15

STATISTICS

15.1 DESCRIBING THE DISPERSION

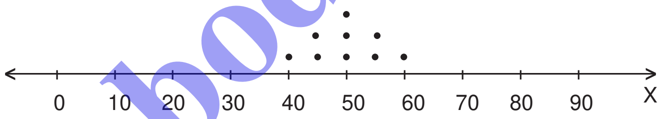
You have already studied the locational statistics — which give us a sort of central value around which the values of the variable are located. These measures of central tendency give a rough idea of where data points are centred. However, in order to interpret the data better, we should also know *how far these values spread around the central value.*

Consider an example of a cricket team and its batting performance in last 9 one day mathces. Let us assume that batsman A scored 60, 55, 50, 50, 40, 45, 55, 45, 50; batsman B scored 70, 30, 60, 20, 50, 90, 40, 80, 10; batsman C scored 20, 25, 15, 25, 15, 20, 20 in seven innings and did not get to bat in 2 innings. The mean and median are :

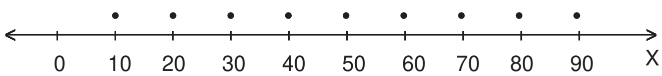
	A	*B*	*C*
Mean	50	50	20
Median	50	50	20

This tells us that average performace of batsman A and B is the same, and much better than that of C. But this is only part of the story. Now let us plot these scores as dots on a number line.

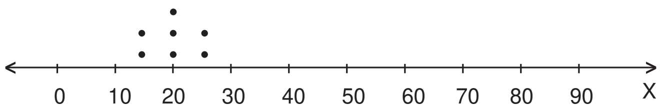
For batsman A :



For batsman B :



For bataman C :



These diagrams show that average score of A and B is the same, the score distribution of batsman A is more 'compact' than that of B whose score distribution is comparatively *dispersed (scattered)* widely. When A goes out to bat, you are quite sure that he will score around a half century, but when batsman B goes out to bat, you keep your fingers crossed. You don't know whether he will be out for a duck or whether he will hit a century. Though his average is also half a century. As far as batsman C is concerned, though his average is low, he is quite *reliable* when he goes out to bat, you are pretty certain that he will score "around" 20 runs.

This leads us to second set of descriptives—**measures of variability** or **spread of distribution**.

Definition. ***Dispersion*** *indicates the extent to which the individual measures differ from an average.*

15.2 DIFFERENT METHODS OF MEASURING DISPERSION

There are many ways in which the dispersion *i.e.* the spread of the data can be measured. These include :

(*i*) Range (*ii*) Quartile deviation

(*iii*) Mean deviation (*iv*) Standard deviation

In this chapter, we shall study the range, mean deviation about the mean, mean deviation about the median, and standard deviation.

15.3 RANGE

The difference between the largest value and smallest value of a data series is called its **range**. Thus the range indicates the total spread of the data *i.e.* 100 %. values lie within the range.

In the example given in section 15.1 :

For batsman A, range = maximum value – minimum value = 60 – 40 = 20,

For batsman B, range = 90 – 10 = 80,

For batsman C, range = 25 – 15 = 10.

For grouped data, range is defined as the difference between the upper boundary of the highest class and lower boundary of the lowest class.

15.4 MEAN DEVIATION

15.4.1 Mean deviation for ungrouped data

Recall that measures of central tendency lie between the maximum and miunimum values of a set of observations. Thus, if we consider the deviations $x_i - a$ from such a measure a, then some of the values of deviations will be positive and some negative. In particular, we know that $\Sigma (x_i - \bar{x}) = 0$. Thus, if we want to measure how far the data is scattered around the mean, we may take absolute values of deviations $x_i - \bar{x}$ and then find their average.

Thus, mean deviation about the mean $(\bar{x})$ is

$$\text{M.D. } (\bar{x}) = \frac{\Sigma\, |x_i - \bar{x}|}{n}, \text{ where } n \text{ is number of observations.}$$

Similarly, mean deviation about the median (M) is

$$\text{M.D. (M)} = \frac{\Sigma\, |x_i - \text{M}|}{n}, \text{ where } n \text{ is number of observations.}$$

Thus, in the example taken in section 15.1, for batsman A the deviation around the mean (50) for 9 values are 10, 5, 0, 0, –10, –5, 5, –5, 0. The absolute values of these deviations are 10, 5, 0, 0, 10, 5, 5, 5, 0, whose sum is 40.

Hence, mean deviation about the mean $= \frac{40}{9} = 4.4$ (approx)

Similarly, for batsman B, the deviations around the mean (50) are 20, –20, 10, –30, 0, 40, –10, 30, –40; absolute deviations are 20, 20, 10, 30, 0, 40, 10, 30, 40 ; their sum is 200. Hence mean deviation about the mean $= \frac{200}{9} = 22.2$ (approx). We immediately notice that dispersion value for batsman B is much higher than that of A.

