15

Statistics is the study of collection, organisation, analysis, interpretation and presentation of data. The word 'statistics' has different meanings in different contexts. Some people regards statistics as data, facts or measurements, while other belive it to be the study of figures. Statistics deals with data collected for specific purposes. We can make decisions about the data by analysing and interpreting it.

STATISTICS

CHAPTER CHECKLIST

- Measure of Dispersion
- Variance and Standard Deviation

Data

Facts or figures, collected with a definite purpose are called data.

Data can be of two types

UNGROUPED DATA

In an ungrouped data, data is listed in series e.g. 1, 4, 5, 6, 12, 3 etc. This is also called individual data.

GROUPED DATA

It is of two types

(i) **Discrete data** In this type, data is presented in such a way that exact measurements of items are clearly show.

e.g. 15 students of class XI have secured the following marks

Marks	Frequency	Marks	Frequency
11	3	14	4
12	1	15	2
13	5		

(ii) **Continuous group data** In this type, data is arranged in groups or classes but they are not exactly measurable, they form a continuous series. e.g.

Marks obtained	Number of students
0-10	5
10-20	7
20-30	13
30-40	20

MEASURES OF CENTRAL TENDENCY

A certain value that represent the whole data and signifying its characteristics is called measure of **central tendency**. **Mean or average, median** and **mode** are the measures of central tendency.

1. Mean (Arithmetic Mean)

The arithmetic mean (or simple mean) of a set of observations is obtained by dividing the sum of the values of observations by the number of observations.

MEAN OF UNGROUPED DATA

The mean of n observations $x_1, x_2, x_3, \ldots, x_n$ is given by

$$\text{Mean } (\bar{x}) = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

n n

MEAN OF GROUPED DATA

(i) **Direct method**

Let $x_1, x_2, \ldots, x_n$ be n observations with respective frequencies $f_1, f_2, \ldots, f_n$.

Then, $$\text{Mean } (\bar{x}) = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

(ii) **Assumed mean method**

$$\text{Mean} = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$$

where, a = assumed mean

$d_i = x_i - a$ = deviation from assumed mean

(iii) **Step deviation method**

$$\text{Mean } (\bar{x}) = a + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h$$

where, a = assumed mean, $u_i = \dfrac{x_i - a}{h}$

and h = width of the class interval

2. Median

Median is defined as the middle most or the central value of the observations, when the observations are arranged either in ascending or descending order of their magnitudes.

MEDIAN OF UNGROUPED DATA

Let n be the number of observations. First, arrange the data in ascending or descending order.

Then,

(i) if n is odd,

$$\text{Median} = \text{Value of the } \left(\frac{n+1}{2}\right)\text{th observation}$$

(ii) if n is even, Median

$$= \frac{\text{Sum of values of the } \left(\frac{n}{2}\right)\text{th and } \left(\frac{n}{2}+1\right)\text{th observations}}{2}$$

MEDIAN OF GROUPED DATA

I. **For Discrete Data** First, arrange the data in ascending or descending order and find cumulative frequency. Now, find $\frac{N}{2}$, where $N = \Sigma f_i$.

After that, find the median by using the following formula.

(i) If $\Sigma f_i = N$ is even, then

(i) If $\Sigma f_i = N$ is even, then

$$\text{Median} = \frac{\text{Value of } \left(\frac{N}{2}\right)\text{th observation} + \text{Value of } \left(\frac{N}{2} + 1\right)\text{ th observation}}{2}$$

(ii) If $\Sigma f_i = N$ is odd, then

$$\text{Median} = \text{Value of } \left(\frac{N+1}{2}\right)\text{th observation.}$$

II. **For Continuous Data** First arrange the data in ascending or descending order and find the cumulative frequencies of all the classes.

Now, find $\frac{N}{2}$, where $N = \Sigma f_i$

Further, find the class interval, whose cumulative frequency is just greater than or equal to $\frac{N}{2}$.

$$\text{Then, median} = l + \frac{\left(\frac{N}{2} - cf\right)}{f} \times h$$

where, l = lower limit of median class

N = number of observations

cf = cumulative frequency of class preceding the median class

f = frequency of the median class

h = class width (assuming class size to be equal)

Note

The cumulative frequency of a class is the frequency obtained by adding the frequencies of all the classes preceding the given class and the frequency of given class

|TOPIC 1|
Measure of Dispersion

The measures of central tendency are not sufficient to give complete information about given data. Variability is another factor which is required to be studied under statistics. The single number that describe variability is called **measure of dispersion**. It is the measure of spreading (scatter) of the data about some central tendency.

The dispersion or scatter in a data is measured on the basis of the observations and the types of measure of central tendency used. *There are following measures of dispersion*

1. Range
2. Quartile deviation
3. Mean deviation
4. Standard deviation.

Note

In this chapter, 'quartile deviation' will not be discussed, it is not in syllabus.

RANGE

Range is the difference of maximum and minimum values of data.

Range = Maximum value − Minimum value

In the previous example

For Ravi, Range $= 79 - 5 = 74$

and for Ramesh, Range $= 60 - 42 = 18$

Here, we see that, Range of Ravi > Range of Ramesh.

Therefore, the scores are scattered or dispersed in case of Ravi while for Ramesh these are close to each other.

The range of data gives us a rough idea of variability or scatter but does not tell about the dispersion of the data from a measure of central tendency.

MEAN DEVIATION

Mean deviation is an important measure of dispersion, which depend upon the deviations of the observations from a central tendency. Mean deviation is defined as the arithmetic mean of the absolute deviations of all the values taken from a central value or a fixed number 'a'. The mean devation from '*a*' is denoted as MD (*a*).

Thus, mean deviation from '*a*',

$$\text{MD}(a) = \frac{\left[\text{Sum of absolute values of deviations from } a\right]}{\text{Number of observations}}$$

Note

Mean deviation may be obtained from any measure of central tendency. But in this chapter, we study deviation from mean and median.

Mean Deviation for Ungrouped Data

Let $x_1, x_2, \ldots, x_n$ be n observations. Then, mean deviation about mean or median can be determined by using following steps

***Step* I** Find the mean or median of given observations using the suitable formula.

***Step* II** Find the deviation of each observation x_i from $\bar{x}$ (mean) or M (median) and then take their absolute values i.e. $|x_i - \bar{x}|$ or $|x_i - M|$.

***Step* III** Find the sum of absolute values of deviations obtained in step III

i.e. $$\sum_{i=1}^{n} |x_i - \bar{x}| \text{ or } \sum_{i=1}^{n} |x_i - M|$$

***Step* IV** Now, find mean deviation about mean or median by using the formula $\dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$ or $\dfrac{\sum_{i=1}^{n} |x_i - M|}{n}$, where n is the number of observations.

EXAMPLE |1| Find the mean deviation about the mean for the following data. **[NCERT]**

38, 70, 48, 40, 42, 55, 63, 46, 54, 44

Sol. Given observations are

38, 70, 48, 40, 42, 55, 63, 46, 54 and 44

Here, number of observations, $n = 10$

∴ Mean,

$$\bar{x} = \frac{(38 + 70 + 48 + 40 + 42 + 55 + 63 + 46 + 54 + 44)}{10} = \frac{500}{10}$$

$= 50$

Let us make the table for deviation and absolute deviation

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
38	$38 - 50 = -12$	12
70	$70 - 50 = 20$	20
48	$48 - 50 = -2$	2
40	$40 - 50 = -10$	10
42	$42 - 50 = -8$	8
55	$55 - 50 = 5$	5
63	$63 - 50 = 13$	13
46	$46 - 50 = -4$	4
54	$54 - 50 = 4$	4
44	$44 - 50 = -6$	6
Total		$\sum_{i=1}^{10} \lvert x_i - \bar{x}\rvert = 84$

Now, $MD = \dfrac{\sum_{i=1}^{10} |x_i - \bar{x}|}{10} = \dfrac{84}{10} = 8.4$

EXAMPLE |2| Find the mean deviation from the mean for the following data

6.5, 5, 5.25, 5.5, 4.75, 4.5, 6.25, 7.75, 8.5

Sol. Given observations are

6.5, 5, 5.25, 5.5, 4.75, 4.5, 6.25, 7.75, 8.5

Here number of observation $n = 9$

Let $\bar{x}$ be the mean of given data.

Then, $\bar{x} = \dfrac{6.5 + 5 + 5.25 + 5.5 + 4.75 + 4.5 + 6.25 + 7.75 + 8.5}{9}$

$= \dfrac{54}{9} = 6$

Let us make the table for deviation and absolute deviation.

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x} \rvert$
6.5	0.5	0.50
5.0	− 1	1.00
5.25	− 0.75	0.75
5.5	− 0.5	0.50
4.75	− 1.25	1.25
4.5	− 1.50	1.50
6.25	0.25	0.25
7.75	1.75	1.75
8.5	2.5	2.50
Total		$\sum_{i=1}^{9} \lvert x_i - \bar{x} \rvert = 10.00$

∴ Mean deviation about mean,

$$MD(\bar{x}) = \frac{\sum_{i=1}^{9} |x_i - \bar{x}|}{9} = \frac{10}{9} = 1.1$$

Hence, the mean deviation about mean is 1.1.

EXAMPLE |3| Find the mean deviation about the median for the data 34, 66, 30, 38, 44, 50, 40, 60, 42, 51.

Sol. The given data can be arranged in ascending order as 30, 34, 38, 40, 42, 44, 50, 51, 60, 66. Here, total number of observations are 10. i.e. $n = 10$, which is even.

∴ Median

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

$$= \frac{\left(\frac{10}{2}\right)\text{th observation} + \left(\frac{10}{2}+1\right)\text{th observation}}{2}$$

$$= \frac{(\text{5th observation} + \text{6th observation})}{2}$$

$$= \frac{42 + 44}{2} = \frac{86}{2} = 43$$

Let us make the table for absolute deviation

x_i	$\lvert x_i - M \rvert$
30	$\lvert 30 - 43 \rvert = 13$
34	$\lvert 34 - 43 \rvert = 9$
38	$\lvert 38 - 43 \rvert = 5$
40	$\lvert 40 - 43 \rvert = 3$
42	$\lvert 42 - 43 \rvert = 1$
44	$\lvert 44 - 43 \rvert = 1$
50	$\lvert 50 - 43 \rvert = 7$
51	$\lvert 51 - 43 \rvert = 8$
60	$\lvert 60 - 43 \rvert = 17$
66	$\lvert 66 - 43 \rvert = 23$
Total	$\sum_{i=1}^{10} \lvert x_i - M \rvert = 87$

Now, mean deviation about median,

$$MD = \frac{\sum_{i=1}^{10} |x_i - M|}{10} = \frac{87}{10} = 8.7$$

EXAMPLE |4| The mean of 6, 8, 5, 7, a and 4 is 7. Find the mean deviation about median of these observation.

Sol. Here, number of observations $n = 6$

By the given condition, $\dfrac{6 + 8 + 5 + 7 + a + 4}{6} = 7$

$\Rightarrow \quad 30 + a = 6 \times 7 \Rightarrow a = 42 - 30 = 12$

Now, on arranging the observations in ascending order, we have 4, 5, 6, 7, 8, 12.

Now,

$$\text{Medain, } M = \frac{\left(\frac{6}{2}\right)\text{th observation} + \left(\frac{6}{2}+1\right)\text{th observation}}{2}$$

$$= \frac{\text{3rd observation} + \text{4th observation}}{2} = \frac{6 + 7}{2} = \frac{13}{2} = 6.5$$

Let us make the table for deviation and absolute deviation.

Let us make the table for deviation and absolute deviation.

x_i	$x_i - M$	$\lvert x_i - M \rvert$
4	−2.5	2.5
5	− 1.5	1.5
6	−0.5	0.5
7	0.5	0.5
8	1.5	1.5
12	5.5	5.5
Total		$\sum_{i=1}^{6} \lvert x_i - M \rvert = 12$

∴ Mean deviation about median $= \dfrac{\sum_{i=1}^{6} |x_i - M|}{6} = \dfrac{12}{6} = 2$

EXAMPLE |5| Calculate the mean deviation about the mean of the set of first n natural numbers, when n is an odd number.

