311en17

17

MEASURES OF DISPERSION

You have learnt various measures of central tendency. Measures of central tendency help us to represent the entire mass of the data by a single value.

Can the **central tendency** describe the data fully and adequately?

In order to understand it, let us consider an example.

The daily income of the workers in two factories are :

Factory A	:	35	45	50	65	70	90	100
Factory B	:	60	65	65	65	65	65	70

Here we observe that in both the groups the mean of the data is the same, namely, 65

(i) In group A, the observations are much more scattered from the mean.

(ii) In group B, almost all the observations are concentrated around the mean.

Certainly, the two groups differ even though they have the same mean.

Thus, there arises a need to differentiate between the groups. We need some other measures which concern with the measure of scatteredness (or spread).

To do this, we study what is known as **measures of dispersion.**

OBJECTIVES

After studying this lesson, you will be able to :

- explain the meaning of dispersion through examples;
- define various measures of dispersion – range, mean deviation, variance and standard deviation;
- calculate mean deviation from the mean of raw and grouped data;
- calculate mean deviation from the median of raw and grouped data.
- calculate variance and standard deviation of raw and grouped data; and
- illustrate the properties of variance and standard deviation.
- Analyses the frequencys distributions with equal means.

EXPECTED BACKGROUND KNOWLEDGE

- Mean of grouped data
- Median of ungrouped data

17.1 MEANING OF DISPERSION

To explain the meaning of dispersion, let us consider an example.

Two sections of 10 students each in class X in a certain school were given a common test in Mathematics (maximum marks 40). The scores of the students are given below :

Section A :	6	9	11	13	15	21	23	28	29	35
Section B:	15	16	16	17	18	19	20	21	23	25

The average score in section A is 19.

The average score in section B is 19.

Let us construct a dot diagram, on the same scale for section A and section B (see Fig. 17.1)

The position of mean is marked by an arrow in the dot diagram.

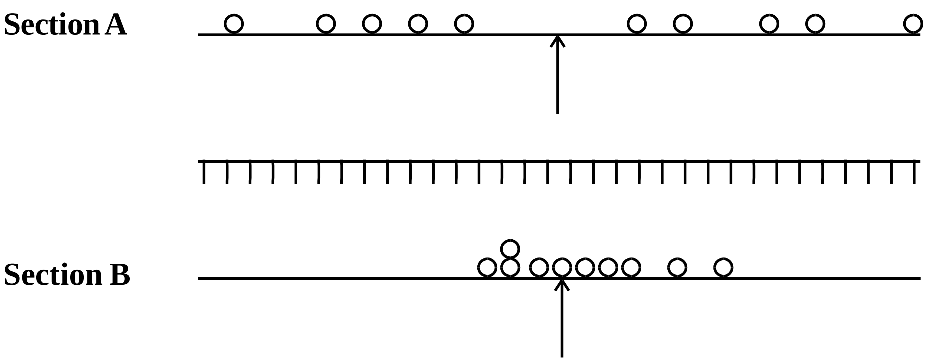

Fig. 17.1

Clearly, the extent of spread or dispersion of the data is different in section A from that of B. The measurement of the scatter of the given data about the average is said to be a measure of dispersion or scatter.

In this lesson, you will read about the following measures of dispersion :

(a) Range

(b) Mean deviation from mean

(c) Mean deviation from median

(d) Variance

(e) Standard deviation

17.2 DEFINITION OF VARIOUS MEASURES OF DISPERSION

(a) **Range :** In the above cited example, we observe that

(i) the scores of all the students in section A are ranging from 6 to 35;

(ii) the scores of the students in section B are ranging from 15 to 25.

The difference between the largest and the smallest scores in section A is 29 (35 – 6)

The difference between the largest and smallest scores in section B is 10 (25 – 15).

Thus, the difference between the largest and the smallest value of a data, is termed as the range of the distribution.

(b) **Mean Deviation from Mean :** In Fig. 17.1, we note that the scores in section B cluster around the mean while in section A the scores are spread away from the mean. Let us take the deviation of each observation from the mean and add all such deviations. If the sum is 'large', the dispersion is 'large'. If, however, the sum is 'small' the dispersion is small.

Let us find the sum of deviations from the mean, i.e., 19 for scores in section A.

Observations (x_i)	Deviations from mean $(x_i - \overline{x})$
6	–13
9	–10
11	–8
13	–6
15	–4
21	+2
23	+4
28	+9
29	+10
35	16
190	0

Here, the sum is zero. It is neither 'large' nor 'small'. Is it a coincidence ?

Let us now find the sum of deviations from the mean, i.e., 19 for scores in section B.

Observations (x_i)	Deviations from mean $(x_i - \overline{x})$
15	–4
16	–3
16	–3
17	–2
18	–1
19	0
20	1
21	2
23	4
25	6
190	0

Again, the sum is zero. Certainly it is not a coincidence. In fact, we have proved earlier that **the sum of the deviations taken from the mean is always zero for any set of data.** Why is the sum always zero ?

On close examination, we find that the signs of some deviations are positive and of some other deviations are negative. Perhaps, this is what makes their sum always zero. In both the cases, we get sum of deviations to be zero, so, we cannot draw any conclusion from the sum of deviations. But this can be avoided if we take only the **absolute value of the deviations** and then take their sum.

If we follow this method, we will obtain a measure (descriptor) called the mean deviation from the mean.

The mean deviation is the sum of the absolute values of the deviations from the mean divided by the number of items, (i.e., the sum of the frequencies).

(c) **Variance :** In the above case, we took the absolute value of the deviations taken from mean to get rid of the negative sign of the deviations. Another method is to square the deviations. Let us, therefore, square the deviations from the mean and then take their sum. If we divide this sum by the number of observations (i.e., the sum of the frequencies), we obtain the average of deviations, which is called variance. **Variance is usually denoted by** σ^2.

(d) **Standard Deviation :** If we take the positive square root of the variance, we obtain the root mean square deviation or simply called standard deviation and is denoted by σ.

17.3 MEAN DEVIATION FROM MEAN OF RAW AND GROUPED DATA

Mean Deviation from mean of raw data $= \dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{N}$

Mean deviation from mean of grouped data $= \dfrac{\sum_{i=1}^{n} [f_i |x_i - \bar{x}|]}{N}$

where $N = \sum_{i=1}^{n} f_i, \bar{x} = \dfrac{1}{N} \sum_{i=1}^{n} (f_i x_i)$

The following steps are employed to calculate the mean deviation from mean.

Step 1 : Make a column of deviation from the mean, namely $x_i - \bar{x}$ (In case of grouped data take x_i as the mid value of the class.)

Step 2 : Take absolute value of each deviation and write in the column headed $|x_i - \bar{x}|$.

For calculating the mean deviation from the mean of raw data use

Mean deviation of Mean $= \dfrac{\sum_{i=1}^{n} |x_i - \bar{x}|}{N}$

For grouped data proceed to step 3.

Step 3 : Multiply each entry in step 2 by the corresponding frequency. We obtain $f_i (x_i - \bar{x})$ and write in the column headed $f_i |x_i - \bar{x}|$.

Step 4 : Find the sum of the column in step 3. We obtain $\sum_{i=1}^{n}[f_i\,|x_i - \bar{x}|]$

Step 5 : Divide the sum obtained in step 4 by N.

Now let us take few examples to explain the above steps.

