PERIOD–II

TOPIC 5

Standard Deviation

5.1. MEANING

Suppose there are two classes of students with average percentage of marks 56% each, it would not be justified to take for guaranteed that the students of both classes are approximately of same level of intelligence. It may be just possible that the students of first class are getting marks between 50% and 59% and the students of the second class getting between 8% and 90%. Thus, we see that even if the average value of two series are same, the items may be of quite different in nature, in two series. The measures of central tendency are inadequate to give the complete idea about the items in the series, we would be very much helped, if we are also given some idea about the scatter of items about the average value. The measure of scatteredness of items about some average is also called measure of dispersion. We may define measure of dispersion for a statistical data as the measurement of the spread of the items about an average value.

5.2. METHODS OF MEASURING DISPERSION

I. Range
II. Quartile Deviation (Q.D.)
III. Mean Deviation (M.D.)
IV. Standard Deviation (S.D.)
V. Lorenz Curve.

I. RANGE

DEFINITION

The **range** of a statistical data is defined as the difference between the largest and the smallest values of the variable.

$\therefore$ **Range = L – S**,

where L = largest value of the variable

S = smallest value of the variable.

In case, the values of the variable are given in the form of classes, then L is taken as the upper limit of the largest value class and S as the lower limit of the smallest value class.

Example 1. *Find the range of the series :*

4, 2, 6, 8, 10.

Solution. Here L = 10, S = 2.

$\therefore$ Range = L – S = 10 – 2 = **8.**

Example 2. *Find the range of the following distribution* :

No. of students	*16–18*	*18–20*	*20–22*	*22–24*	*24–26*	*26–28*
No. of students	*0*	*4*	*6*	*8*	*2*	*2*

Solution. Here L = 28, S = 18

$\therefore$ Range = L – S = 28 – 18 = **10 years.**

It may be noted that S $\neq$ 16, though it is the lower limit of the smallest value class, but there is no item in this class and so this class is meaningless so far as the calculation of range is concerned.

Let us consider the market value of shares of companies A and B, during a particular week.

Day	***Mon-day***	***Tues-day***	***Wednes-day***	***Thurs-day***	***Friday***	***Satur-day***
M.V. of shares of company A (in L$)	12	11	10	13	16	20
M.V. of shares of company B (in L$)	60	50	55	62	70	75

From the data, we see that Range (A) = 20 – 10 = L$ 10 and Range (B) = 75 – 50 = L$ 25. From these results, one is likely to infer that there is more variability in the II series. But this is not so, because the M.V. of shares of A has increased by 100% in the week, whereas there is only 50% rise in the M.V. of shares of B, during that week. Thus, variability is more in the first series. Thus, we see that range may give misleading results if used for comparing two or more series for variability (scatteredness, dispersion). For comparison purpose, we use its corresponding relative measure, called '*coefficient of range*'. This is defined as

$$\textbf{Coeff. of Range} = \frac{\mathbf{L - S}}{\mathbf{L + S}}.$$

Now Coeff. of Range for A = $\dfrac{\mathbf{20 - 10}}{\mathbf{20 + 10}} = \dfrac{\mathbf{10}}{\mathbf{30}}$ = **0.3333.**

Coeff. of Range for B = $\dfrac{\mathbf{75 - 50}}{\mathbf{75 + 50}} = \dfrac{\mathbf{25}}{\mathbf{125}}$ = **0.2000.**

∴ Coeff. of Range (A) > Coeff. of Range (B)

∴ Variability is more in the M.V. of shares of company A.

II. QUADRTILE DEVIATION (Q.D.)

5.3. INADEQUACY OF RANGE

Consider the series

I : 4, 4, 4, 5, 5, 6, 4, 5, 5, 1000.

II : 4, 4, 4, 5, 5, 6, 4, 5, 5.

For series I, Coeff. of Range = $\dfrac{1000 - 4}{1000 + 4} = \dfrac{996}{1004}$ = 0.992

For series II, Coeff. of Range = $\dfrac{6 - 4}{6 + 4} = \dfrac{2}{10}$ = 0.200.

On comparing the values of coeff. of range for these series, one is likely to conclude that these is marked difference in variability in the series. In fact, the series II is obtained from the series I, just by ignoring the extreme item 1000. Thus, we see that extreme items can distort the value of range and even the coefficient of range. If we have a glance at the definitions of these measures, we would find that only extreme items are required in their calculation, if at all extreme items are present. Even if extreme items are present in a series, the middle 50% values of the variable would be expected to vary quite smoothly, keeping this in view, we define another measure of dispersion, called 'Quartile Deviation'.

5.4. QUARTILE DEVIATION

The **quartile deviation** of a statistical data is defined as

$$\frac{Q_3 - Q_1}{3}$$ and is denoted as Q.D.