ILLUSTRATIVE EXAMPLES

Example 1. *In a test with maximum score 25, eleven students scored 3, 9, 5, 3, 12, 10, 17, 4, 7, 19, 21 marks respectively. Calculate the*

(i) range *(ii) mean deviation about the mean*

(iii) mean deviation about the median.

Solution. Here $n = 11$. The marks can be arranged in ascending order as

3, 3, 4, 5, 7, 9, 10, 12, 17, 19, 21.

(*i*) Range = maximum value – minimum value = 21 – 3 = 18.

(*ii*) Mean $= \dfrac{\text{sum of values}}{\text{number of observations}}$

$$= \frac{3+3+4+5+7+9+10+12+17+19+21}{11}$$

$$= \frac{110}{11} = 10.$$

The deviations of marks from mean (10) are –7, –7, –6, –5, –3, –1, 0, 2, 7, 9, 11. Absolute values of deviations are 7, 7, 6, 5, 3, 1, 0, 2, 7, 9, 11

∴ Mean deviation about the mean,

$$\text{M.D. } (\bar{x}) = \frac{\Sigma\,|x_i - \bar{x}|}{n}$$

$$= \frac{7+7+6+5+3+1+0+2+7+9+11}{11}$$

$$= \frac{58}{11} = 5.27 \text{ (approx)}$$

(*iii*) Here, the data values in ascending order are

3, 3, 4, 5, 7, 9, 10, 12, 17, 19, 21.

Number of observations, $n = 11$, which is odd.

∴ Median (M) $= \dfrac{n+1}{2}$th observation = 6th observation, which is 9.

The deviations of marks from median (9) are –6, –6, –5, –4, –2, 0, 1, 3, 8, 10, 12. Absolute values of deviations are 6, 6, 5, 4, 2, 0, 1, 3, 8, 10, 12.

The required mean deviation about the median M is

$$\text{M.D. (M)} = \frac{\Sigma\,|x_i - \text{M}|}{n} = \frac{6+6+5+4+2+0+1+3+8+10+12}{11}$$

$$= \frac{57}{11} = 5.18 \text{ (approx.)}$$

Example 2. *The mean of 2, 7, 4, 6, 8 and p is 7. Find the mean deviation about the median of these observations.*

Solution. Here the oberservations are 2, 7, 4, 6, 8 and p, which are 6 in number *i.e.* $n = 6$.

Given, the mean of these observations is 7

$$\Rightarrow \quad \frac{2+7+4+6+8+p}{6} = 7$$

$$\Rightarrow \quad 27 + p = 42 \;\Rightarrow p = 15.$$

Arranging the observations in ascending order, we get

2, 4, 6, 7, 8, 15.

$$\therefore \quad \text{Median (M)} = \frac{\frac{n}{2}\text{th observation} + \left(\frac{n}{2}+1\right)\text{th observation}}{2}$$

$$= \frac{\text{3rd observation + 4th observation}}{2}$$

$$= \frac{6+7}{2} = \frac{13}{2} = 6.5$$

Calculation of mean deviation about median :

x_i	$x_i - M$	$\lvert x_i - M \rvert$
2	−4.5	4.5
4	−2.5	2.5
6	−0.5	0.5
7	0.5	0.5
8	1.5	1.5
15	8.5	8.5
Total		18

$\therefore$ Mean deviation about median $= \dfrac{18}{6} = 3$.

Example 3. *Calculate the mean deviation about the mean of first n natural numbers when n is an odd number.* (*NCERT Examplar Problems*)

Solution. First n natural numbers are 1, 2, 3, …, n. Here, n is **odd**.

$$\text{Mean } = \bar{x} = \frac{1+2+3+\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}.$$

The deviations of numbers from mean $\left(\dfrac{n+1}{2}\right)$ are

$$1-\frac{n+1}{2},\ 2-\frac{n+1}{2},\ 3-\frac{n+1}{2},\ \ldots,\ n-\frac{n+1}{2}$$

i.e. $\quad -\dfrac{n-1}{2}, -\dfrac{n-3}{2},\ \ldots, -2, -1, 0, 1, 2, \ldots, \dfrac{n-1}{2}$.

The absolute values of deviation from mean *i.e.* $\lvert x_i - \bar{x} \rvert$ are

$$\frac{n-1}{2},\ \frac{n-3}{2},\ \ldots, 2, 1, 0, 1, 2, \ldots, \frac{n-1}{2}.$$

The sum of absolute values of deviations from the mean *i.e.* $\lvert x_i - \bar{x} \rvert$

$$= 2\left(1 + 2 + 3 + \ldots \text{ to } \frac{n-1}{2} \text{ terms}\right)$$

$$= 2 \cdot \frac{\frac{n-1}{2}\left(\frac{n-1}{2}+1\right)}{2} = \frac{n-1}{2} \cdot \frac{n+1}{2} = \frac{n^2-1}{4}.$$

$\therefore$ Mean deviation about the mean $= \dfrac{\Sigma \lvert x_i - \bar{x} \rvert}{n} = \dfrac{\frac{n^2-1}{4}}{n} = \dfrac{n^2-1}{4n}$.

