16$$

$$\Rightarrow \quad x - y = 4 \quad \text{...(iii)}$$

Solving (i) and (iii), we obtain $x = 8$, $y = 4$.

7. We have,

$n = 200$, Incorrected mean $= 40$ and, Incorrected S.D. $= 5.1$

Now, Incorrected mean $= 40$

$$\Rightarrow \quad \frac{\text{Incorrected } \Sigma x_i}{200} = 40 \qquad \left[\because \overline{X} = \frac{1}{n}\Sigma x_i\right]$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i = 8000$$

$$\therefore \quad \text{Corrected } \Sigma x_i = 8000 - (34+53) + (43+35) = 7991$$

So, $\text{Corrected mean} = \frac{7991}{200} = 39.955$

Now, Incorrected S.D. $= 15$

$$\Rightarrow \quad \text{Incorrected variance} = 225$$

$$\Rightarrow \quad \frac{1}{200}(\text{Incorrected } \Sigma x_i^2) - (40)^2 = 225$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i^2 = 365000$$

$$\therefore \quad \text{Corrected } \Sigma x_i^2 = \text{Incorrected } \Sigma x_i^2 - (34^2 + 53^2) + (43^2 + 35^2)$$

$$= 365000 - (1156 + 2809) + (1849 + 1225) = 364109$$

So, Corrected variance $= \frac{1}{200}(\text{Corrected } \Sigma x_i^2) - (\text{Corrected mean})^2$

$$= \frac{364109}{200} - (39.955)^2 = 1820.545 - 1596.402 = 224.143$$

$$\therefore \quad \text{Corrected S.D.} = \sqrt{224.143} = 14.971$$

8. We have,

$n = 100$, Incorrected mean $= 40$, Incorrected S.D. $= 5.1$

Now,

$$\text{Incorrected mean} = 40$$

$$\Rightarrow \quad \frac{\text{Incorrected } \Sigma x_i}{100} = 40$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i = 4000$$

$$\Rightarrow \quad \text{Corrected } \Sigma x_i = 4000 - 50 + 40 = 3990$$

$$\therefore \quad \text{Corrected mean} = \frac{3990}{100} = 39.90$$

Now,

$$\text{Incorrected S.D.} = 5.1$$

$$\Rightarrow \quad \text{Incorrected variance} = 26.01$$

$$\Rightarrow \quad \frac{1}{100}(\text{Incorrected } \Sigma x_i^2) - (\text{Incorrected mean})^2 = 26.01$$

$$\Rightarrow \quad \frac{1}{100}(\text{Incorrected } \Sigma x_i^2) - 40^2 = 26.01$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i^2 = 162601$$

$$\Rightarrow \quad \text{Corrected } \Sigma x_i^2 = 162601 - 50^2 + 40^2 = 161701$$

$$\therefore \quad \text{Corrected Variance} = \frac{161701}{100} - (39.9)^2 = 1617.01 - 1592.01 = 25$$

Hence, Corrected S.D. $= \sqrt{25} = 5$

9. We have, $n = 20$, Incorrected mean $= 10$, Incorrected S.D. $= 2$.

Now,

$$\text{Incorrected Mean} = 10$$

$$\Rightarrow \quad \frac{\text{Incorrected } \Sigma x_i}{20} = 10$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i = 200$$

and, Incorrected S.D. $= 2$

$$\Rightarrow \quad \text{Incorrected Variance} = 4$$

$$\Rightarrow \quad \frac{\text{Incorrected } \Sigma x_i^2}{20} - (10)^2 = 4$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i^2 = 2080$$

(i) When wrong item is omitted: In this case, $n = 19$.

$$\text{Corrected } \Sigma x_i = \text{Incorrected } \Sigma x_i - 8 = 200 - 8 = 192$$

$$\text{Corrected } \Sigma x_i^2 = \text{Incorrected } \Sigma x_i^2 - 8^2 = 2080 - 64 = 2016$$

$$\therefore \quad \text{Corrected mean} = \frac{\text{Corrected } \Sigma x_i}{19} = \frac{192}{19} = 10.105$$

$$\text{Corrected Variance} = \frac{\text{Corrected } \Sigma x_i^2}{19} - (\text{Corrected mean})^2$$

$$= \frac{2016}{19} - \left(\frac{192}{19}\right)^2 = \frac{38304 - 36864}{361} = \frac{1440}{361} = 3.9889$$

$$\therefore \quad \text{Corrected S.D.} = \sqrt{3.9889} = 1.997$$

(ii) When wrong observations 8 is replaced by 12: It this case, $n = 20$.

$$\text{Corrected } \Sigma x_i = \text{Incorrected } \Sigma x_i - 8 + 12 = 200 + 4 = 204$$

$$\text{Corrected } \Sigma x_i^2 = \text{Incorrected } \Sigma x_i^2 - 8^2 + 12^2 = 2080 - 64 + 144 = 2160$$

$$\therefore \quad \text{Corrected mean} = \frac{\text{Corrected } \Sigma x_i}{20} = \frac{204}{20} = 10.2$$

$$\text{Corrected Variance} = \frac{\text{Corrected } \Sigma x_i^2}{20} - (\text{Corrected mean})^2$$

$$= \frac{2160}{20} - \left(\frac{204}{20}\right)^2 = 108 - 104.04 = 3.96$$

$$\therefore \quad \text{Corrected S.D.} = \sqrt{3.96} = 1.98$$

10. We have,

$$n = 100, \text{ Incorrected mean} = 20, \text{ Incorrected S.D.} = 3$$

$$\therefore \quad \text{Incorrected mean} = 20$$

$$\Rightarrow \quad \text{Incorrected } \Sigma x_i = 20 \times 100$$

$$\Rightarrow \quad \text{Corrected } \Sigma x_i = 2000 - 21 - 21 - 18 = 1940$$

$$\therefore \quad \text{Corrected mean} = \frac{1940}{97} = 20$$

Now,

$$\text{Incorrected S.D.} = 3$$

$$\Rightarrow \quad \text{Incorrected Variance} = 9$$

$$\Rightarrow \quad \frac{1}{100}(\text{Incorrected } \Sigma x_i^2) - (\text{Incorrected Mean})^2 = 9$$

$$\Rightarrow \quad \frac{1}{100}(\text{Incorrected } \Sigma x_i^2) - 400 = 9$$

$$\Rightarrow \quad \text{Incorrected } x_i^2 = 40900$$

$$\Rightarrow \quad \text{Corrected } \Sigma x_i^2 = 40900 - 21^2 - 21^2 - 18^2 = 39694$$

$$\therefore \quad \text{Variance of the remaining observations} = \frac{39694}{97} - (20)^2 = 409.216 - 400 = 9.216$$

$$\therefore \quad \text{Corrected S.D.} = \sqrt{9.216} = 3.035$$

32.5.2 VARIANCE OF A DISCRETE FREQUENCY DISTRIBUTION

If x_i / f_i ; $i = 1, 2, ..., n$ is a discrete frequency distribution of a variate X, then

$$\text{Var}(X) = \frac{1}{N}\left\{ \sum_{i=1}^{n} f_i (x_i - \overline{X})^2 \right\} \quad \text{...(i)}$$

Also, $$\text{Var}(X) = \frac{1}{N}\left[\sum_{i=1}^{n} f_i (x_i^2 - 2x_i \overline{X} + \overline{X}^2) \right]$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{N}\left(\sum_{i=1}^{n} f_i x_i^2 \right) - 2\overline{X}\left(\frac{1}{N} \sum_{i=1}^{n} f_i x_i \right) + \frac{N\overline{X}^2}{N}$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{N}\left(\sum_{i=1}^{n} f_i x_i^2 \right) - 2\overline{X}^2 + \overline{X}^2 \qquad \left[\because \frac{1}{N} \sum_{i=1}^{n} f_i x_i = \overline{X}\right]$$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{N}\left(\sum_{i=1}^{n} f_i x_i^2 \right) - \overline{X}^2$$

or, $$\text{Var}(X) = \frac{1}{N}\left(\sum_{i=1}^{n} f_i x_i^2 \right) - \left(\frac{1}{N} \sum_{i=1}^{n} f_i x_i \right)^2 \quad \text{...(ii)}$$