Example 17.1 Find the mean deviation from the mean of the following data :

Size of items x_i	4	6	8	10	12	14	16
Frequency f_i	2	5	5	3	2	1	4

Mean is 10

Solution :

x_i	f_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$
4	2	–5.7	5.7	11.4
6	4	–3.7	3.7	14.8
8	5	–1.7	1.7	8.5
10	3	0.3	0.3	0.9
12	2	2.3	2.3	4.6
14	1	4.3	4.3	4.3
16	4	6.3	6.3	25.2
	21			69.7

$$\text{Mean deviation from mean} = \frac{\sum[f_i\,|x_i - \bar{x}|]}{21} = \frac{69.7}{21} = 3.319$$

Example 17.2 Calculate the mean deviation from mean of the following distribution :

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of Students	5	8	15	16	6

Mean is 27 marks

Solution :

Marks	Class Marks x_i	f_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$	$f_i\lvert x_i - \bar{x}\rvert$
0 – 10	5	5	–22	22	110
10 – 20	15	8	–12	12	96
20 – 30	25	15	–2	2	30
30 – 40	35	16	8	8	128
40 – 50	45	6	18	18	108
Total		50			472

$$\text{Mean deviation from Mean} = \frac{\sum [f_i |x_i - \bar{x}|]}{N} = \frac{472}{50} \text{ Marks} = 9.44 \text{ Marks}$$

CHECK YOUR PROGRESS 17.1

1. The ages of 10 girls are given below :

   3 5 7 8 9 10 12 14 17 18

   What is the range ?

2. The weight of 10 students (in Kg) of class XII are given below :

   45 49 55 43 52 40 62 47 61 58

   What is the range ?

3. Find the mean deviation from mean of the data

   45 55 63 76 67 84 75 48 62 65

   Given mean = 64.

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

Salary (in rupees)	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
No. of employees	4	6	8	12	7	6	4	3

Given mean = Rs. 57.2

5. Calculate the mean deviation for the following data of marks obtained by 40 students in a test

Marks obtained	20	30	40	50	60	70	80	90	100
No. of students	2	4	8	10	8	4	2	1	1

6. The data below presents the earnings of 50 workers of a factory

Earnings (in rupees) :	1200	1300	1400	1500	1600	1800	2000
No. of workers :	4	6	15	12	7	4	2

   Find mean deviation.

7. The distribution of weight of 100 students is given below :

Weight (in Kg)	50 – 55	55 – 60	60 – 65	65 – 70	70 – 75	75 – 80
No. of students	5	13	35	25	17	5

   Calculate the mean deviation.

8. The marks of 50 students in a particular test are :

Marks	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
No. of students	4	6	9	12	8	6	4	1

Find the mean deviation for the above data.

17.4 MEDIAN

17.4.1 MEDIAN OF GROUPED DATA

Median of Discrete Frequency Distribution :

Step 1 : Arrange the data in ascending order.

Step 2 : Find cumulative frequencies

Step 3 : Find $\frac{N}{2}$

Step 4 : The observation whose cumulative frequency is just greater than $\frac{N}{2}$ is the median of the data.

Example 17.3 Find the median of the data

x_i	8	9	10	12	14	16
f_i	6	2	2	2	6	8

Solution : The given data are already in ascending order. Let us now write the cumulative frequencies of observations

x_i	8	9	10	12	14	16
f_i	6	2	2	2	6	8
c.f.	6	8	10	12	18	26

$$N = 26, \qquad \therefore \frac{N}{2} = 13.$$

The observation whose c.f. is just greater than 13 is 14 (whose c.f. is 18)

$\therefore$ Median = 14.

17.4.2 MEDIAN OF CONTINUOUS FREQUENCY DISTRIBUTION

Step 1 : Arrange the data in ascending order

Step 2 : Write cumulative frequencies of the observations

Step 3 : Identify the class whose cumulative frequency is just greater than $\frac{N}{2}$. Call this class-interval as median class.

Step 4 : Find median by the formula

$$\text{Median} = l + \frac{\frac{N}{2} - C}{f} \times i$$

Where

$l \rightarrow$ Lower limit of the median class
$N \rightarrow$ Number of observations N = Σf_i
$C \rightarrow$ Cumulative frequency of the class just preceding the median class
$f \rightarrow$ Frequency of the median class
$i \rightarrow$ Width of the median class

Example 17.4 Find the median marks obtained by 50 students from the following distribution :

Marks	0-10	10-20	20-30	30-40	40-50
Number of Students	8	8	14	16	4

Solution : The given intervals are already in ascending order. The following table has the row corresponding to the cumulative frequencies.

Marks	0-10	10-20	20-30	30-40	40-50
Number of students	8	8	14	16	4
Cummulative frequency	8	16	30	46	50

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

The class corresponding to the c.f. just greater than 25 is 20-30.

$\therefore$ Median class is 20-30

where $l = 20$, N = 50, C = 16, $f = 14$, $i = 10$.

$$\therefore \quad \text{Median} = l + \frac{\frac{N}{2} - C}{f} \times i = 20 + \frac{25 - 16}{14} \times 10$$

$$= 20 + \frac{9}{14} \times 10 = 20 + 6.43 = 26.43$$

Example 17.5 Find the median of the following:

Marks	Number of Students
0 – 9	3
10 – 19	5
20 – 29	8
30 – 39	9
40 – 49	13
50 – 59	6

Solution : The given class intervals are inclusive series Before finding the median we have to convert the inclusive series into exclusive series.

Method of converting an inclusive series into exclusive series.

(1) Find the half of the difference between the upper limit of a class and the lower limit of its succeeding (next) class.

(2) Subtract this half from the lower limit and add into the upper limit.

Mark	Exclusive Series	f.	c.f.
0-9	0.5-9.5	3	3
10-19	9.5-19.5	5	8
20-29	19.5-29.5	8	16
30-39	29.5-39.5	9	25
40-49	39.5-49.5	13	38
50-59	49.5-59.5	6	44

$$\frac{N}{2} = \frac{44}{2} = 22$$

$\therefore$ Median class is 29.5 – 39.5 as its c.f. is 25, which is just greater than 22.

Now, $l = 29.5$, N = 44, C = 16, f = 9, $i = 39.5 - 29.5 = 10$

$$\therefore \quad \text{Median} = l + \frac{\frac{N}{2} - C}{f} \times i = 29.5 + \frac{22 - 16}{9} \times 10$$

$$= 29.5 + \frac{6}{9} \times 10 = 29.5 + \frac{20}{3}$$

$$= 29.5 + 6.66 = 36.16$$

CHECK YOUR PROGRESS 17.2

Find the median of the following data :

1.

x_i	6	11	16	21	26
f_i	5	3	6	4	7

2.

x_i	5	10	15	20	25
f_i	5	25	29	17	9

3.

Marks	0-5	5-10	10-15	15-20	20-25
Number of Boys	5	9	10	14	12

4.

Age (in years)	17-21	21-26	26-31	31-36	36-41
Number of Boys	5	6	12	7	4

17.5 MEAN DEVIATION FROM MEDIAN

We know that for observations in data the central tendency give us the values about which the data concentrate or cluster. It is also essential to know that how far all observation are, from a measure of central tendency. In other words, in data it is required to know how dispersed the observations are from a given point (or a measure of central tendency). In most of the cases mean deviation from mean and median give us the desired disperson or deviation of the observations. Recall that mean deviation for data is defined as the mean of the absolute values of deviations from '*a*'.

Recall that the deviation of an observation x from a fixed point 'a' is the difference $x - a$.