This is also called *semi-inter quartile* range. We have already studied the method of calculating quartiles. The value of Q.D. is obtained by subtracting Q_1 from Q_3 and then dividing it by 2.

For comparing two or more series for variability, the absolute measure Q.D. would not work. For this purpose, the corresponding relative measure, called coeff. of Q.D. is calculated. This is defined as :

$$\textbf{Coeff. of Q.D.} = \frac{\mathbf{Q_3 - Q_1}}{\mathbf{Q_3 + Q_1}}.$$

Example 3. *Find Q.D. and its coefficient for the following series:*

x (in L$) : 4, 7, 6, 5, 9, 12, 19.

Sol. The values of the variable arranged in ascending order are

x (in L$) : 4, 5, 6, 7, 9, 12, 19.

Here $n = 7$.

$$\mathbf{Q_1} : \frac{n+1}{4} = \frac{7+1}{4} = 2 \qquad \therefore \quad Q_1 = \text{size of 2nd item} = \text{in L\$ } 5$$

$$\mathbf{Q_3} : 3\left(\frac{n+1}{4}\right) = 3\left(\frac{7+1}{4}\right) 6 \qquad \therefore \quad Q_3 = \text{size of 6th item} = \text{in L\$ } 12$$

$$\therefore \qquad \text{Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{12 - 5}{2} = \textbf{in L\$ 3.5.}$$

$$\text{Coeff. of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{12 - 5}{12 + 5} = \frac{7}{17} \ \textbf{0.4118.}$$

Example 4. *Find the quartile deviation for the following distribution:*

Marks	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
No. of students	*10*	*11*	*12*	*13*	*5*	*12*	*7*	*5*

Sol. **Calculation of Quartiles**

Marks	*No. of students* f	*c.f.*
2	10	10
3	11	21
4	12	33
5	13	46
6	5	51
7	12	63
8	7	70
9	5	75 = N
	N = 75	

$$Q_1 : \frac{N+1}{4} = \frac{75+1}{4} = 19$$

$\therefore$ Q_1 = size of 19th item = 3 marks

$$Q_3 : 3\left(\frac{N+1}{4}\right) = 3\left(\frac{75+1}{4}\right) = 57$$

$\therefore$ Q_3 = size of 57th item = 7 marks

$$\therefore \quad \text{Q.D.} = \frac{Q_3 - Q_2}{2} = \frac{7-3}{2} = \mathbf{2\ marks.}$$

III. MEAN DEVIATION (M.D.)

5.5. MEAN DEVIATION (M.D.)

Mean deviation is also called **average deviation**. The **mean deviation** of a statistical data is defined as the arithmetic mean of the numerical values of the deviations of items from some average. Generally, A.M. and median are used in calculating mean deviation. Let 'a' stand for the average used for calculating M.D.

For an **individual series**, the M.D. is given by

$$\textbf{M.D.} = \frac{\sum_{i=1}^{n} |\mathbf{x_i - a}|}{\mathbf{n}} = \frac{\Sigma|\mathbf{x - a}|}{\mathbf{n}}$$

where x_1, x_2, ..., x_n are the values of the variable, under consideration.

For a **frequency distribution,**

$$\textbf{M.D.} = \frac{\sum_{i=1}^{n} f_i \, | x_i - a |}{N} = \frac{\Sigma f \, | x - a |}{N}$$

where f_i is the frequency of x_i $(1 \le i \le n)$.

When the values of the variable are given in the form of classes, then their respective mid-points are taken as the values of the variable.

Median is used in calculating M.D., because of its property that the sum of numerical values of deviations of items from median is always least. So, if median is used in the calculation of M.D., its value would come out to be least. M.D. is also calculated by using A.M. because of its simplicity and popularity. In problems, it is generally given as to which average is to be used in the calculation of M.D. If it is not given, then either of the two can be made use of.

5.6. COEFFICIENT OF M.D.

For comparing two or more series for variability, the corresponding relative measure, 'Coefficient of M.D.', is used. This is defined as :

$$\textbf{Coeff. of M.D.} = \frac{\textbf{M.D.}}{\textbf{Average}}.$$

If M.D. is calculated about A.M., then M.D. is written as M.D.. Similarly, M.D.(Median) would mean that median has been used in calculating M.D.