If the values x_i of variable X or (and) frequencies f_i are large the calculation of variance from the above formulae is quite tedious and time consuming. In such a case, we take deviations of the values of variable X from an arbitrary point A(say). If $d_i = x_i - A,\ i = 1, 2, \dots, n$, then the above formula reduces to

$$\text{Var}(X) = \frac{1}{N}\left(\Sigma f_i d_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{n} f_i d_i\right)^2 \quad \text{...(iii)}$$

Sometimes $d_i = x_i - A$ are divisible by a common number h(say). If we define $u_i = \frac{x_i - A}{h} = \frac{d_i}{h}, i = 1, 2, \dots, n$, then we obtain the following formula for variance.

$$\text{Var}(X) = h^2\left[\left(\frac{1}{N}\sum_{i=1}^{n} f_i u_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{n} f_i u_i\right)^2\right] \quad \text{...(iv)}$$

In order to compute variance by using the following formula

$\text{Var}(X) = \frac{1}{N}\left[\sum_{i=1}^{n} f_i (x_i - \overline{X})^2\right]$, we may use the following algorithm.

ALGORITHM

STEP I *Obtain the given frequency distribution.*

STEP II *Find the mean $\overline{X}$ of the given frequency distribution.*

STEP III *Compute deviations $(x_i - \overline{X})$ from the mean $\overline{X}$.*

STEP IV *Find the squares of deviations obtained in step III.*

STEP V *Multiply the squared deviations by respective frequencies and obtain the total $\Sigma f_i (x_i - \overline{X})^2$.*

STEP VI *Divide the total obtained in step V by $N = \Sigma f_i$ to obtain the variance.*

Following example illustrates the above algorithm.

ILLUSTRATIVE EXAMPLES

LEVEL-1

EXAMPLE 1 *Find the variance and standard deviation of the following frequency distribution:*

Variable (x_i)	2	4	6	8	10	12	14	16
Frequency (f_i)	4	4	5	15	8	5	4	5

SOLUTION *Calculation of Variance and Standard Deviation*

Variable x_i	Frequency f_i	$f_i x_i$	$x_i - \overline{X} = x_i - 9$	$(x_i - \overline{X})^2$	$f_i (x_i - \overline{X})^2$
2	4	8	−7	49	196
4	4	16	−5	25	100
6	5	30	−3	9	45
8	15	120	−1	1	15
10	8	80	1	1	8
12	5	60	3	9	45
14	4	56	5	25	100
16	5	80	7	49	245
	$N = \Sigma f_i = 50$	$\Sigma f_i x_i = 450$			$\Sigma f_i (x_i - \overline{X})^2 = 754$

Here, $N = 50$, $\Sigma f_i x_i = 450$ and, $\Sigma f_i (x_i - \overline{X})^2 = 754$

$$\therefore \quad \overline{X} = \frac{1}{N} \Sigma f_i x_i = \frac{450}{50} = 9$$

and, $$\text{Var}(X) = \frac{1}{N} \left\{ \Sigma f_i (x_i - \overline{X})^2 \right\} = \frac{754}{50} = 15.08$$

Hence, $S.D. = \sqrt{\text{Var}(X)} = \sqrt{15.08} = 3.88$

NOTE: *In practice the calculation of S.D. and variance by the above algorithm is rarely used, because if the actual mean is in fractions the calculation is quite tedious and time consuming.*

In order to compute the variance by using the following formula

$\text{Var}(X) = \left[\left(\frac{1}{N} \Sigma f_i d_i^2 \right) - \left(\frac{1}{N} \Sigma f_i d_i \right)^2 \right]$, where $d_i = x_i - A$, we may use the following algorithm.

ALGORITHM

STE I *Take the deviations of observations from an assumed mean A(say) and denote these deviations by d_i.*

STEP II *Multiply the deviations by the respective frequencies and obtain the total $\Sigma f_i d_i$.*

STEP III *Obtain the squares of deviations obtained in step I i.e. d_i^2.*

STEP IV *Multiply the squared deviations by respective frequencies and obtain the total $\Sigma f_i d_i^2$.*

STEP V *Substitute the values in the formula*

$$\text{Var}(X) = \left(\frac{1}{N} \Sigma f_i d_i^2 \right) - \left(\frac{1}{N} \Sigma f_i d_i \right)^2$$

Following examples illustrate the above algorithm.

EXAMPLE 2 *Calculate the variance and standard deviation from the data given below:*

Size of item	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Frequency	3	7	22	60	85	32	8

SOLUTION Let the assumed mean be $A = 6.5$

Calculation of Variance and Standard Deviation

Size of item x_i	f_i	$d_i = x_i - 6.5$	d_i^2	$f_i d_i$	$f_i d_i^2$
3.5	3	−3	9	−9	27
4.5	7	−2	4	−14	28
5.5	22	−1	1	−22	22
6.5	60	0	0	0	0
7.5	85	1	1	85	85
8.5	32	2	4	64	128
9.5	8	3	9	24	72
	$N = \Sigma f_i = 217$			$\Sigma f_i d_i = 128$	$\Sigma f_i d_i^2 = 362$

Here, $N = 217$, $\Sigma f_i d_i = 128$ and $\Sigma f_i d_i^2 = 362$

$$\therefore \quad \text{Var}(X) = \left(\frac{1}{N}\Sigma f_i d_i^2\right) - \left(\frac{1}{N}\Sigma f_i d_i\right)^2 = \frac{362}{217} - \left(\frac{128}{217}\right)^2 = 1.668 - 0.347 = 1.321$$

Hence, $\quad \text{S.D.} = \sqrt{\text{Var}(X)} = \sqrt{1.321} = 1.149$

REMARK *Sometimes deviations d_i in the algorithm given above are divisible by a common number h. In such a case, we define $u_i = \frac{x_i - a}{h} = \frac{d_i}{h}$, $i = 1, 2, \ldots, n$ and the formula for computing variance is*

$$\text{Var}(X) = h^2\left[\frac{1}{N}\left(\sum_{i=1}^{n} f_i u_i^2\right) - \left(\frac{1}{N}\sum_{i=1}^{n} f_i u_i\right)^2\right].$$

EXAMPLE 3 *Find the variance and standard deviation for the following distribution:*

X:	4.5	14.5	24.5	34.5	44.5	54.5	64.5
f:	1	5	12	22	17	9	4

SOLUTION *Calculation of Variance and Standard Deviation*

x_i	f_i	$d_i = x_i - 34.5$	$u_i = \frac{x_i - 34.5}{10}$	$f_i u_i$	u_i^2	$f_i u_i^2$
4.5	1	–30	–3	–3	9	9
14.5	5	–20	–2	–10	4	20
24.5	12	–10	–1	–12	1	12
34.5	22	0	0	0	0	0
44.5	17	10	1	17	1	17
54.5	9	20	2	18	4	36
64.5	4	30	3	12	9	36
	$N = \Sigma f_i = 70$			$\Sigma f_i u_i = 22$		$\Sigma f_i u_i^2 = 130$

Here, $N = 70$, $\Sigma f_i u_i = 20$, $\Sigma f_i u_i^2 = 130$ and $h = 10$

$$\therefore \quad \text{Var}(X) = h^2\left[\left(\frac{1}{N}\Sigma f_i u_i^2\right) - \left(\frac{1}{N}\Sigma f_i u_i\right)^2\right]$$

$$\Rightarrow \quad \text{Var}(X) = 100\left[\frac{130}{70} - \left(\frac{22}{70}\right)^2\right] = 100\left[\frac{13}{7} - \left(\frac{11}{35}\right)^2\right] = 100\,[1.857 - 0.098] = 175.822$$

Hence, $\quad \text{S.D.} = \sqrt{Var(X)} = \sqrt{175.822} = 13.259$

EXAMPLE 4 *The following table gives the number of finished articles turned out per day by different number of workers in a factory. Find the standard deviation of the daily output of finished articles.*