Sol. Consider first n natural number when n is an odd, i.e. 1, 2, 3, 4, ..., n (odd).

Clearly, Mean $(\bar{x}) = \dfrac{1+2+3+...+n}{n} = \dfrac{n(n+1)}{2n} = \dfrac{n+1}{2}$

$\left[\because \text{sum of first } N \text{ Natural number} = \dfrac{N(N+1)}{2}\right]$

Now, $$MD = \frac{\left|1-\frac{n+1}{2}\right| + \left|2-\frac{n+1}{2}\right| + \left|3-\frac{n+1}{2}\right| + ... + \left|n-\frac{n+1}{2}\right|}{n}$$

$$\left[\because MD = \frac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}\right]$$

$$= \frac{\left|1-\frac{n+1}{2}\right| + \left|2-\frac{n+1}{2}\right| + ... + \left|\frac{n-1}{2}-\frac{n+1}{2}\right| + \left|\frac{n+1}{2}-\frac{n+1}{2}\right| + \left|\frac{n+3}{2}-\frac{n+1}{2}\right| + ... + \left|\frac{2n-2}{2}-\frac{n+1}{2}\right| + \left|n-\frac{n+1}{2}\right|}{n}$$

$$= \frac{\left|\frac{1-n}{2}\right| + \left|\frac{3-n}{2}\right| + ... + |-1| + 0 + |1| + |2| + ... + \left|\frac{n-3}{2}\right| + \left|\frac{n-1}{2}\right|}{n}$$

$$= \frac{2}{n}\left[\underbrace{1+2+...+\frac{n-3}{2}+\frac{n-1}{2}}_{\left(\frac{n-1}{2}\right)\text{ terms}}\right] = \frac{2}{n}\left[\frac{\left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}+1\right)}{2}\right]$$

$$\left[\because \text{sum of first } N \text{ natural numbers} = \frac{N(N+1)}{2}\right]$$

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

EXAMPLE |6| Calculate the mean deviation about the mean of the set of first n natural numbers when n is an even number.

Sol. Consider first n natural number, when n is even i.e. 1, 2, 3, 4, ..., n (even).

Clearly, Mean $(\bar{x}) = \dfrac{1+2+3+...+n}{n} = \dfrac{n(n+1)}{2n} = \dfrac{n+1}{2}$

Now, $$MD = \frac{1}{n}\left[\left|1-\frac{n+1}{2}\right| + \left|2-\frac{n+1}{2}\right| + \left|3-\frac{n+1}{2}\right| + \cdots + \left|\frac{n-2}{2}-\frac{n+1}{2}\right| + \left|\frac{n}{2}-\frac{n+1}{2}\right| + \left|\frac{n+2}{2}-\frac{n+1}{2}\right| + ... + \left|n-\frac{n+1}{2}\right|\right]$$

$$= \frac{1}{n}\left[\left|\frac{1-n}{2}\right| + \left|\frac{3-n}{2}\right| + \left|\frac{5-n}{2}\right| + ... + \left|\frac{-3}{2}\right| + \left|\frac{1}{2}\right| + ... + \left|\frac{n-1}{2}\right|\right]$$

$$= \frac{2}{n}\left[\underbrace{\frac{1}{2}+\frac{3}{2}+...+\frac{n-1}{2}}_{\left(\frac{n}{2}\right)\text{ terms}}\right]$$

$$= \frac{1}{n}\cdot\left(\frac{n}{2}\right)^2 \quad [\because \text{sum of first } N \text{ odd natural numbers} = N^2]$$

$$= \frac{1}{n}\cdot\frac{n^2}{4} = \frac{n}{4}$$

EXAMPLE |7| If $\bar{x}$ is the mean and $MD(\bar{x})$ is mean deviation from mean then find the number of observations lying between $\bar{x} - MD(\bar{x})$ and $\bar{x} + MD(\bar{x})$. Use the data 22, 24, 30, 27, 29, 31, 25, 28, 41, 42.

Sol. On arranging the given data in ascending order, we have 22, 24, 25, 27, 28, 29, 30, 31, 41, 42

Now, mean

$$(\bar{x}) = \frac{22+24+25+27+28+29+30+31+41+42}{10} = \frac{299}{10} = 29.9$$

and mean deviation from the mean, $MD(\bar{x}) = \dfrac{\sum_{i=1}^{n}|x_i - \bar{x}|}{n}$

$$= \frac{\left[\begin{array}{l}|22-29.9| + |24-29.9| + |25-29.9| + |27-29.9| \\ + |28-29.9| + |29-29.9| + |30-29.9| \\ + |31-29.9| + |41-29.9| + |42-29.9|\end{array}\right]}{10}$$

$$= \frac{7.9+5.9+4.9+2.9+1.9+0.9+0.1+1.1+11.1+12.1}{10}$$

$$= \frac{48.8}{10} = 4.88 = 4.9 \text{ (approx)}.$$

$\therefore \quad \bar{x} - MD(\bar{x}) = 29.9 - 4.9 = 25$

and $\quad \bar{x} + MD(\bar{x}) = 29.9 + 4.9 = 34.8$

On examining the arranged data, we find that observations between 25 and 34.8 are 27, 28, 29, 30, 31.

Hence, there are 5 observations lying between $\bar{x} - MD(\bar{x})$ and $\bar{x} + MD(\bar{x})$.

Mean Deviation for Grouped Data

(A) FOR DISCRETE FREQUENCY DISTRIBUTION

Let the given data have n distinct values $x_1, x_2, \ldots, x_n$ and their corresponding frequencies are $f_1, f_2, \ldots, f_n$, respectively. Then, this data can be represented in the tabular form, as

x_i	x_1	x_2	x_3	...	x_n
f_i	f_1	f_2	f_3	...	f_n

and is called discrete frequency distribution.

Here, mean deviation about mean or median is given by

$$= \frac{\sum_{i=1}^{n} f_i |x_i - A|}{N}, \text{ where, } N = \sum_{i=1}^{n} f_i = \text{total frequency}$$

and A = mean or median.

WORKING RULE FOR FINDING THE MEAN DEVIATION ABOUT MEAN

For finding the mean deviation about mean, we use the following working steps

Step I Find the mean of given observations using the formula $\bar{x} = \dfrac{\Sigma f_i x_i}{\Sigma f_i}$.

Step II Find the deviation of each observation x_i from $\bar{x}$ and take their absolute values i.e. $|x_i - \bar{x}|$ and then find $f_i|x_i - \bar{x}|$.

Step III Find the sum of absolute values of deviations obtained in step II i.e. $\Sigma f_i |x_i - \bar{x}|$.

Step IV Now, find the mean deviation about mean by using the formula, $\dfrac{\Sigma f_i |x_i - \bar{x}|}{\Sigma f_i}$,

EXAMPLE |8| Find the mean deviation about the mean for the following data.

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

Sol. Let us make the following table from the given data.

x_i	f_i	$f_i x_i$	$\|x_i - \bar{x}\|$	$f_i\|x_i - \bar{x}\|$
2	2	4	5.5	11
5	8	40	2.5	20
6	10	60	1.5	15
8	7	56	0.5	3.5
10	8	80	2.5	20
12	5	60	4.5	22.5
Total	40	300		92

Here, $N = \Sigma f_i = 40, \Sigma f_i x_i = 300$

Now, mean $(\bar{x}) = \dfrac{1}{N} \Sigma f_i x_i = \dfrac{1}{40} \times 300 = 7.5$

$\therefore$ Mean deviation about the mean,

$$\text{MD}(\bar{x}) = \frac{1}{N} \Sigma f_i |x_i - \bar{x}| = \frac{1}{40} \times 92 = 2.3$$

Hence, the mean deviation about mean is 2.3.

WORKING RULE FOR FINDING THE MEAN DEVIATION ABOUT MEDIAN

For finding the mean deviation about median, we use the following working steps

Step I Arrange the given data either in ascending order or descending order.

Step II Make a cumulative frequency table.

Step III Find the median by using the formula,

(i) If $\Sigma f_i = N$ is even, then

Median
$$= \frac{\text{Value of } \left(\frac{N}{2}\right)\text{th observation} + \text{Value of } \left(\frac{N}{2} + 1\right)\text{th observation}}{2}$$

(ii) If $\Sigma f_i = N$ is odd, then

$$\text{Median} = \text{Value of } \left(\frac{N+1}{2}\right)\text{th observation}$$

Step IV Find the absolute values of deviations of observations x_i from M.

Step V Find the product of frequency with absolute deviations i.e. $f_i |x_i - M|$.

Step VI Find the mean deviation from median by using the formula, $\text{MD} = \dfrac{\Sigma f_i |x_i - M|}{\Sigma f_i}$.

EXAMPLE |9| Find the mean deviation from the median of the following frequency distribution.

Age (in years)	10	12	15	18	21	23
Frequency	3	5	4	10	8	4

Sol. The given observations are already in ascending order. Now, let us make the cumulative frequency.

Age (x_i)	Frequency (f_i)	cf
10	3	3
12	5	8
15	4	12
18	10	22
21	8	30
23	4	34
Total	$N = 34$	

Here, $\Sigma f_i = N = 34$, which is even.

∴ Median

$$= \frac{\text{Value of} \left(\frac{34}{2}\right)\text{th observation} + \text{Value of} \left(\frac{34}{2}+1\right)\text{th observation}}{2}$$

$$= \frac{\text{Value of 17th observation + Value of 18th observation}}{2}$$

$$= \frac{18+18}{2} = 18$$

[∵ both of these observation lies in the cumulative frequency 22 and its corresponding observation is 18.]

Now, let us make the following table from the given data.

$\|x_i - 18\|$	8	6	3	0	3	5	Total
$f_i\|x_i - 18\|$	24	30	12	0	24	20	110

$$= \frac{\Sigma f_i |x_i - M|}{\Sigma f_i}$$

$$= \frac{110}{34} = 3.24 \text{ years}$$

EXAMPLE |10| Find the mean deviation from the median for the following data. **[NCERT]**

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

Sol. The given observations are already in ascending order. Now, let us make the following table from the given data.

x_i	f_i	cf	$\|x_i - M\|$	$f_i\|x_i - M\|$
15	3	3	15	45
21	5	8	9	45
27	6	14	3	18
30	7	21	0	0
35	8	29	5	40
Total	29		32	148

Here, $N = \Sigma f_i = 29$, which is odd.

So, the median $= \left(\frac{N+1}{2}\right)$th observation

i.e. $\frac{29+1}{2} = 15$th observation, which is equal to 30.

[∵ 15th observation lie in the cumulative frequency 21 and its corresponding observation is 30.]

Thus, the median $(M) = 30$.

Hence, mean deviation from the median

$$= \frac{\Sigma f_i |x_i - 30|}{\Sigma f_i}$$

$$= \frac{148}{29} = 5.1$$

EXAMPLE |11| Find the mean deviation from the median of the following frequency distribution.

Age (in years)	10	11	12	13	14	15	16
Frequency	3	8	14	19	7	6	3

Sol. The given observations are already in ascending order. Now, let us make the following table from the given data.

Age x_i	Frequency (f_i)	cf	$\|x_i - M\|$ $= \|x_i - 13\|$	$f_i\|x_i - M\|$
10	3	3	3	9
11	8	11	2	16
12	14	25	1	14
13	19	44	0	0
14	7	51	1	7
15	6	57	2	12
16	3	60	3	9
Total	$N = 60$			67

Here, $N = \Sigma f_i = 60$, which is even.