So mean deviation about 'a' denoted by M.D (a) is given by

$$\text{M.D. (a)} = \frac{\text{Sum of the absolute values of deviations from 'a'}}{\text{Number of observations}}$$

Methematically we can write

$$\text{M.D.}(a) = \frac{\sum_{i=1}^{n} |x_i - a|}{n}$$

Like wise

$$\text{M.D. (Mean} = \overline{\text{X}}\text{)} = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n}$$

and

$$\text{M.D.(Median M)} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \text{M}|$$

Example 17.6 Find mean deviation about median for the observation

7, 10, 15, 16, 8, 9, 5, 17, 14

Solution : In order to find median, arrange the given values in ascending order, so we have

5, 7, 8, 9, 10, 14, 15, 16, 17,

> Algorithm to find mean deviation about mean/median :
>
> **Step 1 :** Calculate the mean or median of the data
>
> **Step 2 :** Find deviations of each observation x_i from mean/median
>
> **Step 3 :** Find the absolute values of the deviations.
> Assolute values can be obtained by dropping the minus sign if it is there
>
> **Step 4 :** Calculate the mean of the obsolute values of the deviations. This mean will be the required Mean deviation.

$n = 9,$

$$\text{Median} = \frac{n+1}{2}\text{th observation}$$

= 5th observation

M = 10.

Deviations of the observation from median i.e. 10 are

	5–10	7–10	8–10	9–10	10–10	14–10	15–10	16–10	17–10
i.e x_i-M are	–5	–3	–2	–1	0	4	5	6	7

Absolute values of the deviations i.e. $|x_i - \text{M}|$ are

5, 3, 2, 1, 0, 4, 5, 6, 7

Now $$\text{M.D. (M)} = \frac{\sum_{i=1}^{n} |x_i - \text{M}|}{n}$$

$$= \frac{5+3+2+1+0+4+5+6+7}{10} = \frac{33}{10} = 3.3.$$

17.5.1 MEAN DEVIATION OF GROUPED DATA FROM MEDIAN

Recall that data presented in the following form are called grouped data

(a) Discrete frequency distribution

Observation :	x_1	x_2	x_3	...	x_n
Frequencies :	f_1	f_2	f_3	...	f_n

(b) Continuous frequency distribution :

Observations	$l_1 - u_1$	$l_2–u_2$	$l_3 - u_3$	...	$l_n–u_n$
Frequencies	f_1	f_2	f_3	...	f_n

For example, marks obtained by 50 students

Marks	0-5	5-10	10-15	15-20	20-25	25-30
Number of Students	8	6	12	10	10	4

Let us now learn to find mean deviation about median by following examples.

Example 17.7 Find the mean deviation about the median for the following data :

x_i	25	20	15	10	5
f_i	7	4	6	3	5
c.f.	7	11	17	20	25

Here N = 25, and we know that median is the $\frac{25+1}{2}$ = 13th observation. This observation lies in the C.f 17, for which corresponding observation is 15.

∴ Median M = 15

Now deviations and their absolute values are given in following table.

x_i	f_i	x_i – M	$\|x_i$ – M$\|$	$f_i\|x_i$ – M$\|$
25	7	25 – 15 = 10	10	7 × 10 = 70
20	4	20 – 15 = 5	5	4 × 5 = 20
15	6	15 – 15 = 0	0	6 × 0 = 0
10	3	10 – 15 = – 5	5	3 × 5 = 15
5	5	5 – 15 = – 10	10	5 × 10 = 50
	N= Σf_i = 25			$\Sigma f_i\ \|x_i - M\|$ = 155

$$\therefore \quad \text{Mean Deviation (M)} = \frac{\sum_{i=1}^{n} f_i\,|x_i - M|}{\sum_{i=1}^{n} f_i} = \frac{155}{25} = 6.2$$

Example 17.8 Find the mean deviation about median for the following data :

Heights (in cm)	95-105	105-115	115-125	125-135	135-145	145-155
Number of Girls	9	15	23	30	13	10

Solution : Let us first find median :

Height (in cm)	Number of Girls (f)	Cumulative frequncy (c.f)
95-105	9	9
105-115	15	24
115-125	23	47
125-135	30	77
135-145	13	90
145-155	10	100

$$N = 100 \Rightarrow \frac{N+1}{2} = \frac{101}{2} = 50.5$$

$\frac{N}{2} = 50.5$ lies in c.f. 77.

∴ Median class is corresponding to the c.f. 77 i.e., 125 – 135

Now, Median = $l+\dfrac{\frac{N}{2}-C}{f}\times i$

where l = lower limit of the median class

N = sum of frequencies

C = c.f. of the class just preceding the median class

f = frequency of the median class

and i = width or class-size of the median class

Here, l = 125, N = 100, C = 47, f = 30, i = 10

$$\therefore \quad M = 125+\frac{50-47}{30}\times 10 = 125+\frac{3}{3}=126$$

To find mean deviation let us form the following table :

Height (in cm)	Number of Girls (f)	Mid-value of the heights	Absolute Deviation (x_i – M|)	$f_i\|x_i$ – M|
95-105	9	100	|100-126| = 26	9 × 26 = 234
105-115	15	110	|110-126| = 16	15 × 16 = 240
115-125	23	120	|120 – 126| = 6	23 × 6 = 138
125-135	30	130	|130-126| = 4	30 × 4 = 120
135-145	13	140	|140-126| = 14	13 × 14 = 182
145-155	10	150	|150-126| = 24	10 × 24 = 240
	Σf_i = 100			Σf_i |xi – M| = 1154

$$\therefore \quad \text{Mean Deviation (Median)} = \text{M.D.(M)} = \frac{\sum_{i=1}^{n} f_i\,|x_i-M|}{\sum_{i=1}^{n} f_i} = \frac{1154}{100} = 11.54.$$

17.5.2 STEP TO FIND MEAN DEVIATION FROM MEDIAN OF A CONTINUOUS FREQUENCY DISTRIBUTION.

Step 1 : Arrange the intervals in ascending order

Step 2 : Write cumulative frequencies

Step 3 : Identify the median class, as the class having c.f. just greater than $\frac{N}{2}$, where N is the total number of obsservations (i.e. sum of all frequencies)

Step 4 : Find the corresponding values for the median class and put in the formula :

$$\text{Median} = l + \frac{\frac{N}{2} - C}{f} \times i$$

where

$l \rightarrow$ lower limit of the madian class

$N \rightarrow$ Sum of frequencies

$C \rightarrow$ c.f. of the class just preceding the median class

$f \rightarrow$ frequency of the median class

$i \rightarrow$ width of the median class

Step 5 : Now form the table for following columns :

Given intervals	Frequencies	Mid-value x_i	Absolute Deviation from Median $\|x_i - M)$	$f_i\|x_i - M\|$

Step 6 : Now calculate

$$\text{M.D.(M)} = \frac{\sum_{i=1}^{n} f_i \, | x_i - M |}{\sum_{i=1}^{n} f_i}$$

CHECK YOUR PROGRESS 17.3

Find the mean deviation about median of the following data.

1.

x_i	11	12	13	14	16	17	18
f_i	2	3	2	3	1	2	1

2.

x_i	3	6	7	9	11	13
f_i	3	9	11	8	9	6

3.

Weight (in kg)	40-42	42-44	44-46	46-48	48-50
No. of Students	9	13	24	28	6

4.