$\therefore$ We can write

$$\textbf{Coeff. of M.D. } (\bar{x}) = \frac{\textbf{M.D.}(\bar{x})}{\bar{x}}$$

$$\textbf{Coeff. of M.D.(Median)} = \frac{\textbf{M.D.(Median)}}{\textbf{Median}}$$

WORKING RULES TO FIND M.D. ($\bar{x}$)

Rule I. *In case of an individual series, first find* $\bar{x}$ *by using the formula* $\bar{x} = \frac{\Sigma x}{n}$. *In the second step, find the values of* $x - \bar{x}$. *In the* *next step, find the numerical values* $|x-\bar{x}|$ *of* $x-\bar{x}$. *Find the sum* $\Sigma|x-\bar{x}|$ *of these numerical values* $|x-\bar{x}|$. *Divide this sum by n to get the value of M.D.*$(\bar{x})$.

Rule II. *In case of a frequency distribution, first find* $\bar{x}$ *by using the formula* $\bar{x}=\frac{\Sigma fx}{N}$. *In the second step, find the values of* $x-\bar{x}$. *In the next step, find the numerical values* $|x-\bar{x}|$ *of* $x-\bar{x}$. *Find the products of the values of* $|x-\bar{x}|$ *and their corresponding frequencies. Find the sum* $\Sigma f|x-\bar{x}|$ *of these products. Divide this sum by N to get the value of M.D.*$(\bar{x})$.

Rule III. *If the values of the variable are given in the form of classes, then their respective mid-points are taken as the values of the variable.*

Rule IV. *To find the coefficient of M.D.*$(\bar{x})$, *divide M.D.*$(\bar{x})$ *by* $(\bar{x})$.

Example 5. *Find M.D. and M.D.(median) for the following statistical series :*

7, 10, 12, 13, 15, 20, 21, 27, 30, 35.

Solution. **Calculation of M.D.**

S. No.	x	$x-\bar{x}$ $\bar{x}=19$	$\lvert x-\bar{x}\rvert$
1	7	– 12	12
2	10	– 9	9
3	12	– 7	7
4	13	– 6	6
5	15	– 4	4
6	20	1	1
7	21	2	2
8	27	8	8
9	30	11	11
10	35	16	16
n = 10	Σx = 190		$\Sigma\lvert x-\bar{x}\rvert$ = 76

$$\bar{x} = \frac{\Sigma x}{n} = \frac{190}{10} = 19$$

$$\therefore \quad \text{M.D.}(\bar{x}) = \frac{\Sigma|x - \bar{x}|}{n} = \frac{76}{10} = 7.6.$$

Calculation of M.D.(median)

S. No.	***x***	***x – median*** ***median = 17.5***	**| *x – median* |**
1	7	– 10.5	10.5
2	10	– 7.5	7.5
3	12	– 5.5	5.5
4	13	– 4.5	4.5
5	15	– 2.5	2.5
6	20	2.5	2.5
7	21	3.5	3.5
8	27	9.5	9.5
9	30	12.5	12.5
10	35	17.5	17.5
$n = 10$			$\Sigma|x - \text{median}| = 76$

$$\frac{n+1}{2} = \frac{10+1}{2} = \mathbf{5.5}$$

$$\therefore \quad \text{Median} = \frac{\text{Size of 5th item + size of 6th item}}{2} = \frac{15+20}{20}$$

$$= 17.5$$

$$\therefore \quad \text{M.D.(median)} = \frac{\Sigma|x - \text{median}|}{n} = \frac{76}{10} = \mathbf{7.6.}$$

Example 6. *Find the coeff. of M.D.(Median) for the following frequency distribution:*

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

Solution. **Calculation of M.D.(Median)**

Marks	*No. of students*	*c.f.*	*Mid-points of classes x*	*x-median (med. = 28)*	*| x – med. |*	*f| x – med. |*
0–10	5	5	5	– 23	23	115
10–20	8	13	15	– 13	13	104
20–30	15	28	25	– 3	3	45
30–40	16	44	35	7	7	112
40–50	6	50 = N	45	17	17	102
	N = 50					$\Sigma f \mid x - med. \mid = 478$

Median = size of 50/2th item = size of 25th item.

∴ Median class is 20–30

$$\text{Median} = L + \left(\frac{N/2 - c}{f}\right)h = 20 + \left(\frac{25-13}{15}\right)10$$

$$= 28$$

Now $$\text{M.D.(Median)} = \frac{\Sigma f \mid x - \text{median} \mid}{N} = \frac{478}{50} = 9.56 \text{ marks.}$$

$$\therefore \text{ Coeff. of M.D.(Median)} = \frac{\text{M.D.(Median)}}{\text{Median}} = \frac{9.56}{28} = \mathbf{0.3414.}$$

IV. STANDARD DEVIATION (S.D.)

5.7. STANDARD DEVIATION (S.D.)

It is the most important measure of dispersion. It finds indispensable place in advanced statistical methods. The **standard deviation** of a statistical data is defined as the positive square root of the A.M. of the squared deviations of items from the A.M. of the series under consideration. The S.D. is often denoted by the greek letter 'σ'.