∴ Median (M)

$$= \frac{\left(\frac{N}{2}\right)\text{th observation} + \left(\frac{N}{2}+1\right)\text{th observation}}{2}$$

$$= \frac{\left(\frac{60}{2}\right)\text{th observation} + \left(\frac{60}{2}+1\right)\text{observation}}{2}$$

$$= \frac{\text{30th observation + 31st observation}}{2}$$

$$= \frac{13+13}{2} = 13$$

[∵ Both of these observations lie in the cumulative frequency 44 and its corresponding observation is 13.]

∴ Mean deviation from the median $= \frac{\Sigma f_i |x_i - M|}{N}$

$$= \frac{67}{60} = 1.12 \text{ yr}$$

(B) FOR CONTINUOUS FREQUENCY DISTRIBUTION

A continuous frequency distribution is a series in which the data is classified into different class intervals without gaps alongwith their respective frequencies.

MEAN DEVIATION ABOUT MEAN

For calculating mean deviation from mean 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 mean.

EXAMPLE |12| 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

Sol. Let us make the following table from the given data.

Marks obtained	Number of students (f_i)	Mid-points (x_i)	$f_i x_i$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$
10-20	2	15	30	30	60
20-30	3	25	75	20	60
30-40	8	35	280	10	80
40-50	14	45	630	0	0
50-60	8	55	440	10	80
60-70	3	65	195	20	60
70-80	2	75	150	30	60
Total	40		1800		400

Here, $N = \sum_{i=1}^{7} f_i = 40$ and $\sum_{i=1}^{7} f_i x_i = 1800,$

Therefore, mean $(\bar{x}) = \frac{1}{N}\sum_{i=1}^{7} f_i x_i = \frac{1800}{40} = 45$

Now, mean deviation,

$$\text{MD}(\bar{x}) = \frac{1}{N}\sum_{i=1}^{7} f_i|x_i - \bar{x}| \quad \left[\because \sum_{i=1}^{7} f_i|x_i - \bar{x}| = 400\right]$$

$$= \frac{1}{40} \times 400 = 10$$

Hence, the mean deviation about mean is 10.

MEAN DEVIATION ABOUT MEDIAN

For calculating mean deviation from median of a continuous frequency distribution, the procedure is same as about mean. The only difference is that, here we replace mean by median and median is calculated by the following formula,

$$M = l + \frac{\frac{N}{2} - cf}{f} \times h$$

l, f, h and cf are respectively the lower limit, the frequency, the width of median class and cumulative frequency of class just preceding the median class.

EXAMPLE |13| 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

Sol. Let us make the following table from the given data.

Class	Mid value (x_i)	Frequency (f_i)	Cumulative frequency (cf)	$\lvert x_i - 14\rvert$	$f_i\lvert x_i - 14\rvert$
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
Total		$N = \Sigma f_i = 44$			$\Sigma f_i\lvert x_i - 14\rvert = 278$

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

Now, $\quad \text{median} = l + \frac{\frac{N}{2} - cf}{f} \times h$

where, $l = 12$, $f = 12$, $cf = 18$ and $h = 6$

$\therefore \quad \text{Median} = 12 + \frac{22 - 18}{12} \times 6 = 12 + \frac{4 \times 6}{12} = 14$

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

$= \frac{278}{44} = 6.318$

EXAMPLE |14| Calculate the mean deviation about median for the following data. **[NCERT]**

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

First, calculate the median by using the formula,

$$M = l + \frac{\frac{N}{2} - cf}{f} \times h.$$

Further, calculate the mean deviation about median, using

$$\text{MD} = \frac{\Sigma f_i|x_i - M|}{N}$$

Sol. Let us make the following table from the given data.

Class	f_i	cf	Mid-point (x_i)	$\lvert x_i - M\rvert$	$f_i\lvert x_i - M\rvert$
0-10	6	6	5	23	138
10-20	7	13	15	13	91
20-30	15	28	25	3	45
30-40	16	44	35	7	112
40-50	4	48	45	17	68
50-60	2	50	55	27	54
Total	50				508

Here, $\frac{N}{2} = \frac{50}{2} = 25$

which item lies in the cumulative frequency 28 . Therefore, 20-30 is the median class.

So, we have, $l = 20$, $cf = 13$, $f = 15$, $h = 10$ and $N = 50$

Now, Median, $M = l + \frac{\frac{N}{2} - cf}{f} \times h = 20 + \frac{25 - 13}{15} \times 10$

$= 20 + 8 = 28$

Hence, the mean deviation about median is given by

$$\text{MD}(M) = \frac{1}{N} \sum_{i=1}^{6} f_i |x_i - M| = \frac{1}{50} \times 508 = 10.16$$

Hence, the mean deviation about median is 10.16.

EXAMPLE |15| Calculate the mean deviation about median for the age distribution of 100 persons given below. **[NCERT]**

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

First, make the class intervals are of uniform length and then use the following formula.

Median $(M) = l + \frac{\frac{N}{2} - cf}{f} \times h$ and $MD\ (M) = \frac{\Sigma f_i\,|x_i - M|}{\Sigma f_i}$

Sol. Let us make the following table from the given data.

Class	f_i	cf	Mid-value (x_i)	$\|x_i - M\|$	$f_i\|x_i - M\|$
15.5-20.5	5	5	18	$\|18 - 38\| = 20$	100
20.5-25.5	6	11	23	$\|23 - 38\| = 15$	90
25.5-30.5	12	23	28	$\|28 - 38\| = 10$	120
30.5-35.5	14	37	33	$\|33 - 38\| = 5$	70
35.5-40.5	26	63	38	$\|38 - 38\| = 0$	0
40.5-45.5	12	75	43	$\|43 - 38\| = 5$	60
45.5-50.5	16	91	48	$\|48 - 38\| = 10$	160
50.5-55.5	9	100	53	$\|53 - 38\| = 15$	135
Total	$N = 100$				735

Here, $\frac{N}{2} = 50$

So, the median class is 35.5-40.5.

$\therefore$ $l = 35.5$, $cf = 37$, $f = 26$ and $h = 5$

Now, Median $(M) = l + \frac{\frac{N}{2} - cf}{f} \times h$

$= 35.5 + \frac{50 - 37}{26} \times 5 = 35.5 + \frac{13}{26} \times 5 = 35.5 + 2.5 = 38$

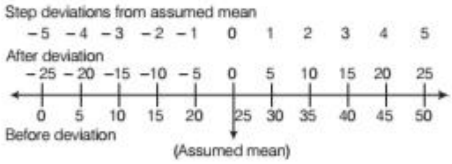
and Mean deviation about the median

$= \frac{\Sigma f_i\,|x_i - M|}{\Sigma f_i} = \frac{735}{100} = 7.35$

Hence, the mean deviation about median is 7.35.

Shortcut Method for Calculating the Mean Deviation about Mean

Sometimes, the data is too large and then the calculation by the previous method is tedious. So, we apply the step deviation method. In this method, we take an assumed mean, which is in the middle or just close to it, in the data. The process of taking step deviation is the change of scale on the number line as shown in the figure given below



For step deviation method, we denote the new variable by

u_i and it is defined as $u_i = \frac{x_i - a}{h}$

where, a is the assumed mean and h is the common factor or length of the class interval.

The mean $\bar{x}$ by step deviation method is given by

$$\bar{x} = a + \frac{\sum_{i=1}^{n} f_i u_i}{N} \times h$$

EXAMPLE |16| Find the mean deviation from the mean of the following data by shortcut or step deviation method.

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

Sol. Let us make the following table for step deviation and product of frequency with absolute deviation.

Class	(f_i)	Mid-points (x_i)	$u_i = \frac{x_i - 450}{100}$ ($a = 450$, $h = 100$)	$f_i u_i$	$\|x_i - \bar{x}\|$ $= \|x_i - 358\|$	$f_i\|x_i - \bar{x}\|$
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
Total	$N = \Sigma f_i = 50$			−46		7896

Here, $a = 450$

$\Sigma f_i u_i = -46$ and $h = 100$.

$\therefore \quad \bar{x} = a + h\left(\frac{1}{N}\Sigma f_i u_i\right)$

$= 450 + 100 \times \left(-\frac{46}{50}\right) = 358$

Now, mean deviation $= \frac{1}{N}\Sigma f_i |x_i - \bar{x}| = \frac{7896}{50} = 157.92$

LIMITATIONS OF MEAN DEVIATION

The following limitations of mean deviations are given below

- If the data is more scattered or the degree of variability is very high, then the median is not a valid representative. Thus, the mean deviation about the median is not fully relied.
- The sum of the deviations from the mean is more than the sum of the deviations from the median. Therefore, the mean deviation about mean is not very scientific.
- The mean deviation is calculated on the basis of absolute values of the deviations and so cannot be subjected to further algebraic treatment. Sometime, it gives unsatisfactory results.

TOPIC PRACTICE 1

OBJECTIVE TYPE QUESTIONS

1. The number which indicates variability of data or observations, is called
(a) measure of central tendency
(b) mean
(c) median
(d) measure of dispersion

2. Which of the following is/are true about the range of the data?
I. It helps to find the variability in the observations on the basis of maximum and minimum value of observations.
II. Range of series = Minimum value – Maximum value.
III. It tells us about the dispersion of the data from a measure of central tendency.
(a) Only I is true (b) II and III are true
(c) I and II are true (d) All are true

3. Mean deviation about the median for the data 3, 9, 5, 3, 12, 10, 18, 4, 7, 19, 21 is
(a) 4.27 (b) 5.24 (c) 5.27 (d) 4.24

4. The mean deviation from the mean of the set of observations –1, 0 and 4 is
(a) 3 (b) 1 (c) –2 (d) 2

5. For the arithmetic progression $a, (a+d), (a+2d), (a+3d), \ldots, (a+2nd)$, the mean deviation from mean is
(a) $\frac{n(n+1)d}{2n-1}$ (b) $\frac{n(n+1)d}{2n+1}$
(c) $\frac{n(n-1)d}{2n+1}$ (d) $\frac{(n+1)d}{2}$

SHORT ANSWER Type Questions

6 Find the mean deviation from the mean for the following data
6, 5, 5.25, 5.5, 4.75, 4.5, 6.25, 7.75, 9.

7 Find the mean deviation of the data 3, 10, 10, 4, 7, 10, 5 from the mean.

8 Find the mean deviation of the data 2, 9, 9, 3, 6, 9, 4 from the mean.

9 The scores of a batsman in 10 innings are 48, 80, 58, 44, 52, 65, 73, 56, 64, 54. Find the mean deviation from the median.

10 Find the mean deviation from the median for the given data 3, 9, 5, 3, 12, 10, 18, 4, 7, 9, 21 **[NCERT]**

11 Following are the marks obtained by 9 students in a mathematics test: 50, 69, 20, 33, 53, 39, 40, 65, 59 find the mean deviation from the median.

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

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

13 Find the mean deviation from the median for the following data. **[NCERT]**

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

14 Find the mean deviation about median for the following data.

x_1	3	6	9	12	13	15	21	22
f_i	3	4	5	2	4	5	4	3

15 Calculate the mean deviation from the mean of the following distribution.

Marks	0-10	10-20	20-30	30-40	40-50
Number of students	5	8	15	16	6

16 Find the mean deviation from the mean for the following data. **[NCERT]**

Class	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	6	8	14	16	4	2

17 Calculate the mean deviation from the mean of the following distribution.

Marks	0-10	10-20	20-30	30-40	40-50
Number of students	5	8	15	16	6

LONG ANSWER Type Questions

18 Find the mean deviation from the mean for the following data.

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

19 Find the mean deviation from the median for the following data.

x_1	74	89	42	54	91	94	35
f_1	20	12	2	4	5	3	4

20 Calculate the mean deviation from the median for the following data.

Wages per day	Number of workers
20-30	3
30-40	8
40-50	12
50-60	9
60-70	8

21 Find the mean deviation about the mean for the following data. **[NCERT]**

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

22 Find the mean deviation from the mean for the following data by shortcut method.

Class	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	8	10	15	25	20	18	9	5

23 Calculate the mean deviation from the median for the given data.

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

HINTS & ANSWERS

1. (d) Variability is another factor which is required to be studied under Statistics. Like *'measures of central tendency'* we want to have a single number to describe variability. This single number is called a *'measure of dispersion'*.