Age (in years)	0-1	1-2	2-3	3-4	4-5
No. of Children given polio drops	100	155	210	315	65

17.6 VARIANCE AND STANDARD DEVIATION OF RAW DATA

If there are n observations, $x_1, x_2...., x_n$, then

$$\text{Variance}\left(\sigma^2\right) = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + ... + (x_n - \overline{x})^2}{n}$$

or $$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}; \text{ where } \overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

The standard deviation, denoted by σ, is the positive square root of σ^2. Thus

$$\sigma = +\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}$$

The following steps are employed to calculate the variance and hence the standard deviation of raw data. The mean is assumed to have been calculated already.

Step 1 : Make a column of deviations from the mean, namely, $x_i - \overline{x}$.

Step 2 (check) : Sum of deviations from mean must be zero, i.e., $\sum_{i=1}^{n}(x_i - \overline{x}) = 0$

Step 3: Square each deviation and write in the column headed $(x_i - \overline{x})^2$.

Step 4 : Find the sum of the column in step 3.

Step 5 : Divide the sum obtained in step 4 by the number of observations. We obtain σ^2.

Step 6 : Take the positive square root of σ^2. We obtain σ (Standard deviation).

Example 17.9 The daily sale of sugar in a certain grocery shop is given below :

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
75 kg	120 kg	12 kg	50 kg	70.5 kg	140.5 kg

The average daily sale is 78 Kg. Calculate the variance and the standard deviation of the above data.

Solution : $\bar{x} = 78$ kg (Given)

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
75	– 3	9
120	42	1764
12	– 66	4356
50	– 28	784
70.5	– 7.5	56.25
140.5	62.5	3906.25
	0	10875.50

Thus $\sigma^2 = \dfrac{\sum_{i=1} (x_i - \bar{x})^2}{n} = \dfrac{10875.50}{6} = 1812.58$ (approx.)

and $\sigma = 42.57$ (approx.)

Example 17.10 The marks of 10 students of section A in a test in English are given below :

7 10 12 13 15 20 21 28 29 35

Determine the variance and the standard deviation.

Solution : Here $\bar{x} = \dfrac{\sum x_i}{10} = \dfrac{190}{10} = 19$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
7	– 12	144
10	– 9	81
12	– 7	49
13	– 6	36
15	– 4	16
20	+1	1
21	+2	4
28	+9	81
29	+10	100
35	+16	256
	0	768

Thus $\sigma^2 = \dfrac{768}{10} = 76.8$ and $\sigma = +\sqrt{76.8} = 8.76$ (approx)

CHECK YOUR PROGRESS 17. 4

1. The salary of 10 employees (in rupees) in a factory (per day) is

   50 60 65 70 80 45 75 90 95 100

   Calculate the variance and standard deviation.

2. The marks of 10 students of class X in a test in English are given below :

   9 10 15 16 18 20 25 30 32 35

   Determine the variance and the standard deviation.

3. The data on relative humidity (in %) for the first ten days of a month in a city are given below:

   90 97 92 95 93 95 85 83 85 75

   Calculate the variance and standard deviation for the above data.

4. Find the standard deviation for the data

   4 6 8 10 12 14 16

5. Find the variance and the standard deviation for the data

   4 7 9 10 11 13 16

6. Find the standard deviation for the data.

   40 40 40 60 65 65 70 70 75 75 75 80 85 90 90 100

17. 7 STANDARD DEVIATION AND VARIANCE OF RAW DATA AN ALTERNATE METHOD

If $\overline{x}$ is in decimals, taking deviations from $\overline{x}$ and squaring each deviation involves even more decimals and the computation becomes tedious. We give below an alternative formula for computing σ^2. In this formula, we by pass the calculation of $\overline{x}$.

We know
$$\sigma^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n} = \sum_{i=1}^{n} \frac{x_i^2 - 2x_i\overline{x} + \overline{x}^2}{n}$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n} - \frac{2\overline{x}\sum_{i=1}^{n} x_i}{n} + \overline{x}^2 = \frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2, \quad \left(\because \overline{x} = \frac{\sum x_i}{n}\right)$$

i.e.
$$\sigma^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}$$

And $\sigma = +\sqrt{\sigma^2}$

The steps to be employed in calculation of σ^2 and, hence σ by this method are as follows :

Step 1 : Make a column of squares of observations i.e. x_i^2.

Step 2 : Find the sum of the column in step 1. We obtain $\sum_{i=1}^{n} x_i^2$

Step 3 : Substitute the values of $\sum_{i=1}^{n} x_i^2$, n and $\sum_{i=1}^{n} x_i$ in the above formula. We obtain σ^2.

Step 4 : Take the positive sauare root of σ^2. We obtain σ.

Example 17.11 We refer to Example 17.10 of this lesson and re-calculate the variance and standard deviation by this method.

Solution :

x_i	x_i^2
7	49
10	100
12	144
13	169
15	225
20	400
21	441
28	784
29	841
35	1225
190	4378

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

$$= \frac{4378 - \frac{(190)^2}{10}}{10}$$

$$= \frac{4378 - 3610}{10} = \frac{768}{10} = 76.8$$

and $\sigma = +\sqrt{76.8} = 8.76$ (approx)

We observe that we get the same value of σ^2 and σ by either methods.

17.8 STANDARD DEVIATION AND VARIANCE OF GROUPED DATA : METHOD - I

We are given k classes and their corresponding frequencies. We will denote the variance and the standard deviation of grouped data by σ_g^2 and σ_g respectively. The formulae are given below :

$$\sigma_g^2 = \frac{\sum_{i=1}^{K}\left[f_i(x_i - \bar{x})^2\right]}{N}, \quad N = \sum_{i=1}^{K} f_i \quad \text{and} \quad \sigma_g = +\sqrt{\sigma_g^2}$$

The following steps are employed to calculate σ_g^2 and, hence σ_g : (The mean is assumed to have been calculated already).

Step 1 : Make a column of class marks of the given classes, namely x_i

Step 2 : Make a column of deviations of class marks from the mean, namely, $x_i - \bar{x}$. Of course the sum of these deviations need not be zero, since x_i's are no more the original observations.

Step 3 : Make a column of squares of deviations obtained in step 2, i.e., $(x_i - \bar{x})^2$ and write in the column headed by $(x_i - \bar{x})^2$.

Step 4 : Multiply each entry in step 3 by the corresponding frequency.

We obtain $f_i(x_i - \bar{x})^2$.

Step 5 : Find the sum of the column in step 4. We obtain $\sum_{i=1}^{k}\left[f_i(x_i - \bar{x})^2\right]$

Step 6 : Divide the sum obtained in step 5 by N (total no. of frequencies). We obtain σ_g^2.

Step 7 : $\sigma_g = +\sqrt{\sigma_g^2}$

Example 17.12 In a study to test the effectiveness of a new variety of wheat, an experiment was performed with 50 experimental fields and the following results were obtained :

Yield per Hectare (in quintals)	Number of Fields
31 – 35	2
36 – 40	3
41 – 45	8
46 – 50	12
51 – 55	16
56 – 60	5
61 – 65	2
66 – 70	2

The mean yield per hectare is 50 quintals. Determine the variance and the standard deviation of the above distribution.