Number of articles:	18	19	20	21	22	23	24	25	26	27
No. of workers:	3	7	11	14	18	17	13	8	5	4

SOLUTION *Calculation of Standard Deviation*

x	f	$d_i = x_i - 23$	d_i^2	$f_i\, d_i$	$f_i\, d_i^2$
18	3	−5	25	−15	75
19	7	−4	16	−28	112
20	11	−3	9	−33	99
21	14	−2	4	−28	56
22	18	−1	1	−18	18
23	17	0	0	0	0
24	13	1	1	13	13
25	8	2	4	16	32
26	5	3	9	15	45
27	4	4	16	16	64
	$N = \Sigma f_i = 100$			$\Sigma f_i d_i = -62$	$\Sigma f_i d_i^2 = 514$

Clearly, $N = 100$, $\Sigma f_i d_i = -62$ and $\Sigma f_i d_i^2 = 514$

$$\therefore \quad \sigma^2 = \frac{1}{N}\left(\Sigma f_i\, d_i^2\right) - \left(\frac{1}{N}\,\Sigma f_i\, d_i\right)^2 = \frac{514}{100} - \left(-\frac{62}{100}\right)^2 = \frac{47556}{10000}$$

Hence, $\sigma = \sqrt{\frac{47556}{10000}} = \frac{218.07}{100} = 2.1807$

LEVEL-2

EXAMPLE 5 *If a is a positive integer and the frequency distribution:*

$x:$	a	$2a$	$3a$	$4a$	$5a$	$6a$
$f:$	2	1	1	1	1	1

has a variance of 160. Determine the value of a. **[NCERT EXEMPLAR]**

SOLUTION *Computation of Variance*

x_i	f_i	$f_i x_i$	$f_i x_i^2$
a	2	$2a$	$2a^2$
$2a$	1	$2a$	$4a^2$
$3a$	1	$3a$	$9a^2$
$4a$	1	$4a$	$16a^2$
$5a$	1	$5a$	$25a^2$
$6a$	1	$6a$	$36a^2$
	$N = \Sigma f_i = 7$	$\Sigma f_i x_i = 22a$	$\Sigma f_i x_i^2 = 92a^2$

Here, $N = 7$, $\Sigma f_i x_i = 22a$, $\Sigma f_i x_i^2 = 92a^2$ and Variance $= 160$

Now,

Variance $= 160$

$$\Rightarrow \quad 160 = \left(\frac{1}{N}\Sigma f_i x_i^2\right) - \left(\frac{1}{N}\Sigma f_i x_i\right)^2$$

$$\Rightarrow \quad 160 = \frac{92a^2}{7} - \left(\frac{22a}{7}\right)^2 \Rightarrow 160 = \frac{644a^2 - 484a^2}{49} \Rightarrow 160 = \frac{160a^2}{49} \Rightarrow a^2 = 49 \Rightarrow a = 7$$

EXAMPLE 6 *There are 60 students in a class. The following is the frequency distribution of marks obtained by the students in a test:*

Marks:	0	1	2	3	4	5
Frequency:	$x-2$	x	x^2	$(x+1)^2$	$2x$	$x+1$

where x is a positive integer. Determine the mean and standard deviation of the marks.

[NCERT EXEMPLAR]

SOLUTION It is given that there are 60 students in the class.

$\therefore \quad (x-2)+x+x^2+(x+1)^2+2x+(x+1)=60$

$\Rightarrow \quad 2x^2+7x-60=0$

$\Rightarrow \quad (2x+15)(x-4)=0$

$\Rightarrow \quad x-4=0 \qquad [\because\ x>0\ \therefore\ 2x+15 \neq 0]$

$\Rightarrow \quad x=4$

Thus, we obtain the following frequency distribution:

Marks:	0	1	2	3	4	5
Frequency:	2	4	16	25	8	5

Computation of mean and standard deviation

Marks (x_i)	*Frequency* (f_i)	$f_i x_i$	$f_i x_i^2$
0	2	0	0
1	4	4	4
2	16	32	64
3	25	75	225
4	8	32	128
5	5	25	125
	$N=\Sigma f_i=60$	$\Sigma f_i x_i=168$	$\Sigma f_i x_i^2=546$

Here, $N=60,\ \Sigma f_i x_i=168,\ \Sigma f_i x_i^2=546$

$\therefore \quad \text{Mean}=\frac{1}{N}\Sigma f_i x_i=\frac{168}{60}=2.8$

and, $\text{Variance}=\left(\frac{1}{N}\Sigma f_i x_i^2\right)-\left(\frac{1}{N}\Sigma f_i x_i\right)^2=\frac{546}{60}-\left(\frac{168}{60}\right)^2=9.1-7.84=1.26$

Hence, $\text{S.D.}=\sqrt{\text{Variance}}=\sqrt{1.26}=1.122$

EXERCISE 32.5

LEVEL-1

1. Find the standard deviation for the following distribution:

x:	4.5	14.5	24.5	34.5	44.5	54.5	64.5
f:	1	5	12	22	17	9	4

2. Table below shows the frequency f with which 'x' alpha particles were radiated from a diskette

x:	0	1	2	3	4	5	6	7	8	9	10	11	12
f:	51	203	383	525	532	408	273	139	43	27	10	4	2

Calculate the mean and variance.

3. Find the mean, and standard deviation for the following data:

(i)

Year render:	10	20	30	40	50	60
No. of persons (cumulative):	15	32	51	78	97	109

(ii)

Marks:	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Frequency:	1	6	6	8	8	2	2	3	0	2	1	0	0	0	1

[NCERT EXEMPLAR]

4. Find the standard deviation for the following data:

(i)

x:	3	8	13	18	23
f:	7	10	15	10	6

(ii)

x:	2	3	4	5	6	7
f:	4	9	16	14	11	6

[NCERT EXEMPLAR]

ANSWERS

1. 13.26 2. $\overline{X} = 3.88$, $\sigma^2 = 3.64$ 3. (i) $\overline{X} = 37.25$ years, S.D. = 15.5 years. (ii) $X = 5.975$, S.D. = 2.85. 4. (i) 6.12 (ii) 1.38

32.5.3 VARIANCE OF A GROUPED OR CONTINUOUS FREQUENCY DISTRIBUTION

In a grouped or continuous frequency distribution any of the methods discussed above for a discrete frequency distribution can be used. We may use the following algorithm for computing variance of a grouped or continuous frequency distribution.

ALGORITHM

STEP I *Find the mid-points of various classes.*

STEP II *Take the deviations of these mid-points from an assumed mean. Denote these deviations by d_i.*

STEP III *Divide the deviations in step II by the class interval h and denote them by u_i., i.e. $u_i = \frac{d_i}{h}$.*

STEP IV *Multiply the frequency of each class with the corresponding u_i and obtain $\Sigma f_i u_i$.*

STEP V *Square the values of u_i and multiply them with the corresponding frequencies and obtain $\Sigma f_i u_i^2$.*

STEP VI *Substitute the values of $\Sigma f_i u_i$, $\Sigma f_i u_i^2$ h and N = Σf_i in the formula*

$$\text{Var}(X) = h^2 \left\{ \frac{1}{N} \Sigma f_i u_i^2 - \left(\frac{1}{N} \Sigma f_i u_i \right)^2 \right\}$$

Following examples will illustrate the above algorithm.