Example 4. *Calculate the mean deviation about the mean of first n natural numbers when n is an even number.* (*NCERT Examplar Problems*)

Solution. First n natural numbers are 1, 2, 3, …, n. Here, n is **even**.

$$\text{Mean } = \bar{x} = \frac{1+2+3+\ldots+n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}.$$

The deviations of numbers from mean $\left(\dfrac{n+1}{2}\right)$ are

$$1-\frac{n+1}{2},\ 2-\frac{n+1}{2},\ 3-\frac{n+1}{2},\ \ldots,\ n-\frac{n+1}{2}$$

i.e. $\quad -\dfrac{n-1}{2}, -\dfrac{n-3}{2},\ \ldots, -\dfrac{3}{2}, -\dfrac{1}{2}, \dfrac{1}{2}, \dfrac{3}{2}, \ldots, \dfrac{n-1}{2}$.

The absolute values of deviations from the mean *i.e.* $|x_i - \bar{x}|$ are

$$\frac{n-1}{2}, \frac{n-3}{2}, \ldots, \frac{3}{2}, \frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots, \frac{n-1}{2}.$$

The sum of absolute values of deviations from the mean *i.e.* $|x_i - \bar{x}|$

$$= 2\left[\frac{1}{2} + \frac{3}{2} + \frac{5}{2} + \ldots \text{ to } \frac{n}{2} \text{ terms}\right]$$

$$= 2 \cdot \frac{\frac{n}{2}}{2}\left[2 \times \frac{1}{2} + \left(\frac{n}{2} - 1\right) \times 1\right] \qquad \text{(Sum of A.P.)}$$

$$= \frac{n}{2}\left(1 + \frac{n}{2} - 1\right) = \frac{n}{2} \cdot \frac{n}{2} = \frac{n^2}{4}.$$

$\therefore$ Mean deviation about the mean $= \dfrac{\Sigma |x_i - \bar{x}|}{n} = \dfrac{\frac{n^2}{4}}{n} = \dfrac{n}{4}.$

15.4.2 Mean deviation for grouped data

Data can be grouped in two ways :

(*i*) Discrete frequency distribution

(*ii*) Continuous frequency distribution

Let us discuss the mean deviation in these distributions one by one.

15.4.3 Discrete frequency distribution

Let the given data consist of m distinct values $x_1, x_2, \ldots, x_m$ with frequencies $f_1, f_2, \ldots, f_m$ respectively. This data can be represented in tabular form as given below and is called discrete frequency distribution.

x	x_1	x_2	...	x_m
f	f_1	f_2	...	f_m

15.4.4 Mean deviation about the mean

The mean $\bar{x}$ of the above data is calculated as

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1}{n}\Sigma f_i x_i, \text{ where } n = \Sigma f_i.$$

To find the mean deviation about the mean ($\bar{x}$), first we find the absolute values of deviations $x_i - \bar{x}$, and then use the formula

$$\text{M.D. } (\bar{x}) = \frac{\Sigma f_i |x_i - \bar{x}|}{\Sigma f_i} = \frac{1}{n}\Sigma f_i |x_i - \bar{x}|.$$

15.4.5 Mean deviation about the median

Recall that to calculate the median of a discrete frequency distribution, first we arrange the observations $x_1, x_2, \ldots, x_m$ in ascending order, and then obtain cumulative frequencies. If $\Sigma f_i = n$, then

$$\text{median} = \begin{cases} \dfrac{n+1}{2} \text{th observation, if } n \text{ is odd} \\ \dfrac{\dfrac{n}{2} \text{th observation} + \left(\dfrac{n}{2} + 1\right) \text{th observation}}{2}, \text{ if } n \text{ is even}. \end{cases}$$

After finding median (M), we obtain the mean deviation from the median by using the formula

$$\text{M.D. (M)} = \frac{\Sigma f_i |x_i - \text{M}|}{\Sigma f_i} = \frac{1}{n}\Sigma f_i |x_i - \text{M}|.$$

ILLUSTRATIVE EXAMPLES

Example 1. *Find the mean deviation about the mean for the following data :*

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

(NCERT)

Solution. To calculate mean, we require $f_i x_i$ values; then to find mean deviation, we will require $|x_i - \bar{x}|$ values and $f_i|x_i - \bar{x}|$ values. Hence we make the following table. (Note that 4th column is added after calculating $\bar{x}$, then 5th column is added).

x_i	f_i	$f_i x_i$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$
10	4	40	40	160
30	24	720	20	480
50	28	1400	0	0
70	16	1120	20	320
90	8	720	40	320
	80	4000		1280

Here $\quad n = \Sigma f_i = 80$ and $\Sigma f_i x_i = 4000$

$$\therefore \quad \bar{x} = \frac{\Sigma f_i x_i}{n} = \frac{4000}{80} = 50.$$

Mean deviation about the mean,

$$\text{M.D.}(\bar{x}) = \frac{\Sigma f_i\,|x_i - \bar{x}|}{n} = \frac{1280}{80} = 16.$$