Solution :

Yield per Hectare (in quintal)	No. of Fields	Class Marks	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
31 – 35	2	33	– 17	289	578
36 – 40	3	38	– 12	144	432
41 – 45	8	43	– 7	49	392
46 – 50	12	48	– 2	4	48
51 – 55	16	53	+3	9	144
56 – 60	5	58	+8	64	320
61 – 65	2	63	+13	169	338
66 – 70	2	68	+18	324	648
Total	50				2900

Thus $\sigma_g^2 = \dfrac{\sum_{i=1}^{n}\left[f_i(x_i - \bar{x})^2\right]}{N} = \dfrac{2900}{50} = 58$ and $\sigma_g = +\sqrt{58} = 7.61$ (approx)

17.9 STANDARD DEVIATION AND VARIANCE OF GROUPED DATA : – METHOD - II

If $\bar{x}$ is not given or if $\bar{x}$ is in decimals in which case the calculations become rather tedious, we employ the alternative formula for the calculation of σ_g^2 as given below:

$$\sigma_g^2 = \frac{\sum_{i=1}^{k}\left[f_i x_i^2\right] - \dfrac{\left(\sum_{i=1}^{k}\left[f_i x_i\right]\right)^2}{N}}{N}, \qquad N = \sum_{i=1}^{k} f_i$$

and $$\sigma_g = +\sqrt{\sigma_g^2}$$

The following steps are employed in calculating σ_g^2, and, hence σ_g by this method:

Step 1 : Make a column of class marks of the given classes, namely, x_i.

Step 2 : Find the product of each class mark with the corresponding frequency. Write the product in the column $x_i f_i$.

Step 3 : Sum the entries obtained in step 2. We obtain $\sum_{i=1}^{k}(f_i x_i)$.

Step 4 : Make a column of squares of the class marks of the given classes, namely, x_i^2.

Step 5 : Find the product of each entry in step 4 with the corresponding frequency. We obtain $f_i {x_i}^2$.

Step 6 : Find the sum of the entries obtained in step 5. We obtain $\sum_{i=1}^{k}(f_i x_i^2)$.

Step 7 : Substitute the values of $\sum_{i=1}^{k}(f_i x_i^2)$, N and $\left(\sum_{i=1}^{k}(f_i x_i)\right)$ in the formula and obtain σ_g^2.

Step 8 : $\sigma_g = +\sqrt{\sigma_g^2}$.

Example 17.13 Determine the variance and standard deviation for the data given in Example 17.12 by this method.

Solution :

Yields per Hectare (in quintals)	f_i	x_i	$f_i x_i$	${x_i}^2$	$f_i {x_i}^2$
31 – 35	2	33	66	1089	2178
36 – 40	3	38	114	1444	4332
41 – 45	8	43	344	1849	14792
46 – 50	12	48	576	2304	27648
51 – 55	16	53	848	2809	44944
56 – 60	5	58	290	3364	16820
61 – 65	2	63	126	3969	7938
66 – 77	2	68	136	4624	9248
Total	50		2500		127900

Substituting the values of $\sum_{i=1}^{k}\left(f_i x_i^2\right)$, N and $\sum_{i=1}^{k}\left(f_i x_i\right)$ in the formula, we obtain

$$\sigma_g^2 = \frac{127900 - \frac{(2500)^2}{50}}{50} = \frac{2900}{50} = 58$$

and $\qquad \sigma_g = +\sqrt{58} = 7.61$ (approx.)

Again, we observe that we get the same value of σ_g^2, by either of the methods.

CHECK YOUR PROGRESS 17.5

1. In a study on effectiveness of a medicine over a group of patients, the following results were obtained :

Percentage of relief	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of patients	10	10	25	15	40

   Find the variance and standard deviation.

2. In a study on ages of mothers at the first child birth in a village, the following data were available :

Age (in years) at first child birth	18 – 20	20 – 22	22 – 24	24 – 26	26 – 28	28 – 30	30 – 32
No. of mothers	130	110	80	74	50	40	16

   Find the variance and the standard deviation.

3. The daily salaries of 30 workers are given below:

Daily salary (In Rs.)	0 – 50	50 – 100	100 – 150	150 – 200	200 – 250	250 – 300
No. of workers	3	4	5	7	8	3

   Find variance and standard deviation for the above data.

17.10 STANDARD DEVIATION AND VARIANCE: STEP DEVIATION METHOD

In Example 17.12, we have seen that the calculations were very complicated. In order to simplify the calculations, we use another method called the step deviation method. In most of the frequency distributions, we shall be concerned with the equal classes. Let us denote, the class size by h.

Now we not only take the deviation of each class mark from the arbitrary chosen 'a' but also divide each deviation by h. Let

$$u_i = \frac{x_i - a}{h} \qquad .....(1)$$

Then $\qquad x_i = hu_i + a \qquad .....(2)$

We know that $\qquad \bar{x} = h\bar{u} + a \qquad .....(3)$

Subtracting (3) from (2), we get

$$x_i - \bar{x} = h(u_i - \bar{u}) \qquad .....(4)$$

In (4), squaring both sides and multiplying by f_i and summing over k, we get

$$\sum_{i=1}^{k}\left[f_i(x_i - \bar{x})^2\right] = h^2\sum_{i=1}^{k}\left[f_i(u_i - \bar{u})^2\right] \qquad .....(5)$$

Dividing both sides of (5) by N, we get

$$\frac{\sum_{i=1}^{k}\left[f_i(x_i - \bar{x})^2\right]}{N} = \frac{h^2}{N}\sum_{i=1}^{k}\left[f_i(u_i - \bar{u})^2\right]$$

i.e. $\qquad \sigma_x^2 = h^2\sigma_u^2 \qquad .....(6)$

where $\sigma_x{}^2$ is the variance of the original data and $\sigma_u{}^2$ is the variance of the coded data or coded variance. $\sigma_u{}^2$ can be calculated by using the formula which involves the mean, namely,

$$\sigma_u^2 = \frac{1}{N}\sum_{i=1}^{k}\left[f_i(u_i - \bar{u})^2\right], \qquad N = \sum_{i=1}^{k} f_i \qquad .....(7)$$

or by using the formula which does not involve the mean, namely,

$$\sigma_u^2 = \frac{\sum_{i=1}^{k}\left[f_i u_i^2\right] - \frac{\left(\sum_{i=1}^{k}[f_i u_i]\right)^2}{N}}{N}, \qquad N = \sum_{i=1}^{k} f_i \qquad .....(8)$$

Example 17.14 We refer to the Example 17.12 again and find the variance and standard deviation using the coded variance.

Solution : Here h = 5 and let a = 48.

Yield per Hectare (in quintal)	Number of fields f_i	Class marks x_i	$u_i = \frac{x_i - 48}{5}$	$f_i u_i$	$u_i{}^2$	$f_i u_i{}^2$
31 – 35	2	33	–3	–6	9	18
36 – 40	3	38	–2	–6	4	12

41 – 45	8	43	– 1	– 8	1	8
46 – 50	12	48	0	0	0	0
51 – 55	16	53	+1	16	1	16
56 – 60	5	58	+2	10	4	20
61 – 65	2	63	+3	6	9	18
66 – 70	2	68	+4	8	16	32
Total	50			20		124

Thus

$$\sigma_u^2 = \frac{\sum_{i=1}^{k} f_i u_i^2 - \frac{\left(\sum_{i=1}^{k} f_i u_i\right)^2}{N}}{N}$$

$$= \frac{124 - \frac{(20)^2}{50}}{50} = \frac{124 - 8}{50} \quad \text{or} \quad \sigma_u^2 = \frac{58}{25}$$

Variance of the original data will be

$$\sigma_x^2 = h^2 \sigma_u^2 = 25 \times \frac{58}{25} = 58$$

and

$$\sigma_x = +\sqrt{58}$$

$$= 7.61 \text{ (approx)}$$

We, of course, get the same variance, and hence, standard deviation as before.