ILLUSTRATIVE EXAMPLES

LEVEL-1

EXAMPLE 1 *Calculate the mean and standard deviation for the following distribution:*

Marks:	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of students:	3	6	13	15	14	5	4

SOLUTION *Calculation of Standard Deviation*

Class-interval	Frequency (f_i)	Mid-values (x_i)	$u_i = \frac{x_i - 55}{10}$	$f_i u_i$	u_i^2	$f_i u_i^2$
20-30	3	25	−3	−9	9	27
30-40	6	35	−2	−12	4	24
40-50	13	45	−1	−13	1	13
50-60	15	55	0	0	0	0
60-70	14	65	1	14	1	14
70-80	5	75	2	10	4	20
80-90	4	85	3	12	9	36
	$N = \Sigma f_i = 60$			$\Sigma f_i u_i = 2$		$\Sigma f_i u_i^2 = 134$

Here, $N = 60,\ \Sigma f_i u_i = 2,\ \Sigma f_i u_i^2 = 134$ and $h = 10$

$\therefore$ Mean $= \overline{X} = A + h\left(\frac{1}{N}\ \Sigma f_i u_i\right) = 55 + 10\left(\frac{2}{60}\right) = 55.333$

and, $\text{Var}(X) = h^2\left\{\left(\frac{1}{N}\ \Sigma f_i u_i^2\right) - \left(\frac{1}{N}\ \Sigma f_i u_i\right)^2\right\} = 100\left[\frac{134}{60} - \left(\frac{2}{60}\right)^2\right] = 222.9$

$\therefore$ S.D. $= \sqrt{\text{Var}(X)} = \sqrt{222.9} = 14.94.$

EXAMPLE 2 *The following table gives the distribution of income of 100 families in a village. Calculate the standard deviation:*

Income ₹	0-1000	1000-2000	2000-3000	3000-4000	4000-5000	5000-6000
No. of Families	18	26	30	12	10	4

SOLUTION *Calculation of Standard Deviation*

Income ₹	Mid-values x_i	No. of families (frequencies) f_i	$u_i = \frac{x_i - 2500}{1000}$	$f_i u_i$	u_i^2	$f_i u_i^2$
0-1000	500	18	−2	−36	4	72
1000-2000	1500	26	−1	−26	1	26
2000-3000	2500	30	0	0	0	0
3000-4000	3500	12	1	12	1	12
4000-5000	4500	10	2	20	4	40
5000-6000	5500	4	3	12	9	36
		$\Sigma f_i = 100$		$\Sigma f_i u_i = -18$		$\Sigma f_i u_i^2 = 186$

Here, $N = 100,\ \Sigma f_i u_i = -18,\ \Sigma f_i u_i^2 = 186$ and, $h = 1000$

$\therefore$ $\text{Var}(X) = h^2\left\{\frac{1}{N}\left(\Sigma f_i u_i^2\right) - \left(\frac{1}{N}\ \Sigma f_i u_i\right)^2\right\} = (1000)^2\left\{\frac{186}{100} - \left(\frac{-18}{100}\right)^2\right\} = 1827600$

Hence, S.D. $= \sqrt{\text{Var}(X)} = \sqrt{1827600} = 1351.88$

EXAMPLE 3 *Calculate the mean and standard deviation for the following data:*

Wages upto (in ₹)	15	30	45	60	75	90	105	120
No. of workers	12	30	65	107	157	202	222	230

SOLUTION We are given the cumulative frequency distribution. So, first we will prepare the frequency distribution as given below:

Class-interval	*Cumulative frequency*	*Mid-values*	*Frequency*	$u_i = \frac{x_i - 67.5}{15}$	$f_i u_i$	$f_i u_i^2$
0-15	12	7.5	12	–4	–48	192
15-30	30	22.5	18	–3	–54	162
30-45	65	37.5	35	–2	–70	140
45-60	107	52.5	42	–1	–42	42
60-75	157	67.5	50	0	0	0
75-90	202	82.5	45	1	45	45
90-105	222	97.5	20	2	40	80
105-120	230	112.5	8	3	24	72
			$\Sigma f_i = 230$		$\Sigma f_i u_i = -105$	$\Sigma f_i u_i^2 = 733$

Here, $A = 67.5,\ h = 15,\ N = 230,\ \Sigma f_i u_i = -105$ and $\Sigma f_i u_i^2 = 733$

$$\therefore \quad \text{Mean} = A + h\left(\frac{1}{N}\Sigma f_i u_i\right) = 67.5 + 15\left(\frac{-105}{230}\right) = 67.5 - 6.85 = 60.65$$

and, $$\text{Var}(X) = h^2\left\{\frac{1}{N}\Sigma f_i u_i^2 - \left(\frac{1}{N}\Sigma f_i u_i\right)^2\right\}$$

$$\Rightarrow \quad \text{Var}(X) = 225\left\{\frac{733}{230} - \left(\frac{-105}{230}\right)^2\right\} = 225\,(3.18 - 0.2025) = 669.9375$$

$$\therefore \quad \text{S.D.} = \sqrt{\text{Var}(X)} = \sqrt{669.9375} = 25.883$$

EXAMPLE 4 *The measurements of the diameters (in mm) of the heads of 107 screws are given below:*

Diameter (in mm)	33-35	36-38	39-41	42-44	45-47
No. of screws	17	19	23	21	27

Calculate the standard deviation.

SOLUTION Here the class intervals are formed by the inclusive method. But, the mid-points of class-intervals remain same whether they are formed by inclusive method or exclusive method. So there is no need to convert them into an exclusive series.

Calculation of Standard Deviation

Diameter (in mm)	*Mid-values* x_i	*No. of screws* f_i	$u_i = \frac{x_i - 40}{3}$	$f_i u_i$	$f_i u_i^2$
33-35	34	17	–2	–34	68
36-38	37	19	–1	–19	19
39-41	40	23	0	0	0
42-44	43	21	1	21	21
45-47	46	27	2	54	108
		$\Sigma f_i = 107$		$\Sigma f_i u_i = 22$	$\Sigma f_i u_i^2 = 216$

Here $N = \Sigma f_i = 107$, $\Sigma f_i u_i = 22$, $\Sigma f_i u_i^2 = 216$, $A = 40$ and, $h = 3$

$$\therefore \quad \text{Var}(X) = h^2 \left\{ \left(\frac{1}{N} \Sigma f_i u_i^2 \right) - \left(\frac{1}{N} \Sigma f_i u_i \right)^2 \right\} = 9 \left\{ \frac{216}{107} - \left(\frac{22}{107} \right)^2 \right\}$$

$$\Rightarrow \quad \text{Var}(X) = 9(2.0187 - 0.0420) = 9 \times 1.9767 = 17.7903$$

$$\therefore \quad \text{S.D.} = \sqrt{17.7903} = 4.2178.$$

EXAMPLE 5 *Calculate the mean and standard deviation for the following table given the age distribution of a group of people:*

Age:	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of persons:	3	51	122	141	130	51	2

SOLUTION Here $A = 55$, $h = 10$.

Calculation of Mean and Standard Deviation

Age	*Mid-values* (x_i)	*Number of persons* (f_i)	$u_i = \frac{x_i - 55}{10}$	$f_i u_i$	u_i^2	$f_i u_i^2$
20-30	25	3	−3	−9	9	27
30-40	35	51	−2	−102	4	204
40-50	45	122	−1	−122	1	122
50-60	55	141	0	0	0	0
60-70	65	130	1	130	1	130
70-80	75	51	2	102	4	204
80-90	85	2	3	6	9	18
		$N = \Sigma f_i = 500$		$\Sigma f_i u_i = 5$		$\Sigma f_i u_i^2 = 705$

Here, $N = \Sigma f_i = 500$, $\Sigma f_i u_i = 5$ and, $\Sigma f_i u_i^2 = 705$

$$\therefore \quad \overline{X} = A + h \left(\frac{1}{N} \Sigma f_i u_i \right) = 55 + 10 \left(\frac{5}{500} \right) = 55.1$$

$$\text{and,} \quad \sigma^2 = h^2 \left\{ \left(\frac{1}{N} \Sigma f_i u_i^2 \right) - \left(\frac{1}{N} \Sigma f_i u_i \right)^2 \right\}$$

$$\Rightarrow \quad \sigma^2 = 100 \left\{ \frac{705}{500} - \left(\frac{5}{500} \right)^2 \right\} = \frac{100}{50000}(70500 - 5) = \frac{70495}{500} = \frac{14099}{100}$$

$$\Rightarrow \quad \sigma = \frac{\sqrt{14099}}{10} = \frac{118.739}{10} = 11.8739.$$

EXERCISE 32.6

LEVEL-1

1. Calculate the mean and S.D. for the following data:

Expenditure (in ₹):	0-10	10-20	20-30	30-40	40-50
Frequency:	14	13	27	21	15

2. Calculate the standard deviation for the following data:

Class:	0-30	30-60	60-90	90-120	120-150	150-180	180-210
Frequency:	9	17	43	82	81	44	24

3. Calculate the A.M. and S.D. for the following distribution:

Class:	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Frequency:	18	16	15	12	10	5	2	1

4. A student obtained the mean and standard deviation of 100 observations as 40 and 5.1 respectively. It was later found that one observation was wrongly copied as 50, the correct figure being 40. Find the correct mean and S.D.