Example 2. *Find the mean deviation about the median for the following data :*

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

(NCERT)

Solution. The given values are already in ascending order. We construct the following table to calculate the median

x_i	5	7	9	10	12	15
f_i	8	6	2	2	2	6
c.f.	8	14	16	18	20	26

Here $n = 26$, which is even. So median is the average of 13th and 14th item, both of which lie in the cumulative frequency 14 for which the corresponding observation is 7.

$$\therefore \text{ Median M} = \frac{\text{13th observation} + \text{14th observation}}{2} = \frac{7+7}{2} = 7.$$

Now we calculate the M.D. (M) by constructing the following table :

$\lvert x_i - \text{M}\rvert$	2	0	2	3	5	8
f_i	8	6	2	2	2	6
$f_i\lvert x_i - \text{M}\rvert$	16	0	4	6	10	48

Here $\quad n = \Sigma f_i = 26,\ \Sigma f_i|x_i - \text{M}| = 84$

$$\therefore \quad \text{M.D. (M)} = \frac{1}{n}\Sigma f_i|x_i - \text{M}| = \frac{84}{26} = \frac{42}{13} = 3.23 \text{ (approx)}$$

15.4.6 Continuous frequency distribution

In a continuous frequency distribution, the data is classified into different class intervals without gaps along with their respective frequencies. To calculate the mean deviation, first we calculate the mean or median as the case may be. Then to calculate deviations, we assume that the frequency in each class is centred at its mid point. We take this mid point's deviation from mean or median, and proceed as in discrete frequency distribution to calculate M.D. ($\bar{x}$) or M.D. (M). This will be clear from the following illustrative examples.

Note : If the classes are discontinuous, then first convert them into continuous classes.

ILLUSTRATIVE EXAMPLES

Example 1. *Calculate the mean deviation about the mean for the following data.*

Income per day	*0 – 100*	*100 – 200*	*200 – 300*	*300 – 400*	*400 – 500*	*500 – 600*	*600 – 700*	*700 – 800*
Number of persons	*4*	*8*	*9*	*10*	*7*	*5*	*4*	*3*

(NCERT)

Solution. We construct the following table. (5th and 6th columns are filled after calculating the mean.)

Income per day	*Number of persons* f_i	*Mid points* x_i	$f_i x_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0 – 100	4	50	200	308	1232
100 – 200	8	150	1200	208	1664
200 – 300	9	250	2250	108	972
300 – 400	10	350	3500	8	80
400 – 500	7	450	3150	92	644
500 – 600	5	550	2750	192	960
600 – 700	4	650	2600	292	1168
700 – 800	3	750	2250	392	1176
	50		17900		7896

Here $n = \Sigma f_i = 50,\ \Sigma f_i x_i = 17900$

$$\therefore \quad \text{Mean} = \frac{1}{n}\Sigma f_i x_i = \frac{17900}{50} = 358$$

$$\text{M.D. }(\bar{x}) = \frac{1}{n}\Sigma f_i \lvert x_i - \bar{x} \rvert = \frac{7896}{50} = 157.92.$$

REMARK

If the classes are of uniform size, say c, then calculation of mean can be further simplified by using *step deviation method*. Here, we take

$u_i = \dfrac{x_i - a}{c}$, where a is the assumed mean and then using the formula

$$\bar{x} = a + \frac{\Sigma f_i u_i}{n} \times c.$$

In above example, *mean* would be easier to calculate by taking *assumed mean* $a = 350$ and $c = 100$.

Income	Number of persons f_i	Mid points x_i	$u_i = \frac{x_i - 350}{100}$	$f_i u_i$
0 – 100	4	50	– 3	– 12
100 – 200	8	150	– 2	– 16
200 – 300	9	250	– 1	– 9
300 – 400	10	350	0	0
400 – 500	7	450	1	7
500 – 600	5	550	2	10
600 – 700	4	650	3	12
700 – 800	3	750	4	12
	50			4

Here $\bar{x} = a + \frac{\Sigma f_i u_i}{n} \times c = 350 + \frac{4}{50} \times 100 = 358.$

Example 2. *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

(*NCERT Examplar Problems*)

Solution. Taking assumed mean $a = 10$ and class size $c = 4$, we construct the following table. Note that 6th and 7th columns are filled in after calculating the mean of the given distribution.