Example 17.15 Find the standard deviation for the following distribution giving wages of 230 persons.

Wages (in Rs)	No. of persons	Wages (in Rs)	No. of persons
70 – 80	12	110 – 120	50
80 – 90	18	120 – 130	45
90 – 100	35	130 – 140	20
100 – 110	42	140 – 150	8

Solution :

Wages (in Rs.)	No. of persons f_i	class mark x_i	$u_i = \frac{x_i - 105}{10}$	u_i^2	$f_i u_i$	$f_i u_i^2$
70 – 80	12	75	–3	9	–36	108
80 – 90	18	85	–2	4	–36	72
90 – 100	35	95	–1	1	–35	35
100 – 110	42	105	0	0	0	0
110 – 120	50	115	+1	1	50	50
120 – 130	45	125	+2	4	90	180
130 – 140	20	135	+3	9	60	180
140 – 150	8	145	+4	16	32	128
Total	230				125	753

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

$$= 100\left[\frac{753}{230} - \left(\frac{125}{230}\right)^2\right] = 100\,(3.27 - 0.29) = 298$$

$\therefore$ $\sigma = +\sqrt{298} = 17.3$ (approx)

CHECK YOUR PROGRESS 17.6

1. The data written below gives the daily earnings of 400 workers of a flour mill.

Weekly earning (in Rs.)	No. of Workers
80 – 100	16
100 – 120	20
120 – 140	25
140 – 160	40
160 – 180	80
180 – 200	65
200 – 220	60
220 – 240	35
240 – 260	30
260 – 280	20
280 – 300	9

Calculate the variance and standard deviation using step deviation method.

2. The data on ages of teachers working in a school of a city are given below:

Age (in years)	20 – 25	25 – 30	30 – 35	35 – 40
No. of teachers	25	110	75	120
Age (in years)	40 – 45	45 – 50	50 – 55	55 – 60
No. of teachers	100	90	50	30

Calculate the variance and standard deviation using step deviation method.

3. Calculate the variance and standard deviation using step deviation method of the following data :

Age (in years)	25 – 30	30 – 35	35 – 40
No. of persons	70	51	47
Age (in years)	40 – 50	45 – 50	50 – 55
No. of persons	31	29	22

17.11 PROPERTIES OF VARIANCE AND STANDARD DEVIA TION

Property I : The variance is independent of change of origin.

To verify this property let us consider the example given below.

Example : 17.16 The marks of 10 students in a particular examination are as follows:

10 12 15 12 16 20 13 17 15 10

Later, it was decided that 5 bonus marks will be awarded to each student. Compare the variance and standard deviation in the two cases.

Solution : ***Case –I***

x_i	f_i	$f_i x_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f_i(x_i - \bar{x})^2$
10	2	20	– 4	16	32
12	2	24	– 2	4	8
13	1	13	– 1	1	1
15	2	30	1	1	2
16	1	16	2	4	4
17	1	17	3	9	9
20	1	20	6	36	36
	10	140			92

Here $$\bar{x} = \frac{140}{10} = 14$$

Variance $= \dfrac{\sum\left[f_i(x_i - \overline{x})^2\right]}{10}$

$= \dfrac{92}{10} = 9.2$

Standard deviation $= +\sqrt{9.2} = 3.03$

Case – II (By adding 5 marks to each x_i)

x_i	f_i	$f_i x_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$f_i(x_i - \overline{x})^2$
15	2	30	–4	16	32
17	2	34	–2	4	8
18	1	18	–1	1	1
20	2	40	1	1	2
21	1	21	2	4	4
22	1	22	3	9	9
25	1	25	6	36	36
	10	190			92

$$\overline{x} = \frac{190}{10} = 19$$

$\therefore$ Variance $= \dfrac{92}{10} = 9.2$

Standard deviation $= +\sqrt{9.2} = 3.03$

Thus, we see that there is no change in variance and standard deviation of the given data if the origin is changed i.e., if a constant is added to each observation.

Property II : The variance is not independent of the change of scale.

Example 17.17 In the above example, if each observation is multiplied by 2, then discuss the change in variance and standard deviation.

Solution : In case-I of the above example , we have variance = 9.2, standard deviation = 3.03.

Now, let us calculate the variance and the Standard deviation when each observation is multiplied by 2.

x_i	f_i	$f_i x_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$f_i(x_i - \overline{x})^2$
20	2	40	–8	64	128
24	2	48	–4	16	32
26	1	26	–2	4	4
30	2	60	2	4	8
32	1	32	4	16	16
34	1	34	6	36	36
40	1	40	12	144	144
	10	280			368

$$\bar{x} = \frac{280}{10} = 28, \text{ Variance } = \frac{368}{10} = 36.8$$

$$\text{Standard deviation } = +\sqrt{36.8} = 6.06$$

Here we observe that, the variance is four times the original one and consequently the standard deviation is doubled.

In a similar way we can verify that if each observation is divided by a constant then the variance of the new observations gets divided by the square of the same constant and consequently the standard deviation of the new observations gets divided by the same constant.

Property III : Prove that the standard deviation is the least possible root mean square deviation.

Proof : Let $\bar{x} - a = d$

By definition, we have

$$s^2 = \frac{1}{N}\sum\left[f_i (x_i - a)^2\right] = \frac{1}{N}\sum\left[f_i (x_i - \bar{x} + \bar{x} - a)^2\right]$$

$$= \frac{1}{N}\sum f_i\left[(x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - a) + (\bar{x} - a)^2\right]$$

$$= \frac{1}{N}\sum f_i (x_i - \bar{x})^2 + \frac{2}{N}(\bar{x} - a)\sum f_i (x_i - \bar{x}) + \frac{(\bar{x} - a)^2}{N}\sum f_i$$

$$= \sigma^2 + 0 + d^2$$

$\therefore$ The algebraic sum of deviations from the mean is zero

or $$s^2 = \sigma^2 + d^2$$

Clearly s^2 will be least when d = 0 i.e., when $a = \bar{x}$.

Hence the root mean square deviation is the least when deviations are measured from the mean i.e., the standard deviation is the least possible root mean square deviation.

Property IV : The standard deviations of two sets containing n_1, and n_2 numbers are σ_1 and σ_2 respectively being measured from their respective means m_1 and m_2. If the two sets are grouped together as one of $(n_1 + n_2)$ numbers, then the standard deviation σ of this set, measured from its mean m is given by

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{n_1 + n_2} + \frac{n_1 n_2}{(n_1 + n_2)^2}(m_1 - m_2)^2$$

Example 17.18 The means of two samples of sizes 50 and 100 respectively are 54.1 and 50.3; the standard deviations are 8 and 7. Find the standard deviation of the sample of size 150 by combining the two samples.