For an **individual series**, the S.D. is given by

$$\textbf{S.D.} = \sqrt{\frac{\sum_{i=1}^{n}(\mathbf{x_i} - \bar{\mathbf{x}})^2}{\mathbf{n}}} = \sqrt{\frac{\Sigma(\mathbf{x} - \bar{\mathbf{x}})^2}{\mathbf{n}}}$$

where $x_1, x_2, \ldots\ldots, x_n$ are the value of the variable, under consideration.

For a **frequency distribution**,

$$\textbf{S.D.} = \sqrt{\frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N}} = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{N}}$$

where f_i is the frequency of x_i $(1 \le i \le n)$.

When the values of the variable are given in the form of classes, then their respective mid-points are taken as the values of the variable.

WORKING RULES TO FIND S.D.

Rule I. *In case of an individual series, first find* $\bar{x}$ *by using the formula* $\bar{x} = \dfrac{\Sigma x}{n}$. *In the second step, find the values of* $x - \bar{x}$. *In the next step, find the numerical values* $(x - \bar{x})^2$ *the values of* $x - \bar{x}$. *Find the sum* $\Sigma(x - \bar{x})^2$ *of these numerical values* $(x - \bar{x})^2$. *Divide this sum by n. Take the positive square root of this to get the value of S.D.*

Rule II. *In case of a frequency distribution, first find* $\bar{x}$ *by using the formula* $\bar{x} = \dfrac{\Sigma x}{n}$. *In the second step, find the values of* $x - \bar{x}$. *In the next step, find the squares* $(x - \bar{x})^2$ *of the values of* $x - \bar{x}$. *Find the products of the values of* $(x - \bar{x})^2$ *and their corresponding frequencies. Find the sum* $\Sigma f(x - \bar{x})^2$ *of these products. Divide this sum by N. Take the positive square root of this to get the value of S.D.*

Rule III. *If the values of the variable are given in the form of classes, then their respective mid-points are taken as the values of the variable.*

Rule IV. (*i*) *Coeff. of S.D.* $= \dfrac{S.D.}{A.M.}$

Example 7. *Find the S.D. for the following data:*

4, 6, 10, 12, 18.

Sol. **Calculation of S.D.**

S. No.	$\boldsymbol{x}$	$\boldsymbol{x - \bar{x}}$ $\boldsymbol{\bar{x} = 10}$	$(\boldsymbol{x - \bar{x}})^2$
1	4	– 6	36
2	6	– 4	16
3	10	0	0
4	12	2	4
5	18	8	64
$n = 5$	$\Sigma x = 50$		$\Sigma(x - \bar{x}) = 120$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{50}{5} = 10$$

Now $$\text{S.D.} = \sqrt{\frac{\Sigma\,(x - \bar{x})^2}{n}} = \sqrt{\frac{120}{5}} = \sqrt{24} = \mathbf{4.8989.}$$

Example 8. *Calculate S.D. for the following data:*

$\boldsymbol{x}$	5	15	25	35	45	55
$\boldsymbol{f}$	12	18	27	20	17	6

Solution. **Calculation of S.D.**

$\boldsymbol{x}$	$\boldsymbol{f}$	$\boldsymbol{fx}$	$\boldsymbol{x - \bar{x}}$	$(\boldsymbol{x - \bar{x}})^2$	$\boldsymbol{f(x - \bar{x})^2}$
5	12	60	– 23	529	6348
15	18	270	– 13	169	3042
25	27	675	– 3	9	243
35	20	700	7	49	980
45	17	765	17	289	4913
55	6	330	27	729	4374
	N = 100	$\Sigma\, fx = 2000$			$\Sigma\, f\,(x - \bar{x})^2 = 19900$

$$\bar{x} = \frac{\Sigma fx}{\text{N}} = \frac{2800}{100} = 28.$$

Now $$\text{S.D.} = \sqrt{\frac{\Sigma f(x - \bar{x})^2}{\text{N}}} = \sqrt{\frac{19900}{100}} = \sqrt{199} = \mathbf{14.1067.}$$

EXERCISE

1. Find the coeff. of Q.D. for the following distribution:

Marks	0–4	4–8	8–12	12–14
No. of students	10	12	18	7
Marks	14–18	18–20	20–25	25 and above
No. of students	5	8	4	6

2. Find the M.D. from A.M. for the following data:

x	3	5	7	9	11	13
f	*2*	*7*	*10*	*9*	*5*	*2*

3. Calculate S.D. for the following frequency distribution:

Class	***Frequency***	***Class***	***Frequency***
4–8	11	24–28	9
8–12	13	28–32	17
12–16	16	32–36	6
16–20	14	36–40	4
20–24	14		