Class interval	Frequency f_i	Mid-points x_i	$u_i = \frac{x_i - 10}{4}$	$f_i u_i$	$\lvert x_i - \bar{x} \rvert$	$f_i \lvert x_i - \bar{x} \rvert$
0 – 4	4	2	– 2	– 8	7.2	28.8
4 – 8	6	6	– 1	– 6	3.2	19.2
8 – 12	8	10	0	0	0.8	6.4
12 – 16	5	14	1	5	4.8	24.0
16 – 20	2	18	2	4	8.8	17.6
Total	**25**			**– 5**		**96.0**

$$\therefore \quad \text{Mean} = a + \frac{\Sigma f_i u_i}{\Sigma f_i} \times c = 10 + \frac{-5}{25} \times 4 = 10 - 0.8 = 9.2$$

$$\text{Mean deviation about the mean} = \frac{\Sigma f_i \lvert x_i - \bar{x} \rvert}{\Sigma f_i} = \frac{96.0}{25} = 3.84$$

Example 3. *Find the mean deviation about the median for the following data :*

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Number of girls	8	10	10	16	4	2

Solution. We construct the following table. Note that 5th and 6th columns are filled in after calculating the median.

Marks	*Number of girls* f_i	*Cumulative Frequency (c.f.)*	*Mid-points* x_i	$\lvert x_i - M\rvert$	$f_i\lvert x_i - M\rvert$
0 – 10	8	8	5	22	176
10 – 20	10	18	15	12	120
20 – 30	10	28	25	2	20
30 – 40	16	44	35	8	128
40 – 50	4	48	45	18	72
50 – 60	2	50	55	28	56
	50				572

The class containing $\frac{n}{2}$th or 25th item is 20 – 30, which is the median class.

$$\text{Median} = l + \frac{\frac{n}{2} - C}{f} \times c = 20 + \frac{25 - 18}{10} \times 10 = 27$$

$$\text{M.D. (M)} = \frac{1}{n}\Sigma f_i\lvert x_i - \text{M}\rvert = \frac{572}{50} = 11.44.$$

Example 4. *Calculate the mean deviation about the median 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*

(NCERT)

Solution. The given frequency distribution is discontinuous, to convert it into continuous frequency distribution, adjustment factor = $\frac{21 - 20}{2}$ = 0.5.

So, we subtract 0.5 from the lower limit and add 0.5 to the upper limit of each class.

We construct the following table. Note that 5th and 6th columns are filled in after calculating the median.

Age	*Number of persons* f_i	*c.f.*	*Mid-points* x_i	$\lvert x_i - M\rvert$	$f_i\lvert x_i - M\rvert$
15.5 – 20.5	5	5	18	20	100
20.5 – 25.5	6	11	23	15	90
25.5 – 30.5	12	23	28	10	120
30.5 – 35.5	14	37	33	5	70
35.5 – 40.5	26	63	38	0	0
40.5 – 45.5	12	75	43	5	60
45.5 – 50.5	16	91	48	10	160
50.5 – 55.5	9	100	53	15	135
	100				735

The class containing $\frac{n}{2}$th or 50th observation is 35.5 – 40.5, which is the median class.

$$\text{Median} = l + \frac{\frac{n}{2} - C}{f} \times c = 35.5 + \frac{50 - 37}{26} \times 5 = 35.5 + 2.5 = 38$$

$$\therefore \text{ Mean deviation about median} = \frac{1}{n}\Sigma f_i\lvert x_i - \text{M}\rvert = \frac{735}{100}$$

$$= 7.35.$$

15.4.7 Limitations of mean deviation

(*i*) Mean deviation from the median is not fully reliable measure of dispersion where there is a high degree of variability, as the median is not a representative central tendency measure in such a case.

(*ii*) Also, as absolute value of deviations are taken, further algebraic treatment is not possible.

(*iii*) The sum of absolute deviations from the mean is more than the sum of the absolute deviations from the median. In some cases, this is not a reliable measure.

EXERCISE 15.1

Very short answer type questions (1 to 5) :

1. Find the mean deviation about the mean of the following data :
3, 6, 11, 12, 18.

2. Find the mean deviation about the median of the following data :
3, 6, 11, 12, 18.

3. Find the mean deviation about the mean of the following data :
1, 3, 7, 9, 10, 12.

4. Find the mean deviation about the median of the following data :
2, 7, 9, 11, 15, 16.

5. Find the mean deviation of the following data :
1, 2, 3, 4, 5, 6, 7
(*i*) about the mean (*ii*) about the median.

6. Find the range, mean deviation about the mean, as well as median for the following series :
(*i*) 6, 7, 10, 12, 13, 4, 8, 12 (*NCERT*)
(*ii*) 12, 3, 18, 17, 4, 9, 17, 19, 20, 15, 8, 17, 2, 3, 16, 11, 3, 1, 0, 5 (*NCERT*)
(*iii*) 20, 28, 40, 12, 30, 15, 50

7. Find the mean deviation about the mean for the following data :
(*i*) 4, 7, 8, 9, 10, 12, 13, 17 (*NCERT*)
(*ii*) 38, 70, 48, 40, 42, 55, 63, 46, 54, 44 (*NCERT*)
(*iii*) 11, 13, 4, 7, 8, 6, 15, 14, 3, 19

8. Find the mean deviation about the median for the following data :
(*i*) 13, 17, 16, 14, 11, 13, 10, 16, 11, 18, 12, 17 (*NCERT*)
(*ii*) 36, 72, 46, 42, 60, 45, 53, 46, 51, 49 (*NCERT*)
(*iii*) 3, 9, 5, 3, 12, 10, 18, 4, 7, 19, 21 (*NCERT*)