2. (a) Range of a series = Maximum value − Minimum value
The range of data gives us a rough idea of variability or scatter but does not tell about the dispersion of the data from a measure of central tendency.

3. (c) Arranging the data into ascending order, we have 3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21.

Now, median $= \left(\frac{11+1}{2}\right)^{th}$ observation = 9

Now, $|x_i - M|$ are 6, 6, 5, 4, 2, 0, 1, 3, 9, 10, 12.

Therefore, $\sum_{i=1}^{11} |x_i - M| = 58$

and $\text{MD}(M) = \frac{1}{11}\sum_{i=1}^{11}|x_i - M|$

$= \frac{1}{11} \times 58 = 5.27$

4. (d) Mean $(\bar{x}) = \frac{-1+0+4}{3} = 1$

$\therefore$ $\text{MD}(\bar{x}) = \frac{\Sigma|x_i - \bar{x}|}{n}$

$= \frac{|-1-1| + |0-1| + |4-1|}{3} = 2$

5. (b) The mean of the series $a, a+d, \dots, a+2nd$ is

$$\bar{x} = \frac{1}{2n+1}[a + a + d + a + 2d + \cdots + a + 2nd]$$

$$= a + nd$$

$\therefore$ Mean deviation from mean

$$= \frac{1}{2n+1}\sum_{r=0}^{2n}|(a+rd) - (a+nd)|$$

$$= \frac{1}{2n+1}\sum_{r=0}^{2n}|r-n|\, d$$

$$= \frac{1}{2n+1} 2d(1 + 2 + \cdots + n)$$

$$= \frac{n(n+1)}{2n+1} d$$

6. Let $\bar{x}$ be the mean of given data.

Then, $\bar{x} = \frac{6 + 5 + 5.25 + 5.5 + 4.75 + 4.5 + 6.25 + 7.75 + 9}{9}$

$= \frac{54}{9} = 6$

We make the table from the given data.

x_i	$x_i - \bar{x} = x_i - 6$	$\lvert x_i - \bar{x}\rvert$
6	0	0
5.0	− 1	1.00
5.25	− 0.75	0.75
5.5	− 0.5	0.50
4.75	− 1.25	1.25
4.5	− 1.50	1.50
6.25	0.25	0.25
7.75	1.75	1.75
9	3	3
Total		10.00

∴ Mean deviation from mean,

$$\text{MD}(\bar{x}) = \frac{\sum \lvert x_i - \bar{x}\rvert}{n} = \frac{10}{9} = 1.1$$

Hence, mean deviation form mean is 1.1.

7. Solve as Q.6 **Ans.** 2.57

8. Solve as Q.6 **Ans.** 2.57

9. Arranging the data in ascending order, we have
44, 48, 52, 54, 56, 58, 64, 65, 73, 80
Here, $n = 10$. So, median is the mean of 5th and 6th terms.

$$\therefore \quad \text{Median } (M) = \left(\frac{56 + 58}{2}\right) = 57$$

We make the table from the given data.

Scores (x_i)	Deviation from median ($x_i - M$)	$\lvert x_i - M\rvert$
44	44 − 57 = − 13	13
48	48 − 57 = − 9	9
52	52 − 57 = − 5	5
54	54 − 57 = − 3	3
56	56 − 57 = − 1	1
58	58 − 57 = 1	1
64	64 − 57 = 7	7
65	65 − 57 = 8	8
73	73 − 57 = 16	16
80	80 − 57 = 23	23
Total		86

$$\therefore \text{ Mean deviation} = \frac{\sum \lvert x_i - M\rvert}{n} = \frac{86}{10} = 8.6$$

Hence, the mean deviation from the median is 8.6.

10. Solve as Q.9 **Ans.** 4.27

11. Solve as Q.9 **Ans.** 12.67

12. Find the median and then calculate the mean deviation about the median by using the formula $\frac{\sum f_i \lvert x_i - M\rvert}{\sum f_i}$.

Here, $N = 30$, which is even.

So, median is the mean of the 15th and 16th observations. Both of these observations lie in the cumulative frequency 18 for which the corresponding observation is 13.

Therefore,

$$\text{Median } (M) = \frac{\text{15th observation} + \text{16th observation}}{2}$$

$$= \frac{13 + 13}{2} = 13$$

We make the table from the given data.

x_i	f_i	cf	$\lvert x_i - M\rvert$	$f_i\lvert x_i - M\rvert$
3	3	3	10	30
6	4	7	7	28
9	5	12	4	20
12	2	14	1	2
13	4	18	0	0
15	5	23	2	10
21	4	27	8	32
22	3	30	9	27

We have, $\sum f_i = 30$

and $\sum f_i \lvert x_i - M\rvert = 149$

∴ Mean deviation about median,

$$\text{MD}(M) = \frac{1}{N}\sum f_i \lvert x_i - M\rvert$$

$$= \frac{1}{30} \times 149 = 4.97$$

13. Find the median and then calculate the mean deviation about the median by using the formula $\frac{\sum f_i \lvert x_i - M\rvert}{\sum f_i}$.

Here, $N = \sum f_i = 29$, which is odd. So, the median is the $\left(\frac{n+1}{2}\right)$th observation i.e. $\frac{29+1}{2} = 15$th observation, which is equal to 30. Thus, median is 30.

We make the table from the given data.

x_i	f_i	cf	$\lvert x_i - M\rvert$	$f_i\lvert x_i - M\rvert$
15	3	3	15	45
21	5	8	9	45
27	6	14	3	18
30	7	21	0	0
35	8	29	5	40
Total	29			148

$$\therefore \text{ Mean deviation from the median} = \frac{\sum f_i \lvert x_i - 30\rvert}{\sum f_i}$$

$$= \frac{148}{29} = 5.1$$

14. Solve as Q.13 **Ans.** 4.63

15. We make the table from the given data.

Class	Mid value (x_i)	f_i	$u_i = \frac{x_i - 25}{10}$	$f_i u_i$	$\lvert x_i - \bar{x}\rvert = \lvert x_i - 27\rvert$	$f_i \lvert x_i - \bar{x}\rvert$
0-10	5	5	−2	−10	22	110
10-20	15	8	−1	−8	12	96
20-30	25	15	0	0	2	30
30-40	35	16	1	16	8	128
40-50	45	6	2	12	18	108
Total		50		10		472

Here, $\sum f_i = 50, a = 25$

$\sum f_i u_i = 10, h = 10$

$\therefore$ Mean, $\bar{x} = a + \frac{\sum f_i u_i}{\sum f_i} \times h = 25 + \frac{10}{50} \times 10 = 27$

$\therefore$ Mean deviation from mean $= \frac{\sum f_i \lvert x_i - \bar{x}\rvert}{\sum f_i}$

$[\because \sum f_i \lvert x_i - \bar{x}\rvert = 472]$

$= \frac{472}{50} = 9.44$

16. Solve as Q. 15. **Ans.** 10.576

17. Solve as Q. 15. **Ans.** 9.44

18. Solve as Q. 15. **Ans.** 16

19. Solve as Q. 12. **Ans.** 12.5

20. Let us make the following table from the given data.

Wages per day (in ₹)	Mid value (x_i)	f_i	cf	$\lvert x_i - M\rvert = \lvert x_i - 47.5\rvert$	$f_i \lvert x_i - M\rvert$
20-30	25	3	3	22.5	67.5
30-40	35	8	11	12.5	100.0
40-50	45	12	23	2.5	30.0
50-60	55	9	32	7.5	67.5
60-70	65	8	40	17.5	140.0
Total		40			405.0

Here, $\frac{N}{2} = \frac{40}{2} = 20$. The cumulative frequency just greater than 20 is 23, so the median class is 40-50.

So, we have $l = 40$, $f = 12$, $cf = 11$, $h = 10$ and $N = 40$.

Now, Median $(M) = l + \frac{\frac{N}{2} - cf}{f} \times h = 40 + \frac{20 - 11}{12} \times 10$

$= 40 + \frac{90}{12} = 40 + 7.5 = 47.5$

and Mean deviation from the median $= \frac{\Sigma f_i \lvert x_i - M\rvert}{N}$

$= \frac{405}{40} = 10.125$

21. Take the assumed mean $a = 45$ and $h = 10$ Make the table for step deviation and product of frequency with absolute deviation

Marks obtained	Number of students (f_i)	Mid-points(x_i)	$u_i = \frac{x_i - 45}{10}$	$f_i u_i$	$\lvert x_i - \bar{x}\rvert$	$f_i \lvert x_i - \bar{x}\rvert$
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
Total	40			0		400

Here, $\sum_{i=1}^{7} f_i = 40$, $\sum_{i=1}^{7} f_i u_i = 0$ and $\sum f_i \lvert x_i - \bar{x}\rvert = 400$

Now, $\bar{x} = a + \frac{\sum_{i=1}^{7} f_i u_i}{\sum_{i=1}^{7} f_i} \times h = 45 + \frac{0}{40} \times 10 = 45$

$\therefore$ Mean deviation about mean

$= \frac{1}{\sum_{i=1}^{7} f_i} \sum_{i=1}^{7} f_i \lvert x_i - \bar{x}\rvert$

$= \frac{400}{40} = 10$

22. Solve as Q.21. **Ans.** 14.95

23. Solve as Q. 21. **Ans.** 9

|TOPIC 2|
Variance and Standard Deviation

Due to the limitations of mean deviation, some other method is required for measure of dispersion. Standard deviation is such a measure of dispersion.

VARIANCE

The absolute values are considered in calculating the mean deviation about mean or median, otherwise the deviation being negative or positive and may cancel among themselves.

To overcome this difficulty of the signs of the deviations, we take the squares of all the deviations, so that all deviations become non-negative.

Let $x_1, x_2, \ldots, x_n$ be n observations and $\bar{x}$ be their mean. Then, $(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_n - \bar{x})^2$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2$$

An illustration to explain the variance

Let us consider two sets of observations. Set 'A' contains eight observations 4, 8, 12, 16, 20, 24, 28, 32 and set 'B' contains 16 observations 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 24, 26, 28, 31, 33, 36.

Here, the mean obtained by both set of observations have same mean i.e.18

Now, we determine the sum of squares of deviation from mean.

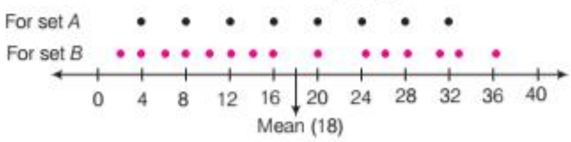
FOR SET A

$$\sum_{i=1}^{8} (x_i - \bar{x})^2 = (4-18)^2 + \ldots + (32-18)^2 = 672$$

FOR SET B

$$\sum_{i=1}^{16} (y_i - \bar{y})^2 = (2-18)^2 + \ldots + (36-18)^2 = 1738$$

If $\sum_{i=1}^{n} (x_i - \bar{x})^2$ is simply our measure of dispersion or scatter about mean, then from the above we can say that set A has lesser dispersion than set B, even though the observations in set A are more scattered about mean, which is clear from the following diagrams.



Thus, we can say that the sum of squares of deviations about mean is not a proper measure of dispersion.

To overcome this difficulty, we take the mean of the squares of the deviations i.e. $\frac{1}{n}\Sigma(x_i - \bar{x})^2$.

Definition The mean of the squares of the deviations from mean is called the **variance** and it is denoted by the symbol σ^2.

Significance of Deviation

1. If the deviation is zero, it means there is no deviation at all and all observations are equal to mean.
2. If deviation is small, this indicates that the observation are close to the mean.
3. If the deviation is large, there is a high degree of dispersion of the observation from the mean.