5. Calculate the mean, median and standard deviation of the following distribution:

Class-interval:	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70
Frequency:	2	3	8	12	16	5	2	3

LEVEL-2

6. Find the mean and variance of frequency distribution given below:

x_i :	$1 \le x < 3$	$3 \le x < 5$	$5 \le x < 7$	$7 \le x < 10$
f_1 :	6	4	5	1

7. The weight of coffee in 70 jars is shown in the following table: [NCERT EXEMPLAR]

Weight (in grams):	200-201	201-202	202-203	203-204	204-205	205-206
Frequency:	13	27	18	10	1	1

Determine the variance and standard deviation of the above distribution.
[NCERT EXEMPLAR]

8. Mean and standard deviation of 100 observations were found to be 40 and 10 respectively. If at the time of calculation two observations were wrongly taken as 30 and 70 in place of 3 and 27 respectively, find the correct standard deviation. [NCERT EXEMPLAR]

9. While calculating the mean and variance of 10 readings, a student wrongly used the reading of 52 for the correct reading 25. He obtained the mean and variance as 45 and 16 respectively. Find the correct mean and the variance. [NCERT EXEMPLAR]

10. Calculate mean, variance and standard deviation of the following frequency distribution:

Class :	1-10	10-20	20-30	30-40	40-50	50-60
Frequency :	11	29	18	4	5	3

ANSWERS

1. $\overline{X} = 26.11, \sigma = 12.86.$ 2. $\overline{X} = 118.7, \sigma = 42.51.$ 3. $AM = 26.01,\ S.D. = 17.47$
4. $\overline{X} = 39.9, \sigma = 5.$ 5. $\overline{X} = 50.35,\ \sigma = 7.94,$ median $= 50.65$
6. Mean $= 5.5$, Variance $= 4.26$ 7. Variance = 1.16 gm, S.D. 1.08 gm
8. 10.24 9. Mean = 42.3, Variance = 43.81 10. Mean = 21.5, Variance = 161, S.D. = 12.7

32.7 ANALYSIS OF FREQUENCY DISTRIBUTIONS

In this section, we shall see how we can use various measures of dispersion to compare two or more series. In the earlier sections of this chapter we have seen that the mean deviation and standard deviation have the same units in which the data are given. Therefore, measures of dispersion are unable to compare two or more series which are measured in different units even if they have the same mean. Thus, we require those measures which are independent of the units. The measure of variability which is independent of units is called coefficient of variation (C.V.) and is defined as

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100, \text{ where } \sigma \text{ and } \overline{X} \text{ are the standard deviation and mean of the data.}$$

For comparing the variability of two series, we calculate the coefficient of variation for each series. The series having greater C.V. is said to be more variable or conversely less consistent, less uniform, less stable or less homogeneous than the other and the series having lesser C.V. is said to be more consistent (or homogeneous) than the other.

Let there be two frequency distributions with standard deviations σ_1 and σ_2 and equal mean $\overline{X}$. Then,

$$\text{C.V. (1st distribution)} = \frac{\sigma_1}{\overline{X}} \times 100 \text{ and, C.V. (2nd distribution)} = \frac{\sigma_2}{\overline{X}} \times 100$$

$$\therefore \quad \frac{\text{C.V. (1st distribution)}}{\text{C.V. (2nd distribution)}} = \frac{\frac{\sigma_1}{\overline{X}} \times 100}{\frac{\sigma_2}{\overline{X}} \times 100} = \frac{\sigma_1}{\sigma_2}$$

This means that the two distributions can be compared on the basis of the values of their standard deviations σ_1 and σ_2 only.

Thus, if two series have equal means then the series with greater standard deviation (or variance) is said to be more variable or dispersed than the other. Also the series with lesser value of the standard deviation (or variance) is said to be more consistent than the other.

ILLUSTRATIVE EXAMPLES

LEVEL-1

EXAMPLE 1 *An analysis of monthly wages paid to the workers of two firms A and B belonging to the same industry gives the following results:*

	Firm A	*Firm B*
Number of workers	1000	1200
Average monthly wages	₹ 2800	₹ 2800
Variance of distribution of wages	100	169

In which firm, A or B is there greater variability in individual wages?

SOLUTION We observe that the average monthly wages in both the firms is same i.e. Rs. 2800. Therefore, the firm with greater variance will have more variability. Thus, firm *B* has greater variability in individual wages.

EXAMPLE 2 *An analysis of monthly wages paid to workers in two firms A and B, belonging to the same industry, gives the following results:* **[NCERT]**

	Firm A	*Firm B*
No. of wage earners	586	648
Mean of monthly wages	₹ 5253	₹ 5253
Variance of the distribution of wages	100	121

(i) *Which firm A or B pays out larger amount as monthly wages?*

(ii) *Which firm A or B shows greater variability in individual wages?*

SOLUTION (i) Firm A:

Number of wage earners (say) $n_1 = 586$

Mean of monthly wages (say) $\overline{X}_1 = ₹\ 5253$

$$\therefore \quad \text{Mean of monthly wages} = \frac{\text{Total monthly wage}}{\text{Number of workers}}$$

$$\Rightarrow \quad 5253 = \frac{\text{Total monthly wages}}{586}$$

$$\Rightarrow \quad \text{Total monthly wages} = ₹\,(5253 \times 586) = ₹\ 3078258$$

Firm B:

Number of wage earners (say) $n_2 = 648$

Mean of monthly wages $= ₹\ 5253$

$\therefore$ Mean of monthly wages $= \dfrac{\text{Total monthly wages}}{\text{Number of workers}}$

$\Rightarrow$ $5253 = \dfrac{\text{Total monthly wages}}{648}$

$\Rightarrow$ Total monthly wages $=$ ₹ (5253×648) $=$ ₹ 3403944

Clearly, firm B pays out larger amount as monthly wages.

(ii) Since firms A and B have the same mean. Therefore, the firm with greater variance will have more variability individual wages.

Clearly, Variance of firm $B >$ Variance of firm A.

Hence, firm B will have greater variability in individual wages.

EXAMPLE 3 *The following values are calculated in respect of heights and weights of the students of a section of class XI:*

	Height	*Weight*
Mean	162.6 cm	52.36 kg
Variance	127.69 cm^2	23.1361 kg^2

Can we say that the weights show greater variation than the heights? **[NCERT]**

SOLUTION In order to compare the variability of height and weight, we have to calculate their coefficients of variation. Let σ_1 and σ_2 denote the standard deviations of height and weight respectively. Further, let $\overline{X}_1$
and $\overline{X}_2$ be the mean height and weight respectively.

We have,

$$\overline{X}_1 = 162.6, \qquad \overline{X}_2 = 52.36$$

$$\sigma_1^2 = 127.69 \quad \text{and} \quad \sigma_2^2 = 23.1361$$

$$\Rightarrow \quad \sigma_1 = \sqrt{127.69} = 11.3 \quad \text{and} \quad \sigma_2 = \sqrt{23.1361} = 4.81$$

Now,

$$\text{Coefficient of variation in heights} = \frac{\sigma_1}{\overline{X}_1} \times 100 = \frac{11.3}{162.6} \times 100 = 6.95$$

and,

$$\text{Coefficient of variation in weights} = \frac{\sigma_2}{\overline{X}_2} \times 100 = \frac{4.81}{52.36} \times 100 = 9.18$$

Clearly, coefficient of variation in weights is greater than the coefficient of variation in heights. So, weights shows more variability than heights.