Solution : Here we have

$$n_1 = 50, n_2 = 100, m_1 = 54.1, m_2 = 50.3$$

$$\sigma_1 = 8 \text{ and } \sigma_2 = 7$$

$$\sigma^2 = \frac{n_1\sigma_1^2 + n_2\sigma_2^2}{(n_1 + n_2)} + \frac{n_1 n_2}{(n_1 + n_2)^2}(m_1 - m_2)^2$$

$$= \frac{(50 \times 64) + (100 \times 49)}{150} + \frac{50 \times 100}{(150)^2}(54.1 - 50.3)^2$$

$$= \frac{3200 + 4900}{150} + \frac{2}{9}(3.8)^2 = 57.21$$

$$\therefore \quad \sigma = 7.56 \text{ (approx)}$$

Example 17.19 Find the mean deviation (M.D) from the mean and the standard deviation (S.D) of the A.P.

$$a, a + d, a + 2d, \ldots\ldots, a + 2n.d$$

and prove that the latter is greater than the former.

Solution : The number of items in the A.P. is (2n + 1)

$$\therefore \quad \bar{x} = a + nd$$

Mean deviation about the mean

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

$$= \frac{1}{(2n+1)}.2[nd + (n-1)d + \ldots\ldots + d]$$

$$= \frac{2}{(2n+1)}[1 + 2 + \ldots.. + (n-1) + n]d$$

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

Now

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

$$= \frac{2d^2}{(2n+1)}\left[n^2 + (n-1)^2 + \ldots. + 2^2 + 1^2\right]$$

$$= \frac{2d^2}{(2n+1)} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{n(n+1)d^2}{3}$$

$$\therefore \qquad \sigma = d.\sqrt{\left(\frac{n(n+1)}{3}\right)} \qquad .....(2)$$

We have further, (2) > (1)

if $$d\sqrt{\left(\frac{n(n+1)}{3}\right)} > \frac{n(n+1)}{(2n+1)}d$$

or if $$(2n+1)^2 > 3n(n+1)$$

or if $n^2 + n + 1 > 0$, which is true for $n > 0$

Hence the result.

Example 17.20 Show that for any discrete distribution the standard deviation is not less than the mean deviation from the mean.

Solution : We are required to show that

$$\text{S.D.} \geq \text{M.D. from mean}$$

or $$(\text{S. D})^2 \geq (\text{M.D. from mean})^2$$

i.e. $$\frac{1}{N}\sum\left[f_i(x_i - \bar{x})^2\right] \geq \left[\frac{1}{N}\sum\left[f_i\,|(x_i - \bar{x})|\right]\right]^2$$

or $$\frac{1}{N}\sum\left[f_i d_i^2\right] \geq \left[\frac{1}{N}\sum\left[f_i\,|d_i|\right]\right]^2, \text{ where } d_i = x_i - \bar{x}$$

or $$N\sum\left(f_i d_i^2\right) \geq \left[\sum\{f_i\,|d_i|\}\right]^2$$

or $$(f_1 + f_2 + ...)\left(f_1 d_1^2 + f_2 d_2^2 + ...\right) \geq [f_1|d_1| + f_2|d_2| + ...]^2$$

or $$f_1 f_2\left(d_1^2 + d_2^2\right) + ... \geq 2 f_1 f_2\, d_1 d_2 + ...$$

or $$f_1 f_2 (d_1 - d_2)^2 + ... \geq 0$$

which is true being the sum of perfect squares.

17.12 ANALYSIS OF FREQUENCY DISTRIBUTIONS WITH EQUAL MEANS

The variability of two series with same mean can be compared when the measures of variation are absolute and are free of units. For this, coefficient of variation (C.V.) is obtained which is defined as

$$\text{C.V.} = \frac{\sigma}{\overline{x}} \times 100, \overline{x} \neq 0$$

where σ and $\overline{x}$ are standard deviation and mean of the data. The coefficients of variation are compared to compare the variability of two series. The series with greater C.V. is said to be more variable than the other. The series having less C.V. is said to be more consistent than the other.

For series with same means, we can have

$$\text{C.V. (1st distribution)} = \frac{\sigma_1}{\overline{x}} \times 100 \qquad \text{...(1)}$$

$$\text{C.V. (2nd distribution)} = \frac{\sigma_2}{\overline{x}} \times 100 \qquad \text{...(2)}$$

where σ_1, σ_2 are standard deviation of the Ist and 2nd distribution respectively, $\overline{x}$ is the equal mean of the distributions.

From (1) and (2), we can conclude that two C.V.'s can be compared on the basis of the values of σ_1 and σ_2 only.

Example 17.21 The standard deviation of two distributions are 21 and 14 and their equal mean is 35. Which of the distributions is more variable?

Solution : Let

σ_1 = Standard dev. of 1st series = 21

σ_2 = Standard dev. of 2nd series = 14

$\overline{x}$ = 35

$$\text{C.V. (Series I)} = \frac{\sigma_1}{\overline{x}} \times 100 = \frac{21}{35} \times 100 = 60$$

$$\text{C.V. (Series II)} = \frac{\sigma_1}{\overline{x}} \times 100 = \frac{14}{35} \times 100 = 40$$

C.V. of series I > C.V. of series II

$\Rightarrow$ Series with S.D = 21 is more variable.

Example 17.22 Monthly wages paid to workers in two factories A and B and other data are given below :

	Factory A	**Factory B**
Mean of monthly wages	` 15550	` 15550
Variance of the distribution of wages	100	121

Which factory A or B shows greater variablility in individual wages?

Solution : Given

$$\sigma_A = \sqrt{\text{variance}} = \sqrt{100} = 10$$

$$\sigma_B = \sqrt{\text{variance}} = \sqrt{121} = 11$$

$$\bar{x} = \text{` } 15550$$

Now,

$$\text{C.V. (A)} = \frac{\sigma_A}{\bar{x}} \times 100 = \frac{10}{15550} \times 100$$

$$= 0.064$$

$$\text{C.V.(B)} = \frac{\sigma_B}{\bar{x}} \times 100 = \frac{11}{15550} \times 100 = 0.07$$

Clearly C.V. (B) > C.V.(A)

∴ Factory B has greater variability in the individual wages.

Example 17.23 Which of the following series X or Y is more consistent?

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

Solution : From the given data we have following table

X	Y	$D_i = X - \bar{X}$	D_i^2	$d_i = Y - \bar{Y}$	d_i^2
58	101	7	49	–4	16
52	104	1	1	–1	1
50	103	–1	1	–2	4
51	104	0	0	–1	1
49	107	–2	4	2	4
35	106	–16	256	1	1
54	105	3	9	0	0
52	105	1	1	0	0
53	107	2	4	2	4
56	108	5	25	3	9
$\Sigma X = 510$	$\Sigma Y = 1050$		$\Sigma D_i^2 = 350$		$\Sigma d_i^2 = 40$

Now,
$$\bar{X} = \frac{\Sigma X_i}{10} = \frac{510}{10} = 51$$

$$\bar{Y} = \frac{\Sigma Y_i}{10} = \frac{1050}{10} = 105$$

$$\sigma_x = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}} = \sqrt{\frac{\Sigma D_i^2}{N}} = \sqrt{\frac{350}{10}}$$

$$= 5.9$$

$$\sigma_y = \sqrt{\frac{\Sigma (Y - \bar{Y})^2}{N}} = \sqrt{\frac{\Sigma d_i^2}{N}} = \sqrt{\frac{40}{10}} = 2$$

Now,
$$\text{C.V.(X)} = \frac{\sigma_X}{X} \times 100 = \frac{5.9}{51} \times 100 = 11.5$$

$$\text{C.V.(Y)} = \frac{\sigma_Y}{Y} \times 100 = \frac{2}{105} \times 100 = 1.9$$

Clearly C.V.(Y) < C.V.(X) $\therefore$ Series Y is more consistent.