STANDARD DEVIATION

Standard deviation is the square root of the arithmetic mean of the squares of deviations from mean and it is denoted by the symbol σ.

Or

The square root of variance, is called standard deviation.

i.e. $\sqrt{\sigma^2}$ or σ.

It is also known as **root mean square deviation.**

VARIANCE AND STANDARD DEVIATION OF UNGROUPED DATA

Variance of n observations $x_1, x_2, x_3, \ldots x_n$ is given by

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} \text{ or } \frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2$$

We know that, standard deviation $= \sqrt{\text{Variance}}$

$$\therefore \quad \sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} \text{ or } \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2}$$

$$\text{or} \quad \frac{1}{n}\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

WORKING RULE TO FIND VARIANCE AND STANDARD DEVIATION

To find the standard deviation or variance when deviations are taken from actual mean, we use the following working steps

***Step* I** Calculate the mean $\bar{x}$ of the given observation $x_1, x_2, \ldots, x_n$.

***Step* II** Take the deviations of the observations from their mean i.e. find $x_i - \bar{x}; i = 1, 2, \ldots, n$.

***Step* III** Square the deviations obtained in step II and then find the sum i.e. $\sum_{i=1}^{n} (x_i - \bar{x})^2$.

***Step* IV** Find the variance and standard deviation by using the formula, Variance, $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$

and standard deviation $= \sqrt{\sigma}$.

EXAMPLE |1| Find the variance and standard deviation for the following data, 6, 7, 10, 12, 13, 4, 8, 12.

Sol. Given observations are 6, 7, 10, 12, 13, 4, 8, 12.

Number of observations = 8

$\therefore \quad \text{Mean } (\bar{x}) = \frac{6 + 7 + 10 + 12 + 13 + 4 + 8 + 12}{8}$

$= \frac{72}{8} = 9$

Now, let us make the following table for deviation.

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
6	−3	9	13	4	16
7	−2	4	4	−5	25
10	1	1	8	−1	1
12	3	9	12	3	9
Total		74	Total		74

$\therefore$ Sum of squares of deviations $= \sum_{i=1}^{8} (x_i - \bar{x})^2 = 74$

Hence, variance, $\sigma^2 = \frac{\sum_{i=1}^{8} (x_i - \bar{x})^2}{n} = \frac{74}{8} = 9.25$

and standard deviation $= \sqrt{\sigma} = \sqrt{9.25} = 3.04$

EXAMPLE |2| Find the variance and standard deviation for the following data.

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

Sol. Let $\bar{x}$ be the mean of the given set of observations.

Number of observations = 10

$$\therefore \quad \bar{x} = \frac{\left[\begin{array}{r} 45 + 60 + 62 + 60 + 50 + 65 + 58 \\ + 68 + 44 + 48 \end{array}\right]}{10} = \frac{560}{10} = 56$$

Make a table from the given data.

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

We have, $\quad n = 10$ and $\Sigma (x_i - \bar{x})^2 = 662$

$\therefore \quad \text{Variance, } \sigma^2 = \frac{1}{n} \Sigma (x_i - \bar{x})^2$

$= \frac{662}{10} = 66.2$

and standard deviation $= \sqrt{\sigma} = \sqrt{66.2} = 8.136$

EXAMPLE |3| Find the standard deviation and the variance of first n natural numbers.

Sol. The first n natural numbers are 1, 2, 3, …, n.

$\because$ Standard deviation,

$$\text{SD} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2}$$

$$\therefore \quad \text{SD} = \sqrt{\frac{n(n+1)(2n+1)}{6n} - \left(\frac{n(n+1)}{2n}\right)^2}$$

$$\left[\because \sum_{i=1}^{n} x_i^2 = \frac{n(n+1)(2n+1)}{6} \text{ and } \sum_{i=1}^{n} x_i = \frac{n(n+1)}{2}\right]$$

$$= \sqrt{(n+1)\left(\frac{2n+1}{6} - \frac{n+1}{4}\right)}$$

$$= \sqrt{(n+1)\left(\frac{4n + 2 - 3n - 3}{12}\right)}$$

$$= \sqrt{\frac{(n+1)(n-1)}{12}} = \sqrt{\frac{n^2 - 1}{12}}$$

$\therefore \text{ Variance} = (\text{SD})^2 = \frac{n^2 - 1}{12}$

Variance and Standard Deviation of a Discrete Frequency Distribution

Let the discrete frequency distribution be
$x : x_1, x_2, x_3, \ldots, x_n$ and $f : f_1, f_2, f_3, \ldots, f_n$. Then, by

DIRECT METHOD

Variance $(\sigma^2) = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2$

or $\sigma^2 = \frac{1}{N}\Sigma f_i x_i^2 - \left(\frac{\Sigma f_i x_i}{N}\right)^2$

and standard deviation,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}$$

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

where, $N = \sum_{i=1}^{n} f_i$

SHORTCUT METHOD

Variance, $\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{\Sigma f_i d_i}{N}\right)^2$

and standard deviation,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i d_i^2 - \left(\frac{\Sigma f_i d_i}{N}\right)^2}$$

where, $d_i = x_i - a$, deviation from assumed mean and a = assumed mean.

EXAMPLE |4| Find the variance and standard deviation of the following data. **[NCERT]**

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

Sol. Let us make the following table from the given data.

x_i	f_i	$f_i x_i$	$x_i - \bar{x} = x_i - 14$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
4	3	12	−10	100	300
8	5	40	−6	36	180
11	9	99	−3	9	81
17	5	85	3	9	45
20	4	80	6	36	144
24	3	72	10	100	300
32	1	32	18	324	324
Total	30	420			1374

Here, we have, $N = \Sigma f_i = 30, \Sigma f_i x_i = 420$

$$\therefore \quad \bar{x} = \frac{\sum_{i=1}^{7} f_i x_i}{N} = \frac{420}{30} = 14$$

Hence, variance $(\sigma^2) = \frac{1}{N}\sum_{i=1}^{7} f_i (x_i - \bar{x})^2$

$$\left[\because \sigma^2 = \frac{1}{N}\left\{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2\right\}\right]$$

$$= \frac{1}{30} \times 1374 = 45.8$$

and standard deviation, $\sigma = \sqrt{\sigma^2} = \sqrt{45.8} = 6.77$

Alternate Method

Let us make the following table from the given data.

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
4	3	12	16	48
8	5	40	64	320
11	9	99	121	1089
17	5	85	289	1445
20	4	80	400	1600
24	3	72	576	1728
32	1	32	1024	1024
Total	30	420		7254

Here, $N = \sum_{i=1}^{7} f_i = 30, \sum_{i=1}^{7} f_i x_i = 420$

and $\sum_{i=1}^{7} f_i x_i^2 = 7254$

$$\therefore \text{Variance, } \sigma^2 = \frac{1}{N}\sum_{i=1}^{7} f_i x_i^2 - \left(\frac{\sum_{i=1}^{7} f_i x_i}{N}\right)^2$$

$$= \frac{1}{30} \times 7254 - \left(\frac{420}{30}\right)^2$$

$$= \frac{7254}{30} - 196$$

$$= 241.8 - 196 = 45.8$$

and standard deviation, $\sigma = \sqrt{45.8} = 6.77$

EXAMPLE |5| Find the variance and standard deviation for the following distribution using shortcut method. **[NCERT]**

x_i	60	61	62	63	64	65	66	67	68
f_i	2	1	12	29	25	12	10	4	5

Sol. Let the assumed mean be $a = 64$.

Now, let as make a table from the given data.

x_i	f_i	$d_i = x_i - 64$	$f_i d_i$	d_i^2	$f_i d_i^2$
60	2	$60-64=-4$	−8	16	32
61	1	$61-64=-3$	−3	9	9
62	12	$62-64=-2$	−24	4	48
63	29	$63-64=-1$	−29	1	29
64	25	$64-64=0$	0	0	0
65	12	$65-64=1$	12	1	12
66	10	$66-64=2$	20	4	40
67	4	$67-64=3$	12	9	36
68	5	$68-64=4$	20	16	80
Total	100		0	60	286

Here, $N = \Sigma f_i = 100$; $\Sigma f_i d_i = 0$ and $\Sigma f_i d_i^2 = 286$.

Now, variance, $(\sigma^2) = \left[\frac{1}{N}\Sigma f_i d_i^2 - \left(\frac{1}{N}\Sigma f_i d_i\right)^2\right]$

$$= \frac{1}{100} \times 286 - \left(\frac{1}{100} \times 0\right)^2 = 2.86$$

and standard deviation, $\sigma = \sqrt{2.86} = 1.69$

EXAMPLE |6| Find the mean, variance and standard deviation of the following data. **[NCERT]**

x_i	92	93	97	98	102	104	109
f_i	3	2	3	2	6	3	3

Sol. Let the assumed mean be $a = 98$.

Now, let us make the following table from the given data.

x_i	f_i	$d_i = x_i - 98$	d_i^2	$f_i d_i$	$f_i d_i^2$
92	3	−6	36	−18	108
93	2	−5	25	−10	50
97	3	−1	1	−3	3
98	2	0	0	0	0
102	6	4	16	24	96
104	3	6	36	18	108
109	3	11	121	33	363
Total	22			44	728

Here, $N = \Sigma f_i = 22$, $\Sigma f_i d_i = 44$ and $\Sigma f_i d_i^2 = 728$

Now, $\bar{x} = a + \frac{\Sigma f_i d_i}{\Sigma f_i}$

$$= 98 + \frac{44}{22} = 98 + 2 = 100$$

Variance, $\sigma^2 = \frac{1}{N}(\Sigma f_i d_i^2) - \left(\frac{\Sigma f_i d_i}{N}\right)^2$

$$= \frac{1}{22}(728) - \left(\frac{1}{22} \times 44\right)^2 = \frac{1}{22}(728) - (2)^2$$

$$= \frac{1}{22}(728 - 22 \times 2^2) = \frac{1}{22}(728 - 88)$$

$$= \frac{1}{22} \times 640 = \frac{320}{11} = 29.09$$

and standard deviation,

$$\sigma = \sqrt{29.09} = 5.394$$

Variance and Standard Deviation of a Continuous Frequency Distribution

DIRECT METHOD

In this method, we first replace each class by its mid-point, then this method becomes similar to the discrete frequency distribution.

If there is a frequency distribution of n classes and each class defined by its mid-point x_i with corresponding frequency f_i, then variance and standard deviation are respectively

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2$$

and $$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}$$

or $$\sigma^2 = \frac{1}{N^2}[N \Sigma f_i x_i^2 - (\Sigma f_i x_i)^2]$$

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

EXAMPLE |7| Calculate the variance and standard deviation for the following distribution.

Class	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	3	7	12	15	8	3	2

Sol. Let us construct the following table.

Class	Frequency (f_i)	Mid-point (x_i)	$f_i x_i$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
30-40	3	35	105	729	2187
40-50	7	45	315	289	2023
50-60	12	55	660	49	588
60-70	15	65	975	9	135
70-80	8	75	600	169	1352
80-90	3	85	255	529	1587
90-100	2	95	190	1089	2178
Total	50		3100		10050

Here, $\quad N = 50$ and $\Sigma f_i x_i = 3100$

$\therefore \quad$ Mean, $\overline{x} = \frac{1}{N} \Sigma f_i x_i = \frac{3100}{50} = 62$

Now, variance, $\sigma^2 = \frac{1}{N} \Sigma f_i (x_i - \overline{x})^2 = \frac{1}{50} \times 10050 = 201$

and standard deviation, $\sigma = \sqrt{201} = 14.18$

Shortcut Method Or Step-Deviation Method

Sometimes the values of mid-points x_i of different classes in a continuous distribution are large and so the calculation of mean and variance becomes tedious and time consuming. For this, we use the **step deviation method** to find mean and variance.