EXAMPLE 4 *The sum and sum of squares corresponding to length x (in cm) and weight y (in gm) of 50 plant products are given below:*

$$\sum_{i=1}^{50} x_i = 212, \quad \sum_{i=1}^{50} x_i^2 = 902.8, \quad \sum_{i=1}^{50} y_i = 261, \quad \sum_{i=1}^{n} y_i^2 = 1457.6$$ **[NCERT]**

Which is more varying, the length or weight?

SOLUTION We have,

$$\sum_{i=1}^{50} x_i = 212 \quad \text{and} \quad \sum_{i=1}^{50} x_i^2 = 902.80$$

$$\therefore \quad \overline{X} = \frac{\sum_{i=1}^{50} x_i}{50} \quad \text{and} \quad \sigma_X^2 = \frac{1}{50}\left(\sum_{i=1}^{50} x_i^2\right) - \left(\frac{1}{50}\sum_{i=1}^{50} x_i\right)^2$$

$$\Rightarrow \quad \bar{X} = \frac{212}{50} \quad \text{and} \quad \sigma_X^2 = \frac{902.80}{50} - \left(\frac{212}{50}\right)^2$$

$$\Rightarrow \quad \bar{X} = 4.24 \quad \text{and} \quad \sigma_X^2 = 18.056 - (4.24)^2 = 18.056 - 17.9776 = 0.0784$$

$$\Rightarrow \quad \bar{X} = 4.24 \quad \text{and} \quad \sigma_X = \sqrt{0.0784} = 0.28$$

It is given that

$$\sum_{i=1}^{50} y_i = 261 \quad \text{and} \quad \sum_{i=1}^{50} y_i^2 = 1457.6$$

$$\therefore \quad \bar{Y} = \frac{\sum_{i=1}^{50} y_i}{50} \quad \text{and} \quad \sigma_Y^2 = \frac{1}{50}\left(\sum_{i=1}^{50} y_i^2\right) - \left(\frac{1}{50}\sum_{i=1}^{50} y_i\right)^2$$

$$\Rightarrow \quad \bar{Y} = \frac{261}{50} \quad \text{and} \quad \sigma_y^2 = \frac{1457.6}{50} - \left(\frac{261}{50}\right)^2$$

$$\Rightarrow \quad \bar{Y} = 5.22 \quad \text{and} \quad \sigma_Y^2 = 29.152 - (5.22)^2 = 1.9036$$

$$\Rightarrow \quad \bar{Y} = 5.22 \quad \text{and} \quad \sigma_Y = 1.3797$$

In order to determine the variability of length and weight, we will have to compute the coefficients of variations in lengths and weights.

$$\text{Coefficient of variation in lengths} = \frac{\sigma_X}{\bar{X}} \times 100 = \frac{0.28}{4.24} \times 100 = 6.60$$

$$\text{Coefficient of variation in weights} = \frac{\sigma_Y}{\bar{Y}} \times 100 = \frac{1.3797}{5.22} \times 100 = 26.43$$

Clearly, coefficient of variation is weights is greater than the coefficient of variation in lengths. Hence, weights have more variability than lengths.

EXAMPLE 5 *The following is the record of goals scored by team A in football session.*

Number of goals scored:	0	1	2	3	4
Number of matches:	1	9	7	5	3

For the team B, mean number of goals scored per match was 2 with a standard deviation 1.25 goals. Find which team may be considered more consistent?

SOLUTION In order to determine the consistency of teams we will have to find the coefficients of variations of two teams.

Computation of mean and standard deviation of goals scored by team A.

No. of goals scored x_i	*No. of matches* f_i	$f_i x_i$	$f_i x_i^2$
0	1	0	0
1	9	9	9
2	7	14	28
3	5	15	45
4	3	12	48
	$\Sigma f_i = 25$	$\Sigma f_i x_i = 50$	$\Sigma f_i x_i^2 = 130$

We have,

$$N = \Sigma f_i = 25, \ \Sigma f_i x_i = 50 \text{ and } \Sigma f_i x_i^2 = 130$$

$\therefore \quad \overline{X}_A = \frac{1}{N}(\Sigma f_i x_i) = \frac{50}{25} = 2$

and, $\quad \sigma_A^2 = \left(\frac{1}{N}\Sigma f_i x_i^2\right) - \left(\frac{1}{N}\Sigma f_i x_i\right)^2 = \frac{130}{25} - \left(\frac{50}{25}\right)^2 = 5.2 - 4 = 1.2$

$\Rightarrow \quad \sigma_A = \sqrt{1.2} = 1.095$

It is given that $\overline{X}_B = 2$ and $\sigma_B = 1.25$

Now,

Coefficient of variation in goals scored by team $A = \frac{\sigma_A}{\overline{X}_A} \times 100 = \frac{1.095}{2} \times 100 = 54.75$

Coefficient of variation of goals scored by team $B = \frac{\sigma_B}{\overline{X}_B} \times 100 = \frac{1.25}{2} \times 100 = 62.50$

We observe that the coefficient of variation of goals scored by team A is lesser than that of team B. Hence, team A is more consistent.

EXAMPLE 6 *Suppose that samples of polythene bags from two manufacturers, A and B, are tested by a prospective buyer for bursting pressure, with the following results:*

Bursting Pressure in kg	*Number of bags manufactured by manufacturer* A	B
5-10	2	9
10-15	9	11
15-20	29	18
20-25	54	32
25-30	11	27
30-35	5	13

Which set of the bags has the highest average brusting pressure? Which has more uniform pressure?

SOLUTION For determining the set of bags having higher average bursting pressure, we compute mean and for finding out set of bags having more uniform pressure we compute coefficient of variation.

Manufacturer A:

Computation of mean and standard deviation

Bursting pressure	*Mid-values* x_i	f_i	$u_i = \frac{x_i - 17.5}{5}$	$f_i u_i$	$f_i u_i^2$
5-10	7.5	2	−2	−4	8
10-15	12.5	9	−1	−9	9
15-20	17.5	29	0	0	0
20-25	22.5	54	1	54	54
25-30	27.5	11	2	22	44
30-35	32.5	5	3	15	45
		$N = \Sigma f_i = 110$	$\Sigma u_i = 3$	$\Sigma f_i u_i = 78$	$\Sigma f_i u_i^2 = 160$

$$\overline{X}_A = a + h\left(\frac{\Sigma f_i u_i}{N}\right)$$

$\Rightarrow \overline{X}_A = 17.5 + 5 \times \frac{78}{110} = 175 + 35 = 21$ $\qquad [\because h = 5, a = 175]$

$$\sigma_A^2 = h^2 \left\{\left(\frac{1}{N}\Sigma f_i u_i^2\right) - \left(\frac{1}{N}\Sigma f_i u_i\right)^2\right\} = 25\left\{\frac{160}{110} - \left(\frac{78}{110}\right)^2\right\} = 25\left(\frac{17600-6084}{110 \times 110}\right) = 23.79$$

$\Rightarrow \sigma_A = \sqrt{23.79} = 4.87$

$\therefore$ Coefficient of variation $= \frac{\sigma_A}{\overline{X}_A} \times 100 = \frac{4.87}{21} \times 100 = 23.19$

Manufacturer B:

Bursting pressure	Mid-values x_i	f_i	$u_i = \frac{x_i - 17.5}{5}$	$f_i u_i$	$f_i u_i^2$
5-10	7.5	9	−2	−18	36
10-15	12.5	11	−1	−11	11
15-20	17.5	18	0	0	0
20-25	22.5	32	1	32	32
25-30	27.5	27	2	54	108
30-35	32.5	13	3	39	117
		$N = \Sigma f_i = 110$		$\Sigma f_i u_i = 96$	$\Sigma f_i u_i^2 = 304$

Here, $N = 110$, $\Sigma f_i u_i = 96$, $a = 17.5$ and $h = 5$

$\therefore \quad \overline{X}_B = a + h\left(\frac{\Sigma f_i u_i}{N}\right) = 17.5 + 5 \times \frac{96}{110} = 17.5 + 4.36 = 21.81$

and, $\sigma_B^2 = h^2 \left\{\left(\frac{1}{N}\Sigma f_i u_i^2\right) - \left(\frac{1}{N}\Sigma f_i u_i\right)^2\right\} = 25\left\{\frac{304}{110} - \left(\frac{96}{110}\right)^2\right\} = 25\left(\frac{33440 - 9216}{110 \times 110}\right) = 50.04$

$\Rightarrow \quad \sigma_B = \sqrt{50.04} = 7.07$

$\therefore \quad$ Coefficient of variation $= \frac{\sigma_B}{\overline{X}_B} \times 100 = \frac{7.07}{21.81} \times 100 = 32.41$

We observe that the average brusting pressure is higher for manufacturer *B*. So, bags manufactured by *B* have higher bursting pressure.