9. Find the mean deviation about the mean for the following data :

(*i*)

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

(*NCERT*)

(*ii*)

Size (x_i)	1	3	5	7	9	11	13	15
Frequency (f_i)	3	3	4	14	7	4	3	4

(*NCERT Examplar Problems*)

(*iii*)

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

(*NCERT Examplar Problems*)

10. Find the mean deviation about the median for the following data :

(*i*)

x_i	3	6	9	12	13	15	21	22
f_i	3	4	5	2	4	5	4	3

(*NCERT*)

(*ii*)

x_i	15	21	27	30	35
f_i	3	5	6	7	8

(*NCERT*)

(*iii*)

Marks obtained	10	11	12	14	15
Number of students	2	3	8	3	4

(*NCERT Examplar Problems*)

11. Find the mean deviation about the mean for the following data :

(*i*)

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*)

(*ii*)

Height in cm	95 – 105	105 – 115	115 – 125	125 – 135	135 – 145	145 – 155
Number of boys	9	13	26	30	12	10

(*NCERT*)

12. Find the mean deviation about the median for the following data :

(*i*)

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

(*NCERT Examplar Problems*)

(*ii*)

Class	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Frequency	6	7	15	16	4	2

(*NCERT*)

(*iii*)

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
Number of girls	6	8	14	16	4	2

(*NCERT*)

15.5 VARIANCE AND STANDARD DEVIATION

In the previous section, you studied the mean deviation from the mean :

$$\text{M.D.} = \frac{\Sigma |x_i - \bar{x}|}{n},$$

where we removed negative signs from the deviations by using the absolute value (modulus).

Another way of removing the negative signs is by squaring the deviations and then finding the average.

The mean of squared deviations is called ***variance,*** *denoted var(x)* or σ^2.

Thus, for ungrouped data :

$$\textbf{Var}(x) \textbf{ or } \sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \textbf{ or simply } \frac{\Sigma (x_i - \bar{x})^2}{n},$$

where $\bar{x}$ is the mean of the given data and n is the total number of observations. Observe that

$$\sigma^2 = \frac{\Sigma (x_i - \bar{x})^2}{n} = \frac{\Sigma (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n}$$

$$= \frac{\Sigma x_i^2}{n} - 2\bar{x}\left(\frac{\Sigma x_i}{n}\right) + n\frac{\bar{x}^2}{n}$$

$$= \frac{\Sigma x_i^2}{n} - \bar{x}^2. \qquad \left(\because \frac{\Sigma x_i}{n} = \bar{x}\right)$$

Thus, we can use the following **short cut method** to calculate variance

$$\boldsymbol{\sigma^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2, \text{ where } \bar{x} = \frac{\Sigma x_i}{n}.}$$

Sometimes, this considerably reduces the calculation work.

A large value of variance indicates greater dispersion of values around the central value (mean), while a smaller value indicates lesser spread. One defect with variance is that its units are different from units of variable x. Thus, when we are talking about heights of students in centimetres, variance is in *square centimetres.* It is difficult to visualise the spread of heights in terms of square centimetres. This leads us to define **standard deviation (S.D.)** or **root mean squared deviation,** which is the square root of variance. It is denoted by σ.

Hence, for ungrouped data :

$$\text{Standard deviation } \sigma = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n}}.$$

$$\text{It can also be taken as } \sigma = \sqrt{\frac{\Sigma x_i^2}{n} - \bar{x}^2}.$$

When values of $(x_i - \bar{x})$ are small, we use first formula; when values of x_i are convenient, we use second formula.

Variance and standard deviation for grouped data

If the variates (observations) $x_1, x_2, \ldots, x_m$ have frequencies $f_1, f_2, \ldots, f_m$ respectively, then the variance is defined as :

$$\text{Var }(x) \text{ or } \sigma^2 = \frac{\sum_{i=1}^{m} f_i (x_i - \bar{x})^2}{\sum_{i=1}^{m} f_i}, \text{ where } \bar{x} = \frac{\sum_{i=1}^{m} f_i x_i}{\sum_{i=1}^{m} f_i}.$$

It is convenient to remember it as

$$\sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}, \text{ where } \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}.$$

It can be shown that this is equivalent to

$$\sigma^2 = \frac{\Sigma f_i x_i^2}{\Sigma f_i} - \bar{x}^2, \text{ where } \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i}.$$

Correspondingly, the standard deviation is

$$\boldsymbol{\sigma = \sqrt{\frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i}} = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \bar{x}^2} = \sqrt{\frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2}.}$$

It can be written as

$$\sigma = \frac{1}{n}\sqrt{n\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2}.$$

Use whatever form is convenient in a given situation.

If the data is *continuous* and each class in the frequency table consists not of single value but an interval, then we use the *mid-point* of each class in the above formulae to get *approximate* values of variance and standard deviation.