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

and standard deviation,

$$\sigma = h \sqrt{\frac{1}{N} \sum_{i=1}^{n} f_i u_i^2 - \left(\frac{\sum_{i=1}^{n} f_i u_i}{N} \right)^2}$$

where, $u_i = \frac{x_i - a}{h}$, a = assumed mean and h = width of class-interval.

WORKING RULE TO FIND VARIANCE AND STANDARD DEVIATION BY SHORTCUT METHOD

To find variance and standard deviation, use the following steps

***Step* I** Select an assumed mean, say a and then calculate $u_i = \frac{x_i - a}{h}$, where h = width of class interval or common factor.

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

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

***Step* IV** Substitute the values of $\Sigma f_i u_i$, $\Sigma f_i u_i{}^2$ and $\Sigma f_i = N$ in the formula,

$$\text{Variance } (\sigma^2) = h^2 \left[\frac{1}{N} \Sigma f_i u_i^2 - \left(\frac{1}{N} \Sigma f_i u_i \right)^2 \right]$$

and standard deviation $= \sqrt{\sigma}$

to get the required values.

EXAMPLE |8| The diameter of circles (in mm) drawn in a design are given below.

Diameter (in mm)	33-36	37-40	41-44	45-48	49-52
Number of circles	15	17	21	22	25

Calculate the standard deviation and mean diameter of circles. **[NCERT]**

The given data is discontinuous, so first make the class intervals in continuous form and then solve it.

Sol. Let us first make the data continuous by making classes as 32.5-36.5, 36.5-40.5, 40.5-44.5, 44.5-48.5, 48.5-52.5.

Let assumed mean is 42.5 and $h = 4$.

Then, $\quad u_i = \frac{x_i - a}{h} = \frac{x_i - 42.5}{4}$

Now, let us make the following table from the given data.

Class interval	Mid-value (x_i)	f_i	$u_i = \frac{x_i - 42.5}{4}$	$f_i u_i$	u_i^2	$f_i u_i^2$
32.5-36.5	34.5	15	−2	−30	4	60
36.5-40.5	38.5	17	−1	−17	1	17
40.5-44.5	42.5	21	0	0	0	0
44.5-48.5	46.5	22	1	22	1	22
48.5-52.5	50.5	25	2	50	4	100
Total		100		25		199

Now, Mean $(\overline{x}) = a + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h = 42.5 + 4 \times \frac{25}{100}$

[$\because$ by using step deviation method]

$= 42.5 + 1 = 43.5$

Standard deviation, (σ) $\sqrt{= h^2 \left[\frac{1}{N} \Sigma f_i u_i^2 - \left(\frac{1}{N} \Sigma f_i u_i \right)^2 \right]}$

[by shortcut method]

$$= \sqrt{(4)^2 \left[\frac{1}{100} \times 199 - \left(\frac{1}{100} \times 25 \right)^2 \right]}$$

$$= \sqrt{16 \left(1.99 - \frac{1}{16} \right)} = \sqrt{16 \left(\frac{31.84 - 1}{16} \right)}$$

$$= \sqrt{30.84} = 5.55$$

Hence, standard deviation and mean diameter of circles are 5.55 and 43.5.

EXAMPLE |9| Calculate the mean and standard deviation of the following cumulative data.

Wages (in ₹)	0-15	15-30	30-45	45-60	60-75	75-90	90-105	105-120
Number of workers	12	30	65	107	157	202	222	230

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

Class interval	cf	Mid value (x_i)	f_i	$u_i = \frac{x_i - 67.5}{15}$	f_iu_i	$f_iu_i^2$
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
Total			230		−105	733

Here, $a = 67.5,\ h = 15,\ N = \Sigma f_i = 230,$

$\Sigma f_iu_i = -105$ and $\Sigma f_iu_i^2 = 733$

Now, Mean $= a + h\left(\frac{1}{N}\Sigma f_iu_i\right)$

$$= 67.5 + 15\left(\frac{-105}{230}\right)$$

$$= 67.5 - 6.85 = 60.65$$

Standard deviation,

$$(\sigma) = \sqrt{h^2\left[\frac{1}{N}\sum f_iu_i^2 - \left(\frac{1}{N}\sum f_iu_i\right)^2\right]}$$

$$= \sqrt{225\left[\frac{733}{230} - \left(-\frac{105}{230}\right)^2\right]}$$

$$= \sqrt{225\,[3.187 - (0.46)^2]}$$

$$= \sqrt{225\,(3.1870 - 0.2116)}$$

$$= \sqrt{669.465} = 25.87$$

TOPIC PRACTICE 2

OBJECTIVE TYPE QUESTIONS

1 The quantity which leads to a proper measure of dispersion, is

(a) $\sum(x_i - \bar{x})^2$ (b) $\frac{1}{n}\sum(x_i - \bar{x})$

(c) $\frac{1}{n}\sum(x_i - \bar{x})^2$ (d) $\sum(x_i - \bar{x})$

2 The mean and variance for the data 6, 7, 10, 12, 13, 4, 8, 12 respectively are

(a) 9, 9.50 (b) 8, 8.50 (c) 9, 9.25 (d) 8, 8.25

3 The variance of the following data

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

(a) 41.5 (b) 40.6
(c) 43.4 (d) 45.2

4 Find the variance of the following data

Class interval	4-8	8-12	12-16	16-20
Frequency	3	6	4	7

(a) 13 (b) 18
(c) 19 (d) 20

5 Variance of the data 2, 4, 5, 6, 8, 17 is 23.33. Then, variance of 4, 8, 10, 12, 16, 34 will be **[NCERT Exemplar]**

(a) 23.33 (b) 25.33
(c) 46.66 (d) 48.66

SHORT ANSWER Type I Questions

6 The marks obtained by 7 students are 8, 9, 11, 13, 14, 15, 21. Find the variance and standard deviation of these marks.

7 Find the variance of the data 6, 5, 9, 13, 12, 8 and 10.

8 Find the standard deviation of first 10 natural numbers.

9 The mean of 100 observations is 50 and their standard deviation is 5. Find the sum of all squares of all the observations. **[NCERT Exemplar]**

10 The following relates to sample of size 60, $\Sigma x^2 = 18000$ and $\Sigma x = 960$, find the variance.

SHORT ANSWER Type II Questions

11 Find the standard deviation and variance of the following data.

x_i	140	145	150	155	160	165	170	175
f_i	4	6	15	30	36	24	8	2

12 Find the mean, variance and standard deviation for the following.

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

[NCERT]

13 Calculate mean and variance of the given data.

x_i	2	4	6	8	10	12	14	16
f_i	4	4	5	15	8	5	4	5

14 Calculate the mean and variance of the following data.

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

15 Calculate the mean and variance for the following data.

Class	0-30	30-60	60-90	90-120	120-150	150-180	180-210
Frequency (f)	2	3	5	10	3	5	2

16 Calculate the mean, variance and standard deviation for the following distribution.

Class	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45
Frequency	20	24	32	28	20	11	26	15	24

17 The following table shows the marks obtained by 100 candidates in an examination. Calculate the mean and standard deviation.

Marks	1-10	11-20	21-30	31-40	41-50	51-60
Number of candidates	3	16	26	31	16	8

HINTS & ANSWERS

1. (c) We can take $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ as a quantity which leads to a proper measure of dispersion.

2. (c)

x_i	x_i^2
6	36
7	49
10	100
12	144
13	169
4	16
8	64
12	144
Total = 72	722

$$\text{Mean} = \frac{\Sigma x_i}{n} = \frac{72}{8} = 9$$

$$\text{Variance} = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$$

$$= \frac{722}{8} - \left(\frac{72}{8}\right)^2 = 9.25$$

3. (c)

x_i	f_i	x_i^2	$f_i x_i$	$f_i x_i^2$
6	2	36	12	72
10	4	100	40	400
14	7	196	98	1372
18	12	324	216	3888
24	8	576	192	4608
28	4	784	112	3136
30	3	900	90	2700
Total = 130	40		760	16176

$$\text{Mean }(\bar{x}) = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{760}{40} = 19$$

$$\text{Variance} = \frac{\Sigma f_i x_i^2}{\Sigma f_i} - \left(\frac{\Sigma f_i x_i}{\Sigma f_i}\right)^2 = \frac{16176}{40} - \left(\frac{760}{40}\right)^2 = 43.4$$

4. (c)

Class interval	Mid-value (x_i)	f_i
4-8	6	3
8-12	10	6
12-16	14	4
16-20	18	7

$$\text{Mean }(\bar{x}) = \frac{\Sigma f_i x_i}{\Sigma f_i}$$

$$= \frac{3\times 6 + 6\times 10 + 4\times 14 + 7\times 18}{20} = 13$$

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

$$= \frac{3(-7)^2 + 6(-3)^2 + 4(1)^2 + 7(5)^2}{20} = 19$$

5. (c) When each observation is multiplied by 2, then variance is also multiplied by 2.
We are given, 2, 4, 6, 8, 17.
When each observation multiplied by 2, we get 4, 8, 12, 16, 34.

∴ Variance of new series = 2 × Variance of given data
$= 2 \times 23.33 = 46.66$

6. Here, $\bar{x} = \frac{8+9+11+13+14+15+21}{7}$

$= \frac{91}{7} = 13$ marks

We make the table from the given data.

Marks (x_i)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
8	–5	25
9	–4	16
11	–2	4
13	0	0
14	1	1
15	2	4
21	8	64
Total		114

Here, $n = 7$, $\sum (x_i - \bar{x})^2 = 114$

$\therefore \quad \sigma^2 = \frac{1}{n}\sum (x_i - \bar{x})^2 = \frac{114}{7} = 16.29$

Also, standard deviation of marks, $\sigma = \sqrt{16.29} = 4.04$

Hence, variance is 16.29 and standard deviation is 4.04

7. Solve as Q. 6. **Ans.** 52/7

8. Similar as Example 3. **Ans.** 2.87

9. Given, $\bar{x} = 50$, $n = 100$ and $\sigma = 5$

We know that, $\sigma^2 = \frac{\Sigma x_i^2}{n} - (\bar{x})^2 \Rightarrow \frac{\Sigma x_i^2}{n} = \sigma^2 + (\bar{x})^2$

$\Rightarrow \quad \Sigma x_i^2 = n[\sigma^2 + (\bar{x})^2] = 100\,[5^2 + (50)^2]$

$= 100\,(25 + 2500) = 252500$

Hence, the sum of all squares of all the observations is 252500.

10. Use, $\sigma^2 = \frac{\Sigma x_i^2}{n} - \left(\frac{\Sigma x_i}{n}\right)^2$ **Ans.** 44

11. Let us take assumed mean, $a = 155$

We make the table from the given data.

x_i	f_i	$u_i = \frac{x_i - 155}{5}$	u_i^2	$f_i u_i$	$f_i u_i^2$
140	4	–3	9	–12	36
145	6	–2	4	–12	24
150	15	–1	1	–15	15
155	30	0	0	0	0
160	36	1	1	36	36
165	24	2	4	48	96
170	8	3	9	24	72
175	2	4	16	8	32
Total	125			77	311

$\therefore$ Variance, $\sigma^2 = \left[\frac{\sum f_i u_i^2}{N} - \left(\frac{\sum f_i u_i}{N}\right)^2\right] \times h^2$

$= \left[\frac{311}{125} - \left(\frac{77}{125}\right)^2\right] \times 25 = 25 \times \frac{311}{125} - 25 \times \frac{77 \times 77}{125 \times 125}$

$= \frac{7775}{125} - \frac{148225}{15625} = 62.200 - 9.4864 = 52.7136$

$\therefore$ Standard deviation $= \sqrt{\text{Variance}}$

$= \sqrt{52.7136} = 7.26$ (approx)

12. Solve as Example 6. **Ans.** 14, 45.85, 6.77

13. Solve as Example 6. **Ans.** 9, 15.08

14. Solve as Example 6. **Ans.** 3.88, 3.64

15. Here, $h = 30$ and $a = 105$

We make the table from the given data.