The coefficient of variation is less for manufacturer *A*. So, bags manufactured by *A* have more uniform pressure.

EXERCISE 32.7

LEVEL-1

1. Two plants *A* and *B* of a factory show following results about the number of workers and the wages paid to them

	Plant A	Plant B
No. of workers	5000	6000
Average monthly wages	₹ 2500	₹ 2500
Variance of distribution of wages	81	100

In which plant *A* or *B* is there greater variability in individual wages?

2. The means and standard deviations of heights ans weights of 50 students of a class are as follows:

	Weights	Heights
Mean	63.2 kg	63.2 inch
Standard deviation	5.6 kg	11.5 inch

Which shows more variability, heights or weights?

3. Coefficient of variation of two distributions are 60% and 70% and their standard deviations are 21 and 16 respectively. What are their arithmetic means?
4. Calculate coefficient of variation from the following data:

Income (in ₹):	1000-1700	1700-2400	2400-3100	3100-3800	3800-4500	4500-5200
No. of families:	12	18	20	25	35	10

5. An analysis of the weekly wages paid to workers in two firms A and B, belonging to the same industry gives the following results:

	Firm A	Firm B
No. of wage earners	586	648
Average weekly wages	₹ 52.5	₹ 47.5
Variance of the distribution of wages	100	121

(i) Which firm A or B pays out larger amount as weekly wages?
(ii) Which firm A or B has greater variability in individual wages?

6. The following are some particulars of the distribution of weights of boys and girls in a class:

	Boys	Girls
Number	100	50
Mean weight	60 kg	45 kg
Variance	9	4

Which of the distributions is more variable?

7. The mean and standard deviation of marks obtained by 50 students of a class in three subjects, mathematics, physics and chemistry are given below:

Subject	Mathematics	Physics	Chemistry
Mean	42	32	40.9
Standard Deviation	12	15	20

Which of the three subjects shows the highest variability in marks and which shows the lowest? [NCERT]

8. From the data given below state which group is more variable G_1 or G_2?

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80
Group G_1	9	17	32	33	40	10	9
Group G_2	10	20	30	25	43	15	7

9. Find the coefficient of variation for the following data:

Size (in cms):	10-15	15-20	20-25	25-30	30-35	35-40
No. of items:	2	8	20	35	20	15

10. From the prices of shares X and Y given below: find out which is more stable in value:

X:	35	54	52	53	56	58	52	50	51	49
Y:	108	107	105	105	106	107	104	103	104	101

11. Life of bulbs produced by two factories A and B are given below:

Length of life (in hours) :	550-650	650-750	750-850	850-950	950-1050
Factory A : (Number of bulbs)	10	22	52	20	16
Factory B : (Number of bulbs)	8	60	24	16	12

The bulbs of which factory are more consistent from the point of view of length of life? **[NCERT EXEMPLAR]**

12. Following are the marks obtained, out of 100, by two students Ravi and Hashina in 10 tests :

Ravi:	25	50	45	30	70	42	36	48	35	60
Hashina:	10	70	50	20	95	55	42	60	48	80

Who is more intelligent and who is more consistent? **[NCERT EXEMPLAR]**

ANSWERS

1. Plant B 2. Heights 3. 35, 22.85 4. 3.21 5. (i) Firm B (ii) Firm B
6. Boys 7. Highest: Chemistry Lowest: Mathematics 8. G_1 9. 21.75 10. Y
11. Factory A 12. Hashina is more intelligent and consistent.

HINTS TO NCERT & SELECTED PROBLEM

1. We observe that S.D. of marks in Mathematics is least and that of Chemistry is highest. Therefore, marks in Mathematics have lowest variability and that in Chemistry have highest variability.

VERY SHORT ANSWER QUESTIONS (VSAQs)

Answer each of the following questions in one word or one sentence or as per exact requirement of the question:

1. Write the variance of first n natural numbers.
2. If the sum of the squares of deviations for 10 observations taken from their mean is 2.5, then write the value of standard deviation.
3. If $x_1, x_2, \ldots, x_n$ are n values of a variable X and $y_1, y_2, \ldots, y_n$ are n values of variable Y such that $y_i = ax_i + b, i = 1, 2, \ldots, n$, then write Var($Y$) in terms of Var(X).
4. If X and Y are two variates connected by the relation $Y = \frac{aX + b}{c}$ and Var (X) $= \sigma^2$, then write the expression for the standard deviation of Y.
5. In a series of 20 observations, 10 observations are each equal to k and each of the remaining half is equal to $-k$. If the standard deviation of the observations is 2, then write the value of k.
6. If each observation of a raw data whose standard deviation is σ is multiplied by a, then write the S.D. of the new set of observations.
7. If a variable X takes values 0, 1, 2,..., n with frequencies ${}^nC_0, {}^nC_1, {}^nC_2, \ldots, {}^nC_n$, then write variance X.

ANSWERS

1. $\frac{n^2-1}{12}$ 2. 0.5 3. Var $(Y) = a^2$ Var (X)

4. $\left|\frac{a}{c}\right| \sigma$ 5. ± 2 6. $|a| \sigma$ 7. $\frac{n}{4}$

MULTIPLE CHOICE QUESTIONS (MCQs)

1. For a frequency distribution mean deviation from mean is computed by
   (a) M.D. $= \frac{\Sigma f}{\Sigma f|d|}$ (b) M.D. $= \frac{\Sigma d}{\Sigma f}$
   (c) M.D. $= \frac{\Sigma fd}{\Sigma f}$ (d) M.D. $= \frac{\Sigma f|d|}{\Sigma f}$
2. For a frequency distribution standard deviation is computed by applying the formula
   (a) $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2}$ (b) $\sigma = \sqrt{\left(\frac{\Sigma fd}{\Sigma f}\right)^2 - \frac{\Sigma fd^2}{\Sigma f}}$
   (c) $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \frac{\Sigma fd}{\Sigma f}}$ (d) $\sqrt{\left(\frac{\Sigma fd}{\Sigma f}\right)^2 - \frac{\Sigma fd^2}{\Sigma f}}$
3. If v is the variance and σ is the standard deviation, then
   (a) $v = \frac{1}{\sigma^2}$ (b) $v = \frac{1}{\sigma}$ (c) $v = \sigma^2$ (d) $v^2 = \sigma$
4. The mean deviation from the median is
   (a) equal to that measured from another value
   (b) maximum if all observations are positive
   (c) greater than that measured from any other value.
   (d) less than that measured from any other value.
5. If $n = 10, \overline{X} = 12$ and $\Sigma x_i^2 = 1530$, then the coefficient of variation is
   (a) 36 % (b) 41 % (c) 25 % (d) none of these
6. The standard deviation of the data:

x:	1	a	a^2	...	a^n
f:	${}^nC_0$	${}^nC_1$	${}^nC_2$	...	${}^nC_n$

   is
   (a) $\left(\frac{1+a^2}{2}\right)^n - \left(\frac{1+a}{2}\right)^n$ (b) $\left(\frac{1+a^2}{2}\right)^{2n} - \left(\frac{1+a}{2}\right)^n$
   (c) $\left(\frac{1+a}{2}\right)^{2n} - \left(\frac{1+a^2}{2}\right)^n$ (d) none of these
7. The mean deviation of the series $a, a+d, a+2d, ..., a+2n$ from its mean is
   (a) $\frac{(n+1)d}{2n+1}$ (b) $\frac{nd}{2n+1}$ (c) $\frac{n(n+1)d}{2n+1}$ (d) $\frac{(2n+1)d}{n(n+1)}$
8. A batsman scores runs in 10 innings as 38, 70, 48, 34, 42, 55, 63, 46, 54 and 44. The mean deviation about mean is
   (a) 8.6 (b) 6.4 (c) 10.6 (d) 7.6