Deviation method

To reduce the calculations further, we can use an assumed mean A, and let d_i be the deviation of x_i from A *i.e.* $d_i = x_i - A$. Then

$$x_i = d_i + A \quad \Rightarrow \Sigma f_i x_i = \Sigma f_i (d_i + A) = \Sigma f_i d_i + A\Sigma f_i$$

$$\Rightarrow \quad \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{\Sigma f_i d_i}{\Sigma f_i} + A \quad \Rightarrow \bar{x} - A = \frac{\Sigma f_i d_i}{\Sigma f_i}.$$

Now $x_i - A = (x_i - \bar{x}) + (\bar{x} - A)$

$$\Rightarrow \quad (x_i - A)^2 = (x_i - \bar{x})^2 + (\bar{x} - A)^2 + 2(x_i - \bar{x})(\bar{x} - A)$$

$$\Rightarrow \quad \Sigma f_i (x_i - A)^2 = \Sigma f_i (x_i - \bar{x})^2 + (\bar{x} - A)^2 \Sigma f_i + 2(\bar{x} - A) \Sigma f_i (x_i - \bar{x})$$

(Note that $\bar{x} - A$, $(\bar{x} - A)^2$ are constants and can be taken out of sigma)

$$\Rightarrow \quad \Sigma f_i d_i^2 = \Sigma f_i (x_i - \bar{x})^2 + \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2 \Sigma f_i + 0$$

$$\left(\text{As } \Sigma f_i (x_i - \bar{x}) = \Sigma f_i x_i - \Sigma f_i \bar{x} = \Sigma f_i \left(\frac{\Sigma f_i x_i}{\Sigma f_i} - \bar{x}\right) = \Sigma f_i (\bar{x} - \bar{x}) = 0\right)$$

$$\Rightarrow \quad \frac{\Sigma f_i d_i^2}{\Sigma f_i} = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i} + \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2.$$

$$\therefore \quad \boldsymbol{\sigma^2 = \frac{\Sigma f_i (x_i - \bar{x})^2}{\Sigma f_i} = \frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2}.$$

For ungrouped data, it becomes

$$\boldsymbol{\sigma^2 = \frac{\Sigma d_i^2}{n} - \left(\frac{\Sigma d_i}{n}\right)^2}.$$

Step deviation method

If in a continuous grouped data, the classes are of uniform size, say c, the calculations can be further simplified.

Putting $t_i = \frac{d_i}{c}$, we get

$$\sigma^2 = \frac{\Sigma f_i d_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i d_i}{\Sigma f_i}\right)^2 = \frac{\Sigma f_i (t_i c)^2}{\Sigma f_i} - \left(\frac{\Sigma f_i t_i c}{\Sigma f_i}\right)^2$$

$$= c^2 \left[\frac{\Sigma f_i t_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i t_i}{\Sigma f_i}\right)^2\right]$$

$$\Rightarrow \quad \boldsymbol{\sigma = c\sqrt{\frac{\Sigma f_i t_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i t_i}{\Sigma f_i}\right)^2}, \text{ where } t_i = \frac{d_i}{c} = \frac{x_i - A}{c}}.$$

ILLUSTRATIVE EXAMPLES

Example 1. *Find the mean, variance and standard deviation for the following data :*

6, 10, 7, 13, 4, 12, 8, 12 *(NCERT)*

Solution. Here the number of observations $n = 8$.

$$\text{Mean } \bar{x} = \frac{\Sigma x_i}{n} = \frac{6 + 10 + 7 + 13 + 4 + 12 + 8 + 12}{8} = \frac{72}{8}$$

$$= 9$$

The respective $x_i - \bar{x}$ are

$6 - 9,\ 10 - 9,\ 7 - 9,\ 13 - 9,\ 4 - 9,\ 12 - 9,\ 8 - 9,\ 12 - 9$

i.e. $-3,\ 1,\ -2,\ 4,\ -5,\ 3,\ -1,\ 3$

$$\therefore \quad \Sigma (x_i - \bar{x})^2 = (-3)^2 + 1^2 + (-2)^2 + 4^2 + (-5)^2 + 3^2 + (-1)^2 + 3^2$$
$$= 9 + 1 + 4 + 16 + 25 + 9 + 1 + 9 = 74$$

$$\therefore \quad \text{Variance} = \frac{\Sigma (x_i - \bar{x})^2}{n} = \frac{74}{8} = 9.25.$$

$$\text{Standard deviation} = \sqrt{9.25} = 3.04.$$

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

57, 64, 43, 67, 49, 59, 44, 47, 61, 59 (*NCERT Examplar Problems*)

Solution. Here, the number of observations = 10.

$$\text{Mean} = \bar{x} = \frac{\Sigma x_i}{n} = \frac{57 + 64 + 43 + 67 + 49 + 59 + 44 + 47 + 61 + 59}{10}$$
$$= \frac{550}{10} = 55.$$

