MEASURES OF DISPERSION

4.1 Introduction :

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.

4.2 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 variance and standard deviation for raw and grouped data;
- Illustrate the properties of variance and standard deviation.

4.3 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 (40 maximum marks). 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.1)

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

Section A

Section B

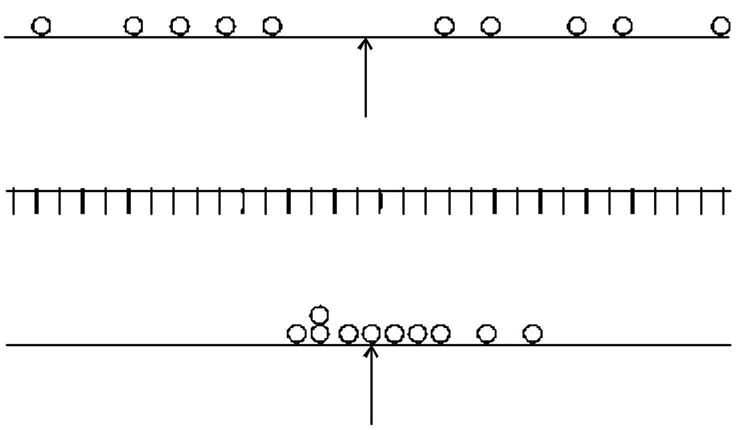

Fig. 1

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) Variance

(d) Standard deviation

(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.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 ($x_i - \bar{x}$)
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 ($x_i - \bar{x}$)
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.**

4.4 MEAN DEVIATION FROM MEAN OF RAW AND GROUPED DATA

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

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

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

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

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

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.

Example1 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

Mean deviation from mean $= \dfrac{\sum[f_i \lvert x_i - \bar{x}\rvert]}{21}$

$$= \frac{69.7}{21} = 3319$$

Example 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

Mean deviation from Mean $= \dfrac{\sum[f_i \lvert x_i - \bar{x}\rvert]}{N}$

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

4.5 VARIANCE AND STANDARD DEVIATION OF RAW DATA

If there are n observations, $x_1, x_2, \dots, x_n$, then

$$\text{Variance}(\sigma^2) = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$$

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

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

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

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, $x_i - \bar{x}$.

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

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

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 3 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 $\sigma^2 = \frac{\sum_{i=1} (x_i - \bar{x})^2}{n} = \frac{10875.50}{6}$

$= 1812.58$ (approx.)

and $\sigma = 42.57$ (approx.)

<u>Example 3</u> 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 $\bar{x} = \frac{\sum x_i}{10} = \frac{190}{10} = 19$

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 $\sigma^2 = \frac{768}{10} = 76.8$,

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

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

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

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

- Variance for grouped data

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

- Standard deviation for grouped data $\sigma_g = +\sqrt{\sigma_g^2}$

Worked example

Find the variance of 6, 7, 10, 11, 11, 13, 16, 18, 25.

Firstly we find the mean, $\bar{x} = \frac{\sum x}{n} = \frac{117}{9} = 13$.

Method 1:

$$\sigma^2 = \frac{\sum(x-\bar{x})^2}{n}$$

It is helpful to show the calculation in a table:

x	6	7	10	11	11	13	16	18	25	Total
$x-\bar{x}$	−7	−6	−3	−2	−2	0	3	5	12	
$(x-\bar{x})^2$	49	36	9	4	4	0	9	25	144	280

$$\begin{aligned}\sum(x-\bar{x})^2 &= 280\\ \sigma^2 &= \frac{\sum(x-\bar{x})^2}{n}\\ &= \frac{280}{9}\\ &= 31.11 \quad \text{(2dp)}\end{aligned}$$

Method 2:

$$\sigma^2 = \frac{\sum x^2}{n} - \bar{x}^2$$

x	6	7	10	11	11	13	16	18	25	Total
x^2	36	49	100	121	121	169	256	324	625	1801

$$\begin{aligned}\sigma^2 &= \frac{\sum x^2}{n} - \bar{x}^2\\ &= \frac{1801}{9} - 13^2\\ &= 200.11 - 169\\ &= 31.11 \quad \text{(2dp)}\end{aligned}$$

Example 1

The five students in a class take a test. Their scores in points are as follows:

80 76 100 83 100

Let's look at three possibilities for measuring the spread or variation of a data set.

Note: All three of these measures are nonnegative in value.

1) Range

Range = Max - Min

= highest value - lowest value in the data set

Range = Max - Min = 100 -76 = 24 points

2) Variance (VAR).
3) Standard Deviation (SD).

Data (x)	**Step 2** Deviations: $(x-\mu)$ values	**Step 3** Squared Deviations: $(x-\mu)^2$ values
80	-7.8	60.84
76	-11.8	139.24
100	12.2	148.84
83	-4.8	23.04
100	12.2	148.84
Step 1: $\mu = 87.8$ points	See Note 2 below.	Sum = 520.8 **Do Steps 4, 5.**

Step 4:

$$\text{VAR, or } \sigma^2 = \text{the average of the squared deviations}$$
$$= \frac{520.8}{5}$$
$$= 104.16 \text{ square points}$$

One reason why we often prefer the SD over the VAR is that units

like "square points" are not natural to us.

Step 5:

$$\text{SD, or } \sigma = \sqrt{\text{VAR}}$$
$$= \sqrt{104.16}$$
$$\approx 10.2 \text{ points}$$

Observe that the SD shares the same units as the original data values.

Exercises

Find the variance and standard deviation of the following correct to 2 decimal places:

1. a) 10, 16, 12, 15, 9, 16, 10, 17, 12, 15
b) 74, 72, 83, 96, 64, 79, 88, 69
c) £326, £438, £375, £366, £419, £424

Answers

1. a) 7.76, 2.79 b) 97.36, 9.87 c) .£ 1531.22, £39.13

4.5 Coefficient of Variation (CV).

The Coefficient of Variation (CV) for a data set defined as the ratio of the standard deviation to the mean:

$$The\ Coefficient\ of\ Variation\ (CV)\ =\ \frac{\pm}{«}\ \ for\ \ population\ or\ \ \frac{s}{\overline{x}}\ \ for\ sample$$

EXAMPLE: Several measurements of the diameter of a ball bearing made with one micrometer had a mean of 2.49 mm and a standard deviation of 0.012 mm, and several measurements of the unstretched length of a spring made with another micrometer had a mean of 0.75 in. with a standard deviation of 0.002 in. Which of the two micrometers is relatively more precise?.

Solution: Calculating the two coefficients of variation, we get.

$$CV_1\ X\frac{0.012}{2.49}\ \lfloor 100 \rceil\ 0.48\%\ \ and\ CV_2\ X\frac{0.002}{0.75}\ \lfloor 100 \rceil\ 0.27\%$$

Thus, the measurements of the length of the spring are relatively less variable, which means that the second micrometer is more precise.

Example:

Is the timing of large earthquakes more regular on the Japan or Cascadia subduction zone?

Japan	Casacadia
Mean interval time = 100 years	Mean interval time = 400 years
Standard deviation = 20 years	Standard deviation = 40 years
Coefficient of variation = 0.2	Coefficient of variation = 0.1

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.

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

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.

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 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.