9. The mean deviation of the numbers 3, 4, 5, 6, 7 from the mean is
(a) 25 (b) 5 (c) 1.2 (d) 0

10. The sum of the squares deviations for 10 observations taken from their mean 50 is 250. The coefficient of variation is
(a) 10 % (b) 40 % (c) 50 % (d) none of these

11. Let $x_1, x_2, ..., x_n$ be values taken by a variable X and $y_1, y_2, ..., y_n$ be the values taken by a variable Y such that $y_i = ax_i + b, i = 1, 2,..., n$. Then,
(a) $\text{Var}(Y) = a^2 \text{Var}(X)$ (b) $\text{Var}(X) = a^2 \text{Var}(Y)$
(c) $\text{Var}(X) = \text{Var}(X) + b$ (d) none of these

12. If the standard deviation of a variable X is σ, then the standard deviation of variable $\frac{aX+b}{c}$ is
(a) $a\sigma$ (b) $\frac{a}{c}\sigma$ (c) $\left|\frac{a}{c}\right|\sigma$ (d) $\frac{a\sigma+b}{c}$

13. If the S.D. of a set of observations is 8 and if each observation is divided by -2, the S.D. of the new set of observations will be
(a) -4 (b) -8 (c) 8 (d) 4

14. If two variates X and Y are connected by the relation $Y = \frac{aX+b}{c}$, where a, b, c are constants such that $ac < 0$, then
(a) $\sigma_Y = \frac{a}{c}\sigma_X$ (b) $\sigma_Y = -\frac{a}{c}\sigma_X$ (c) $\sigma_Y = \frac{a}{c}\sigma_X + b$ (d) none of these

15. If for a sample of size 60, we have the following information $\Sigma x_i^2 = 18000$ and $\Sigma x_i = 960$, then the variance is
(a) 6.63 (b) 16 (c) 22 (d) 44

16. Let a, b, c, d, e be the observations with mean m and standard deviation s. The standard deviation of the observations $a+k, b+k, c+k, d+k, e+k$ is
(a) s (b) ks (c) $s+k$ (d) $\frac{s}{k}$

17. The standard deviation of first 10 natural numbers is
(a) 5.5 (b) 3.87 (c) 2.97 (d) 2.87

18. Consider the first 10 positive integers. If we multiply each number by -1 and then add 1 to each number, the variance of the numbers so obtained is
(a) 8.25 (b) 6.5 (c) 3.87 (d) 2.87

19. Consider the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. If 1 is added to each number, the variance of the numbers so obtained is
(a) 6.5 (b) 2.87 (c) 3.87 (d) 8.25

20. The mean of 100 observations is 50 and their standard deviation is 5. The sum of all squares of all the observations is
(a) 50,000 (b) 250,000 (c) 252500 (d) 255000

21. Let $x_1, x_2, ..., x_n$ be n observations. Let $y_i = ax_i + b$ for $i = 1, 2, ..., n$, where a and b are constants. If the mean of $x_i^{'s}$ is 48 and their standard deviation is 12, the mean of $y_i^{'s}$ is 55 and standard deviation of $y_i^{'s}$ is 15, the values of a and b are
(a) $a = 1.25, b = -5$ (b) $a = -1.25, b = 5$ (c) $a = 2.5, b = -5$ (d) $a = 2.5, b = 5$

22. The mean deviation of the data 3, 10, 10, 4, 7, 10, 5 from the mean is
(a) 2 (b) 2.57 (c) 3 (d) 3.57

23. The mean deviation for n observations $x_1, x_2, \ldots, x_n$ from their mean $\overline{X}$ is given by

(a) $\sum_{i=1}^{n} (x_i - \overline{X})$ (b) $\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{X})$ (c) $\sum_{i=1}^{n} (x_i - \overline{X})^2$ (d) $\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{X})^2$

24. Let $x_1, x_2, \ldots, x_n$ be n observations and $\overline{X}$ be their arithmetic mean. The standard deviation is given by

(a) $\sum_{i=1}^{n} (x_i - \overline{X})^2$ (b) $\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{X})^2$ (c) $\sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{X})^2}$ (d) $\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{X}^2}$

25. The standard deviation of the observations 6, 5, 9, 13, 12, 8, 10 is

(a) 6 (b) $\sqrt{6}$ (c) $\frac{52}{7}$ (d) $\sqrt{\frac{52}{7}}$

ANSWERS

1. (d)	2. (a)	3. (c)	4. (d)	5. (c)	6. (a)	7. (c)	8. (a)
9. (c)	10. (a)	11. (a)	12. (c)	13. (d)	14. (b)	15. (d)	16. (a)
17. (d)	18. (b)	19. (c)	20. (a)	21. (a)	22. (b)	23. (b)	24. (c)
25. (d)							

SUMMARY

1. Dispersion means scatternedness around the central value.
2. Following are the measures of dispersion:
   (i) Range (ii) Quartile diviation (iii) Mean deviation (iv) Standard deviation
3. Range is the difference between the greatest and the least values of the variable.
4. Mean deviation is the arithmetic mean of the absolute values of deviations about some point (mean or median or mode).
   (i) For individual observation, we have
   $$\text{M.D.} = \frac{1}{n}\sum_{i=1}^{n} |x_i - a|, \text{ where } a = \text{mean, median, mode}$$
   $$\text{Also, M.D.} = a + h\left\{\frac{1}{N}\sum_{i=1}^{n} |u_i|\right\}, \text{ where } u_i = \frac{x_i - a}{h}$$
   (ii) For a discrete frequency distribution, we have
   $$\text{M.D.} = \frac{1}{N}\sum_{i=1}^{n} f_i |x_i - a|, a = \text{mean, median, mode}$$
   $$\text{M.D.} = a + h\left\{\frac{1}{N}\sum_{i=1}^{n} f_i u_i\right\}, \text{ where } u_i = \frac{x_i - a}{h}$$
5. Standard deviation is the positive square root of variance.
6. Variance is the arithmetic mean of the squares of deviations about mean $\overline{X}$.
   (i) For individual observations, we have
   $$\text{Variance (X)} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{X})^2$$
   $$\text{Also, Var (X)} = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$$
   $$\text{and, Var (X)} = h^2\left\{\left(\frac{1}{n}\sum_{i=1}^{n} u_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} u_i\right)^2\right\}, \text{ where } u_i = \frac{x_i - a}{h}$$

(ii) For a discrete frequency distribution, we have

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^{n} f_i \, (x_i - \overline{X})^2$$

Also, $$\text{Var}(X) = \left(\frac{1}{N} \sum_{i=1}^{n} f_i \, x_i^{\,2} \right) - \left(\frac{1}{N} \sum_{i=1}^{n} f_i \, x_i \right)^2$$

and, $$\text{Var}(X) = h^2 \left\{ \left(\frac{1}{N} \sum_{i=1}^{n} f_i \, u_i^{\,2} \right) - \left(\frac{1}{N} \sum_{i=1}^{n} f_i \, u_i \right)^2 \right\}$$

7. In order to compare two or more frequency distributions we compare their coefficients of variations. The coefficient of variation is defined as

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

8. The distribution having greater coefficient of variation has more variability around the central value than the distribution having smaller value of the coefficient of variation.