The respective $x_i - \bar{x}$ are

$2,\ 9,\ -12,\ 12,\ -6,\ 4,\ -11,\ -8,\ 6,\ 4$

$$\therefore \quad \Sigma (x_i - \bar{x})^2 = 2^2 + 9^2 + (-12)^2 + 12^2 + (-6)^2 + 4^2 + (-11)^2 + (-8)^2 + 6^2 + 4^2$$
$$= 4 + 81 + 144 + 144 + 36 + 16 + 121 + 64 + 36 + 16$$
$$= 662.$$

$$\therefore \quad \text{Variance} = \frac{\Sigma (x_i - \bar{x})^2}{n} = \frac{662}{10} = 66.2.$$

$$\text{Standard deviation} = \sqrt{66.2} = 8.14 \text{ (approx.)}$$

Example 3. *Find the mean, standard deviation and variance of the first n natural numbers.* *(NCERT)*

Solution. The given numbers are 1, 2, 3, …, n.

$$\text{Mean } \bar{x} = \frac{\Sigma n}{n} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}.$$

$$\text{Variance } \sigma^2 = \frac{\Sigma x_i^2}{n} - \bar{x}^2 = \frac{\Sigma n^2}{n} - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{n(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{4}$$

$$= (n+1)\left[\frac{2n+1}{6} - \frac{n+1}{4}\right] = (n+1)\left(\frac{n-1}{12}\right) = \frac{n^2-1}{12}.$$

$$\therefore \quad \text{Standard deviation } \sigma = \sqrt{\frac{n^2-1}{12}}.$$

Example 4. *Find the mean, variance and standard deviation for the following data :* (NCERT)

x_i	4	8	11	17	20	24	32
f_i	3	5	9	5	4	3	1

ANSWERS

EXERCISE 15.1

1. 4.4 **2.** 4.2 **3.** $\frac{10}{3}$ **4.** 4 **5.** (*i*) $\frac{12}{7}$ (*ii*) $\frac{12}{7}$

6. (*i*) Range = 9, M.D. ($\bar{x}$) = 2.75, M.D. (M) = 2.75

(*ii*) Range = 20, M.D. ($\bar{x}$) = 6.2, M.D. (M) = 6.2

(*iii*) Range = 38, M.D. ($\bar{x}$) = $\frac{512}{49}$, M.D. (M) = $\frac{73}{7}$

7. (*i*) 3 (*ii*) 8.4 (*iii*) 4.4 **8.** (*i*) 2.33 (*ii*) 7 (*iii*) 5.27

9. (*i*) 2.3 (*ii*) 2.95 (*iii*) 1.25 **10.** (*i*) 4.97 (*ii*) 5.1 (*iii*) 1.25

11. (*i*) 10 (*ii*) 11.29 **12.** (*i*) 7 (*ii*) 10.16 (*iii*) 10.34

EXERCISE 15.2

1. 26.8; $\sqrt{26.8}$ **2.** 15; $\sqrt{15}$ **3.** $\frac{70}{3}$; $\sqrt{\frac{70}{3}}$

4. (*i*) 15, 33 (*ii*) 30, 80 (*iii*) 16.5, 74.25 **5.** 55.9; 115.89; 10.77

6. 19; 6.59 **7.** 1.38 **8.** 100; 5.39 **9.** 64; 2.86; 1.69

10. 27; 132; 11.49 **11.** 11.64 years **12.** 62; 201; 14.18 **13.** 56; 422.33; 20.55

14. 1.16 gm^2; 1.08 gm **15.** 172; 7.9 **16.** 3, $\frac{7}{3}$ **17.** 3 and 6

18. 4 and 9 **19.** 20 **20.** 42.3; 43.81

21. (*i*) 10.11; 1.997 (*ii*) 10.2; 1.99

EXERCISE 15.3

1. 36 **2.** Firm A **3.** Plant B **4.** Heights **5.** (*i*) 35 (*ii*) $\frac{160}{7}$

6. 10.47 **7.** 32.6 **8.** 31.24 **9.** 21.89

10. Y is more stable (as C.V. is less) **11.** Team A (as C.V. is lower)

12. Factory A

CHAPTER TEST

1. 1.5, 1.5 **2.** 0.75, 0.75 **3.** 1.48, 1.4 **4.** 6.32

5. 9.44 **6.** 76 cm, 18.17 cm **7.** 22.38, 2.18

8. 5.975; 2.85 **9.** 3.54, 2.23 **10.** 21.5, 164.75 **11.** 93 ; 105.58 ; 10.28

12. 18.49 **13.** (*i*) 69.16 (*ii*) 69.16 (*iii*) 69.16 (*iv*) 276.64

15. 4 and 8

16. Mathematics shows least variability and chemistry shows highest variability.

17. Yes, C.V. inceases from 12.5 to 25.

18. Series A is more consistent as C.V. is 32.73 compared to 39.58 for B.

19. (*i*) B (₹ 85000 monthly against ₹ 60000 for A)

(*ii*) B (C.V. is 1.88 against 0.75 for A).