Class interval	Frequency (f_i)	Mid value (x_i)	$u_i = \frac{x_i - 105}{30}$	$f_i u_i$	u_i^2	$f_i u_i^2$
0-30	2	15	– 3	– 6	9	18
30-60	3	45	– 2	– 6	4	12
60-90	5	75	– 1	– 5	1	5
90-120	10	105	0	0	0	0
120-150	3	135	1	3	1	3
150-180	5	165	2	10	4	20
180-210	2	195	3	6	9	18
Total	30			2		76

Here, $N = 30$, $\sum f_i u_i = 2$, $\sum f_i u_i^2 = 76$

and $\quad h = 30$

$\therefore$ Mean $(\bar{x}) = a + \left(\frac{1}{N}\sum f_i u_i\right) \times h$

$= 105 + \left(\frac{1}{30} \times 2\right) \times 30$

$= 105 + 2 = 107$

Variance, $\sigma^2 = h^2 \left[\frac{1}{N}\sum f_i u_i^2 - \left(\frac{1}{N}\sum f_i u_i\right)^2\right]$

$= (30)^2 \left[\frac{1}{30} \times 76 - \left(\frac{1}{30} \times 2\right)^2\right]$

$= (30)^2 \left[\frac{76}{30} - \frac{1}{225}\right] = 30\left[76 - \frac{30}{225}\right]$

$= 30\,[76 - 0.13] = 30 \times 75.87 = 2276.1$

Hence, mean is 107 and variance is 2276.1.

16. Solve as Q. 15. **Ans.** 21.5, 164.75, 12.835

17. Solve as Example 9. **Ans.** 32, 12.36

SUMMARY

- The single number that describe variability is called **measure of dispersion.**
- **Range** is the difference of maximum and minimum values of data.
- The arithmetic mean of the absolute deviations of all the values taken about any central value is called **mean deviation**.

$$\therefore \qquad \text{MD} = \frac{\text{Sum of absolute values of deviations from } a}{\text{Number of observations}}$$

(i) **Mean Deviation for Ungrouped Data is** $\text{MD} = \frac{\sum_{i=1}^{n} |x_i - a|}{n}$

where, a = mean or median n = number of observations.

(ii) **Mean Deviation for Discrete Frequency Distribution** is

$$\text{MD} = \frac{\sum_{i=1}^{n} f_i |x_i - A|}{N}, \text{ where } N = \sum_{i=1}^{n} f_i = \text{ total frequency and } A = \text{Mean or Median}$$

(iii) **Mean Deviation for Continuous Frequency Distribution** is $\text{MD} = l + \frac{\frac{N}{2} - cf}{f} \times h$

where, h, l and f are width, lower limit and frequency of median class and cf is cumulative frequency of class just preceding the median class.

- The mean of squares of the deviations from mean is called the **variance** i.e. $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$
- The square root of variance is called **standard deviation** (SD)

i.e. $\qquad \sigma = \sqrt{\text{Variance}}$

(i) **SD of Ungrouped Data** is $\sigma = \frac{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}}{n}$ and Variance, $\sigma^2 = (\text{SD})^2$

(ii) **SD of Discrete Frequency Distribution** is $\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{n} f_i (x_i - \overline{x})^2}$

or $\qquad \sigma = \frac{1}{N} \sqrt{N \sum_{i=1}^{h} f_i x_i^2 - (\sum_{i=1}^{h} f_i x_i)^2}$

CHAPTER PRACTICE

OBJECTIVE TYPE QUESTIONS

1. Which of the following is/are used for the measures of dispersion?
(a) Range
(b) Quartile deviation
(c) Standard deviation
(d) All of the above

2. Mean deviation about median for continuous frequency distribution is calculated by using the formula

(a) MD $(M) = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i + M|$

(b) MD $(M) = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - M|$

(c) MD $(M) = \frac{1}{N}\sum_{i=1}^{n} \frac{f_i}{|x_i - M|}$

(d) MD $(M) = \frac{1}{N}\sum_{i=1}^{n} \frac{|x_i - M|}{f_i}$

3. The mean deviation about the mean of the set of first n natural numbers when n is an even number, is
(a) $\frac{n}{2}$ (b) $\frac{n}{4}$ (c) $\frac{n}{6}$ (d) $\frac{n}{8}$

4. If x_1, x_2, x_3, x_4 and x_5 be the observations with mean m and standard deviation s then, the standard deviation of the observations kx_1, kx_2, kx_3, kx_4 and kx_5 is
(a) $k + s$ (b) $\frac{s}{k}$ (c) ks (d) s

5. The standard deviation for the following data is

x_i	3	8	13	18	23
f_i	7	10	15	10	6

(a) 6.21 (b) 6.31 (c) 6.12 (d) 6.13

6. Consider the following data
57, 64, 43, 67, 49, 59, 44, 47, 61, 59
Match the following measures of data in Column I with their corresponding values in Column II and choose the correct option from the codes given below.

	Column I		Column II
A.	Mean of the data	1.	66.2
B.	Variance of the data	2.	8.13
C.	Standard deviation of the data	3.	55

Codes

	A	B	C		A	B	C
(a)	3	2	1	(b)	3	1	2
(c)	2	3	1	(d)	2	1	3

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

8. If the mean deviation of numbers, $1, 1+d, 1+2d, \ldots, 1+100d$ from their mean is 255, then d is equal to
(a) 10.0 (b) 20.0 (c) 10.1 (d) 20.2

SHORT ANSWER Type Questions

9. The frequency distribution

x	A	$2A$	$3A$	$4A$	$5A$	$6A$
f	2	1	1	1	1	1

where, A is a positive integer, has a variance of 160. Determine the value of A.

10. Find the mean and standard deviation for the following data.

Class interval	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
Frequency	3	2	4	6	5	5	5	2	8	5

11. Calculate mean deviation (from median) for the following data. **[NCERT]**

Marks	0-10	10-20	20-30	30-40	40-50
Number of students	5	8	15	16	6

12. Find the mean and standard deviation of the following frequency distribution. **[NCERT]**

x_i	6	10	14	18	24	28	30
f_i	2	4	7	12	8	4	3

13. Find the standard deviation for the following data.

x_i	3	8	13	18	23
f_i	7	10	15	10	6

14. Find the mean and standard deviation for the following data.

Age (in years)	Number of teachers
25-30	30
30-35	23
35-40	20
40-45	14
45-50	10
50-55	3

15. Find the variance and standard deviation for the following distribution.

x_i	4.5	14.5	24.5	34.5	44.5	54.5	64.5
f_i	1	5	12	22	17	9	4

LONG ANSWERS Type Questions

16. An original frequency table with mean 11 and variance 9.9 was lost but the following table derived from it was found. Construct the original table.

Value of deviation (d)	−2	−1	0	1	2
Frequency (f)	1	6	7	4	2

17. The mean and variance of eight observations are 9 and 9.25, respectively. If six of the observations are 6, 7, 10, 12, 12 and 13, then find the remaining two observations. [NCERT]

CASE BASED Question

18. For a group of 200 candidates, the mean and the standard deviation 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.

Answer the following questions.

(i) Find the sum of correct scores.
(a) 7991 (b) 8000 (c) 8550 (d) 6572

(ii) Find the correct mean.
(a) 42.924 (b) 39.955
(c) 38.423 (d) 41.621

(iii) The formula of variance is

(a) $\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$ (b) $\sum_{i=1}^{n}(x_i-\bar{x})^2$

(c) $\dfrac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\Sigma f_i}$ (d) $\sum_{i=1}^{n}f_i(x_i-\bar{x})^2$

(iv) Find the correct variance.
(a) 280.3 (b) 235.6
(c) 224.143 (d) 226.521

(v) Find the correct standard deviation.
(a) 14.971 (b) 11.321
(c) 16.441 (d) 12.824

HINTS & ANSWERS

1. (d) The dispersion or scatter in a data is measured on the basis of the observations and the types of the measure of central tendency, used there. There are following measures of dispersion

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

2. (b) To find the median for a continous frequency distribution. The data is first arranged in ascending order. After finding the median, the absolute values of the deviations of mid-point x_i of each class from the median *i.e.*, $|x_i - M|$ are obtained.

Then, mean deviation about median, i.e.

$$MD(M) = \frac{1}{N}\sum_{i=1}^{n} f_i|x_i - M|$$

3. (b) Consider, first n natural number, which n is even, i.e. 1, 2, 3, 4, ..., n (even).

Clearly, mean $(\bar{x}) = \dfrac{1+2+3+\dots+n}{n} = \dfrac{n(n+1)}{2n} = \dfrac{n+1}{2}$

Now, $MD = \frac{1}{n}\left[\left|1-\frac{n+1}{2}\right| + \left|2-\frac{n+1}{2}\right| + \left|3-\frac{n+1}{2}\right| + \dots + \left|\frac{n-2}{2}-\frac{n+1}{2}\right| + \left|\frac{n}{2}-\frac{n+1}{2}\right| + \left|\frac{n+2}{2}-\frac{n+1}{2}\right| + \dots + \left|n-\frac{n+1}{2}\right|\right]$

$$= \frac{2}{n}\underbrace{\left[\frac{1}{2}+\frac{3}{2}+\dots+\frac{n-1}{2}\right]}_{\left(\frac{n}{2}\right)\text{terms}} = \frac{1}{n}\cdot\left(\frac{n}{2}\right)^2$$

[$\because$ sum of first N odd natural numbers $= N^2$]

$$= \frac{1}{n}\cdot\frac{n^2}{4} = \frac{n}{4}$$

4. (c) Here, $m = \dfrac{\Sigma x_i}{5}$, $s = \sqrt{\dfrac{\Sigma x_i^2}{5} - \left(\dfrac{\Sigma x_i}{5}\right)^2}$

$$SD = \sqrt{\frac{k^2\Sigma x_i^2}{5} - \left(\frac{k\Sigma x_i}{5}\right)^2} = \sqrt{\frac{k^2\Sigma x_i^2}{5} - k^2\left(\frac{\Sigma x_i}{5}\right)^2}$$

$$= \sqrt{\left(\frac{\Sigma x_i^2}{5}\right) - \left(\frac{\Sigma x_i}{5}\right)^2} = ks$$

5. (c) Let us form the following table

x_i	f_i	f_ix_i	x_i^2	$f_ix_i^2$
3	7	21	9	63
8	10	80	64	640
13	15	195	169	2535
18	10	180	324	3240
23	6	138	529	3174
	48	**614**		**9652**

Now, we have

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

$$= \frac{1}{48}\sqrt{48 \times 9652 - (614)^2}$$

$$= \frac{1}{48}\sqrt{463296 - 376996}$$

$$= \frac{1}{48} \times 293.77 = 6.12$$

$\therefore$ Standard deviation $(\sigma) = 6.12$

6. (b) Mean $(\bar{x})$

$$= \frac{57 + 64 + 43 + 67 + 49 + 59 + 61 + 59 + 44 + 47}{10}$$

$$= \frac{550}{10} = 55$$

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

$$= \frac{2^2 + 9^2 + 12^2 + 12^2 + 6^2 + 4^2 + 6^2 + 4^2 + 11^2 + 8^2}{10}$$

$$= \frac{662}{10} = 66.2$$

Standard deviation $(\sigma) = \sqrt{\sigma^2} = \sqrt{66.2} = 8.13$

7. (d) We have the following numbers

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

If 1 is added to each number, we get

2, 3, 4, 5, 6, 7, 8, 9, 10, 11

Sum of these numbers, $\Sigma x_i = 2 + 3 + \dots + 11 = 65$

Sum of squares of these numbers,

$$\Sigma x_i^2 = 2^2 + 3^2 + \dots + 11^2 = 505$$

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

$$= \frac{505}{10} - (6.5)^2$$

$$= 50.5 - 42.25 = 8.25$$

8. (c) $\bar{x} = \dfrac{\text{Sum of quantities}}{n} = \dfrac{\frac{n}{2}(a + l)}{n}$

$$= \frac{1}{2}[1 + 1 + 100d] = 1 + 50d$$

$$\therefore \quad \text{MD}(\bar{x}) = \frac{1}{n}\Sigma|x_i - \bar{x}|$$

$$\Rightarrow \quad 255 = \frac{1}{101}[50d + 49d + \dots + d + 0 + d + \dots + 50d]$$

$$\Rightarrow \quad 255 = \frac{2d}{101}\left[\frac{50 \times 51}{2}\right] \Rightarrow d = \frac{255 \times 101}{50 \times 51} = 10.1$$