CHECK YOUR PROGRESS 17.7

1. From the data given below which section is more variable?

Marks	0-10	10-20	20-30	30-40	40-50
Section A	9	10	40	33	8
Section B	8	15	43	25	9

2. Which of the factory give better consistent wages to workers?

Wages (in `) per day	100-150	150-200	200-250	250-300	300-350
Factory A	35	45	50	42	28
Factory B	16	50	55	13	46

3. Two schools show following results of board examination in a year

	School A	School B
Average Marks Obtained	250	225
No. of Students Appeared	62	62
Variance of distribution of marks	2.25	2.56

Which school has greater variability in individual marks?

LET US SUM UP

- Range : The difference between the largest and the smallest value of the given data.

- Mean deviation from mean $= \dfrac{\sum_{i=1}^{n} (f_i |x_i - \bar{x}|)}{N}$

where $N = \sum_{i=1}^{n} f_i$, $\quad \bar{x} = \dfrac{1}{N}\sum_{i=1}^{n} (f_i x_i)$

- Mean deviation from median $= \dfrac{\sum_{i=1}^{n} fi|xi - m|}{N}$

Where $N = \dfrac{\sum_{i=1}^{n} fi}{N}$,

$$M = l + \frac{\frac{N}{2} - C}{f} \times i$$

- Variance $(\sigma^2) = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$ [for raw data]

- Standard derivation $(\sigma) = +\sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$

- Variance for grouped data

$\sigma_g^2 = \dfrac{\sum_{i=1}^{k} \left[f_i (x_i - \bar{x})^2\right]}{N}$, x_i is the mid value of the class.

Also, $\sigma_x{}^2 = h^2 \sigma_u{}^2$ and $\sigma_u{}^2 = \dfrac{1}{N}\sum_{i=1}^{k} \left[f_i \left(u_i - \bar{u}^2\right)\right]$

$$N = \sum_{i=1}^{k} f_i$$

or $$\sigma_u^2 = \frac{\sum_{i=1}^{k}\left(f_i u_i^2\right) - \frac{\left[\sum_{i=1}^{k}\left(f_i u_i^2\right)\right]^2}{N}}{N} \quad \text{where } N = \sum_{i=1}^{k} f_i$$

- Standard deviation for grouped data $\sigma_g = +\sqrt{\sigma_g^2}$
- If two frequency distributions have same mean, then the distribution with greater Coefficient of variation (C.V) is said to be more variable than the other.

SUPPORTIVE WEB SITES

http:// en.wikipedia.org/wiki/Statistical_**dispersion**

simon.cs.vt.edu/SoSci/converted/**Dispersion**_I/activity.html

TERMINAL EXERCISE

1. Find the mean deviation for the following data of marks obtained (out of 100) by 10 students in a test

55	45	63	76	67	84	75	48	62	65

2. The data below presents the earnings of 50 labourers of a factory

Earnings (in Rs.)	1200	1300	1400	1500	1600	1800
No. of Labourers	4	7	15	12	7	5

Calculate mean deviation.

3. The salary per day of 50 employees of a factory is given by the following data.

Salary (in Rs.)	20 – 30	30 – 40	40 – 50	50 – 60
No. of employees	4	6	8	12
Salary (in rupees)	60 – 70	70 – 80	80 – 90	90 – 100
No. of employees	7	6	4	3

Calculate mean deviation.

4. Find the batting average and mean deviation for the following data of scores of 50 innings of a cricket player:

Run Scored	0 – 20	20 – 40	40 – 60	60 – 80
No. of Innings	6	10	12	18
Run scored	80 – 100	100 – 120		
No. of innings	3	1		

5. The marks of 10 students in a test of Mathematics are given below:

6	10	12	13	15	20	24	28	30	32

Find the variance and standard deviation of the above data.

6. The following table gives the masses in grams to the nearest gram, of a sample of 10 eggs.

46	51	48	62	54	56	58	60	71	75

Calculate the standard deviation of the masses of this sample.

7. The weekly income (in rupees) of 50 workers of a factory are given below:

Income	400	425	450	500	550	600	650
No of workers	5	7	9	12	7	6	4

Find the variance and standard deviation of the above data.

8. Find the variance and standard deviation for the following data:

Class	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
Frequency	7	8	25	15	45

9. Find the standard deviation of the distribution in which the values of x are 1,2,......, N. The frequency of each being one.

10. The following values are calculated in respect of heights and weights of students :

	Weight	Height
Mean	52.5 Kg	160.5 cm
Standard Dev.	11.5	12.2

Which of the attribute weight or height show greater variation?

11. The following are the wickets taken by a bowler in 20 matches, for Player A

No. of Wickets	0	1	2	3	4
No. of Matches	2	6	7	4	1

For the bowler B, mean number of wickets taken in 20 matches is 1.6 with standard deviation 1.25. Which of the players is more consistent?

Find the median of the following distributions (12-14) :

12.

x_i	14	20	26	29	34	46
f_i	4	6	7	8	9	6

13.

Age (in years)	15-19	20-24	25-29	30-34	35-39
Number	8	7	9	11	5

14.

Height (in cm)	95-104	105-114	115-124	125-134	135-144
Number of Boys	10	8	18	8	16

Find mean deviation from median (15-18) :

15.

x_i	5	15	25	35	45	55
f_i	5	23	30	20	16	6

16.

x_i	105	107	109	111	113	115
f_i	8	6	2	2	2	6

17.

Income (per month) (` in '000)	0-5	6-10	11-15	16-20	21-25
Number of Persons	5	6	12	14	26

18.

Age (in years)	0-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40
No. of Persons	5	6	12	14	26	32	16	29

ANSWERS

CHECK YOUR PROGRESS 17.1

1. 15 2. 22 3. 9.4 4. 15.44

5. 13.7 6. 136 7. 5.01 8. 14.4

CHECK YOUR PROGRESS 17.2

1. 16 2. 15 3. 15.35 marks 4. 28 years

CHECK YOUR PROGRESS 17.3

1. 1.85 2. 2.36 3. 3..73 4. 0.977

CHECK YOUR PROGRESS 17.4

1. Variance = 311, Standard deviation = 17.63
2. Variance = 72.9, Standard deviation = 8.5
3. Variance = 42.6, Standard deviation = 6.53
4. Standard deviation = 4
5. Variance = 13.14, Standard deviation = 3.62
6. Standard deviation = 17.6

CHECK YOUR PROGRESS 17.5

1. Variance = 734.96, Standard deviation = 27.1
2. Variance = 12.16, Standard deviation = 3.49
3. Variance = 5489 , Standard deviation = 74.09

CHECK YOUR PROGRESS 17.6

1. Variance = 2194, Standard deviation = 46.84
2. Variance = 86.5 , Standard deviation = 9.3
3. Variance = 67.08 , Standard deviation = 8.19

CHECK YOUR PROGRESS 17.7

1. Section A
2. Factory A
3. School B

TERMINAL EXERCISE

1. 9.4   2. 124.48   3. 15.44   4. 52, 19.8

5. Variance = 72.29, Standard Deviation = 8.5   6. 8.8

7. Variance = 5581.25,Standard Deviation = 74.7

8. Variance = 840, Standard Deviation = 28.9

9. Standard deviation $= \sqrt{\dfrac{N^2 - 1}{12}}$

10. Weight   11. Player B   12. 29   13. 27.27

14. 121.16   15. 10.3   16. 3.38   17. 5.2

18. 0.62