9. Let us make the following table from the given data.

x_i	f_i	f_ix_i	$(x_i - \bar{x})^2 = \left(x_i - \frac{22}{7}A\right)^2$	$(x_i - \bar{x})^2 f_i$
A	2	$2A$	$(225/49)A^2$	$(450/49)A^2$
$2A$	1	$2A$	$(64/49)A^2$	$(64/49)A^2$
$3A$	1	$3A$	$(1/49)A^2$	$(1/49)A^2$
$4A$	1	$4A$	$(36/49)A^2$	$(36/49)A^2$
$5A$	1	$5A$	$(169/49)A^2$	$(169/49)A^2$
$6A$	1	$6A$	$(400/49)A^2$	$(400/49)A^2$
	$N = \Sigma f_i = 7$	$\Sigma f_ix_i = 22A$		$\Sigma f_i(x_i - \bar{x})^2 = \frac{1120}{49}A^2$

Hence, $N = \Sigma f_i = 7$ and $\Sigma f_i x_i = 22A$

$$\therefore \quad \bar{x} = \frac{\sum_{i=1}^{6} f_i x_i}{N} = \frac{22A}{7}$$

Now, variance $= \dfrac{1}{N}\sum_{i=1}^{6} f_i(x_i - \bar{x})^2$

$\Rightarrow 160 = \dfrac{1}{7}\left(\dfrac{1120}{49}\right)A^2$ [$\because$ variance = 60, given]

Ans. $A = 7$

10. Here, $h = 10$ and $a = 55$

Let us make the following table from the given data.

Class interval	f_i	Mid value (x_i)	$u_i = \frac{x_i - 55}{10}$	f_iu_i	u_i^2	$f_iu_i^2$
0-10	3	5	−5	−15	25	75
10-20	2	15	−4	−8	16	32
20-30	4	25	−3	−12	9	36
30-40	6	35	−2	−12	4	24
40-50	5	45	−1	−5	1	5
50-60	5	55	0	0	0	0
60-70	5	65	1	5	1	5
70-80	2	75	2	4	4	8
80-90	8	85	3	24	9	72
90-100	5	95	4	20	16	80
	$N = \Sigma f_i = 45$			$\Sigma f_i u_i = 1$		$\Sigma f_i u_i^2 = 337$

Here, $N = \Sigma f_i = 63, \ \Sigma f_i u_i = 1, \ \Sigma f_i u_i^2 = 337$

$$\therefore \quad \text{Mean } (\bar{x}) = a + \left(\frac{1}{N}\Sigma f_i u_i\right) \times h$$

and standard deviation,

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

Ans. 55.22, 27.36

11. Let us make the following table from the given data.

Marks (Class -interval)	Number of students (f_i)	cf	x_i	$\lvert x_i - M\rvert = \lvert x_i - 28\rvert$	$f_i\lvert x_i - 28\rvert$
0-10	5	5	5	23	115
10-20	8	13	15	13	104
20-30	15	28	25	3	45
30-40	16	44	35	7	112
40-50	6	50	45	17	102
	$N = \Sigma f_i = 50$				$\Sigma f_i \lvert x_i - 28\rvert = 478$

Here, $N = 50$, so $\frac{N}{2} = 25$, $cf = 13$, $l = 20$, $f = 15$ and $h = 10$

Now, median $(M) = \dfrac{l + \frac{N}{2} - cf}{f} \times h = 20 + \dfrac{25-13}{15} \times 10 = 28$

$\therefore$ Mean deviation about median $= \frac{1}{N}(\Sigma f_i \lvert x_i - M\rvert)$

$$= \frac{478}{50} = 9.56$$

12. Let us make the following table from the given data.

x_i	f_i	$f_i x_i$	$(x_i - \bar{x}) = (x_i - 19)$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
6	2	12	−13	169	338
10	4	40	−9	81	324
14	7	98	−5	25	175
18	12	216	−1	1	12
24	8	192	5	25	200
28	4	112	9	81	324
30	3	90	11	121	363
	$\Sigma f_i = 40$	$\Sigma f_i x_i = 760$			$\Sigma f_i(x_i - \bar{x})^2 = 1736$

$\therefore \quad (\bar{x}) = \dfrac{\Sigma f_i x_i}{N} = \dfrac{760}{40} = 19$

and standard deviation $= \sqrt{\frac{1}{N}[\Sigma f_i (x_i - \bar{x})^2]}$

$$= \sqrt{\frac{1736}{40}} = \sqrt{43.4} = 6.59$$

Ans. 19, 6.59

13. Let us make the table from the given data.

x_i	f_i	$f_i x_i$	x_i^2	$f_i x_i^2$
3	7	21	9	63
8	10	80	64	640
13	15	195	169	2535
18	10	180	324	3240
23	6	138	529	3174
Total	$\Sigma f_i = 48$	$\Sigma f_i x_i = 614$		$\Sigma f_i x_i^2 = 9652$

$\therefore$ Standard deviation, $\sigma = \frac{1}{N}\sqrt{N\,\Sigma f_i x_i^2 - (\Sigma f_i x_i^2)^2}$

Ans. 6.12

14. Let assume mean, $a = 42.5$

Let us make the table from the given data.

Class	Mid value (x_i)	f_i	$u_i = \frac{x_i - A}{5}$	u_i^2	$f_i u_i$	$f_i u_i^2$
25-30	27.5	30	−3	9	−90	270
30-35	32.5	23	−2	4	−46	92
35-40	37.5	20	−1	1	−20	20
40-45	42.5	14	0	0	0	0
45-50	47.5	10	1	1	10	10
50-55	52.5	3	2	4	6	12
Total		$\Sigma f_i = 100$			$\Sigma u_i^2 = -140$	$\Sigma f_i u_i^2 = 404$

$\therefore \quad$ Mean, $\bar{x} = a + \frac{1}{N}\Sigma f_i u_i \times h = 42.5 + \frac{1}{100}(-140) \times 5$

$= 35.5$ yr

Standard deviation $= \left[\sqrt{\frac{1}{N}\Sigma f_i u_i^2 - \left(\frac{1}{N}\Sigma f_i u_i\right)^2}\right] \times h$

$$= \left[\sqrt{\frac{404}{100} - \left(\frac{-140}{100}\right)^2}\right] \times 5 = 7.21 \text{ yr}$$

15. Let us make the table from the given data.

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

Variance, $\sigma^2 = \left[\frac{1}{N}\Sigma f_i u_i^2 - \left(\frac{1}{N}\Sigma f_i u_i\right)^2\right] \times h^2$

$$= \left[\frac{130}{70} - \left(\frac{22}{70}\right)^2\right] \times 100 = 175.8$$

and standard deviation $= \sqrt{\text{Variance}} = \sqrt{175.8} = 13.259$

16. Let us make the table from the given data.

d	f	d^2	fd	fd^2
−2	1	4	−2	4
−1	6	1	−6	6
0	7	0	0	0
1	4	1	4	4
2	2	4	4	8
Total	$\Sigma f = 20$		$\Sigma fd = 0$	$\Sigma fd^2 = 22$

We know that, $\bar{x} = a + h\dfrac{\Sigma fd}{\Sigma f}$

$$\Rightarrow \quad 11 = a + h \times \frac{0}{20} \Rightarrow a = 11$$

Also, variance, $\sigma^2 = h^2\left[\dfrac{\Sigma fd^2}{\Sigma f} - \left(\dfrac{\Sigma fd}{\Sigma f}\right)^2\right]$

$$\Rightarrow \quad (9.9)^2 = h^2\left[\frac{22}{20}\right] \Rightarrow h = 3$$

Mid value, $\quad d = \dfrac{x-a}{h} \Rightarrow x = a + dh$

$\therefore$ Different values of x for different values of d, we have

$11 - 2\times 3, 11 - 1\times 3, 11 - 0\times 3, 11 + 1\times 3, 11 + 2\times 3$

i.e. 5, 8, 11, 14, 17.

Thus, the original frequency table is as follows

Class	3.5-6.5	6.5-9.5	9.5-12.5	12.5-15.5	15.5-18.5
Frequency	1	6	7	4	2

17. Let the remaining two observations be a and b.

Given, $\bar{x} = 9$ and $\sigma^2 = 9.25 \quad \because \bar{x} = 9$

$$\Rightarrow \quad \frac{\text{Sum of all observations}}{\text{Number of observations}} = 9$$

$$\Rightarrow \quad \text{Sum of all observations} = 9 \times 8$$

$$\Rightarrow \quad 6 + 7 + 10 + 12 + 12 + 13 + a + b = 72$$

$$\Rightarrow \quad a + b = 12 \qquad \text{...(i)}$$

Again, $\sigma^2 = \dfrac{\Sigma x_i^2}{n} - \left(\dfrac{\Sigma x_i}{n}\right)^2 \Rightarrow \sigma^2 = \dfrac{\Sigma x_i^2}{n} - (\bar{x})^2$

$$\Rightarrow 9.25 = \frac{36 + 49 + 100 + 144 + 144 + 169 + a^2 + b^2}{8} - (9)^2$$

$$\Rightarrow \quad a^2 + b^2 = 80 \qquad \text{...(ii)}$$

Now, from Eq. (i), put $b = (12 - a)$ in Eq. (ii), we get

$$a^2 + (12 - a)^2 = 80 \Rightarrow a = 4 \text{ or } a = 8$$

On putting $a = 4$ or $a = 8$ Eq. (i), we get $b = 8$ or $b = 4$

Hence, two observations are 4 and 8.

18. (i) (a) We have, $n = 200$, incorrect mean $= 40$

and incorrect standard deviation $= 15$

Now, incorrect mean $= 40$

$$\Rightarrow \quad \frac{\text{Incorrect } \Sigma x_i}{200} = 40$$

$$\text{Incorrect } \Sigma x_i = 8000$$

$$\Rightarrow \quad \text{Correct } \Sigma x_i = 8000 - (34 + 53) + (43 + 35)$$

$$= 8000 - 87 + 78 = 7991$$

(ii) (b) Correct mean $= \dfrac{7991}{200} = 39.955$

(iii) (a) $\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$

(iv) (c) Incorrect SD $= 15$

$\Rightarrow$ Incorrect variance $= (15)^2 = 225$

$$\Rightarrow \frac{\text{Incorrect } \Sigma x_i^2}{200} - (\text{Incorrect mean})^2 = 225$$

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

$$\Rightarrow \quad \text{Incorrect } \Sigma x_i^2 = 200(1600 + 225)$$

$$= 200 \times 1825 = 365000$$

$$\Rightarrow \quad \text{Correct } \Sigma x_i^2 = \text{Incorrect } \Sigma x_i^2 - (34^2 + 53^2) + (43^2 + 35^2)$$

$$= 365000 - 3965 + 3074 = 364109$$

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

$$= \frac{1}{200}(364109) - \left(\frac{7991}{200}\right)^2$$

$$= 1820.545 - 1596.402 = 224.143$$

(v) (a) Correct standard deviation $= \sqrt{\text{correct variance}}$

$$= \sqrt{224.143} \text{ [using part (iv)]}$$

$$= 14.971$$