5 Dispersion

The measures of central tendencies (i.e. means, median, mode) indicate the general magnitude of the data and locate only the center of a distribution of measures. They do not establish the degree of variability or the spread out or scatter of the individual items and their deviation from (or the difference with) the means.

i) According to NcisWanger, "Two distributions of statistical data may be symmetrical and have common means, medians and modes and identical frequencies in the modal class. Yet. With these points in common, they may differ widely in the scatter or in their values about the measures of central tendencies".

ii) Simpson and Kafka said "An average alone does not tell the full story. It is hardly fully representative of a mass, unless we know the manner in which the individual item scatter around it ..... a further description of a series is necessary, if we are to gauge, how representative the average is"

From this discussion, we now focus our attention on the scatter or variability which is known as **Disperson**. Let us take the following three sets.

Students	Group X	Group Y	Group Z
1	50	45	30
2	50	50	45
3	50	55	75
$\therefore$ Mean $\overline{X} \Rightarrow$	50	50	50

Thus, the tree groups have same mean i.e. 50. In fact the median of group X and group Y are also equal. Now, if one would say that the students from the three groups are of equal capabilities, it is totally a wrong conclusion then. Close examination reveal that in group 'X' students have equal marks as the mean, students from group 'Y' are very close to the mean but in the third group Z, the marks are widely scattered. It is thus clear that the measure of central tendency is alone not sufficient to describe the data.

Definition of Dispersion : The arithmetic mean of the deviations of the values of the individual item from the measure of a particular central tendency used.

Dispersion also known as scatter, spread or variation measures the extent to which the items vary from the central value. Since measure of dispersion give an average of the differences of various items from an average, they are also called averages of the second order.

In measuring dispersion, it is imperative to know the amount of variation (absolute measure) and the degree of variation (relative measure). In the former case we consider the range, mean deviation, standard deviation etc. In the later case, we consider the coefficient of range, coefficient of mean deviation, the coefficient of variation etc.

SIGNIFICANCE OF MEASURING DISPERSION

Measures of variation are needed for four basic purposes :

1. To determine the reliability of an average.
2. To serve as a basis for the control of the variability.

METHODS OF COMPUTING DISPERSION

1. Range
2. Coefficient of Range
3. Quartile Deviation
4. Coefficient of Quartile Deviation
5. Mean Deviation
6. Standard Deviation

Note that we are going to study some of these and not all.

1

X	f
10-20	4
20-30	5
30-40	6
⋮	⋮

IMP

Range

It is the simplest absolute measure of dispersion. Range in a distribution is the difference between the largest and the smallest values of the variable.

$$\text{Thus Range (R)} = L - S \begin{cases} L - \text{Largest Value of the Series} \\ S - \text{Small Value of the Series} \end{cases}$$

Example : 1

Find the range of the following items :

18, 15, 20, 17, 22, 16

Solution :

Largest Value (L) = 22

Smallest Value (S) = 15

$\therefore$ Range = L – S = 22 – 15 = 7.

Coefficient of Range : The relative measure of the range. It is used in the comparative study of the dispersion.

$$\therefore \text{ coefficient of Range} = \frac{L - S}{L + S}$$

2

Note : A measure of dispersion is the ratio of a measure of absolute to an appropriate average.

IMP

the difference between the mid-values of the highest and the lowest classes.

Example : 2

Find the range and the coefficient of the range of the following items :

110, 117, 129, 197, 190, 100, 100, 178, 255, 790

Solution :

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{790-100}{790+100} = \frac{690}{890} = 0.78$$

Example : 3

Find the range and its coefficient from the following data.

Size :	10–20	20–30	30–40	40–50	500–100
Frequency	2	3	5	4	2

Solution :

$$R = L - S = 100 - 10 = 90$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{100-10}{100+10} = \frac{90}{110} = 0.82$$

Uses of Range

1. Quality Control

The idea basically is that if the range – the difference between the largest and smallest mass produced items – increases beyond a certain point, the production machinery should be examined to find out why the items produced have not followed their usual more consistent pattern.

2. Fluctuations in the Share Prices

Range is useful in studying the variations in the prices of stocks and shares and other commodities that are sensitive to price changes from one period to another.

3. Weather Forecasts

The meteorological department does make use of the range in determining the difference between the minimum temperature and the maximum temperature.

Merits :

1. It can be easily understood.
2. It is easy to calculate and it is the simplest method of measuring dispersion.
3. It lends itself to algebraic treatments.
4. It is an absolute measure of dispersion.

5. It is affected by sampling fluctuations.

QUARTILE DEVIATION OR SEMI-INTER QUARTILE RANGE

Quartile deviation is a measure of dispersion-based on the Upper Quartile (Q_3) and Lower Quartile (Q_1) of a series. It is half of the difference between the upper quartile and the lower quartile. This difference is the range between these two quartiles and is called inter-quartile range. The half of this range is semi-quartile range. The quartile deviation is also known as semi-inter-quartile range.

$$\therefore \quad \text{Interquartile Range} = Q_3 - Q_1$$

$$\text{Quartile Deviation or Q.D} = \frac{Q_3 - Q_1}{2}$$

$$\text{Coefficient of Q.D} = \frac{(Q_3 - Q_1)/2}{(Q_3 + Q_1)/2} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Merits :

4. It is not the representative value of data

Example : 5

Find the quartiles and the interquartile range of the following numbers :

29, 19, 26, 12, 24, 21, 36, 35, 33

Solution :

Arranging the given data in ascending order of magnitude we get

12, 19, 21, 24, 26, 29, 33, 35,36

Here, $N = 9, \dfrac{N+1}{4} = \dfrac{9+1}{4}$; $\dfrac{N+1}{2} = \dfrac{9+1}{2}$ and $\dfrac{3(N+1)}{4} = \dfrac{3(9+1)}{4}$

$$\therefore \quad 1^{st} \text{ Quartile } (Q_1) = \text{Value of } \frac{9+1}{4}\text{th i.e. } 2.5^{th} \text{ term}$$

$$= \text{Value of 2nd term} + \frac{1}{2}\text{(3rd term – 2nd term)}$$

$$= 19 + \frac{1}{2}(21 - 19) = 19 + \frac{1}{2} \times 2 = 20$$

$$3^{rd} \text{ Quartile } (Q_3) = \text{Value of the } \frac{3(9+1)}{4}\text{th i.e. 7.5th term}$$

$$= \text{Value of 7th term} + \frac{1}{2}\text{(8th term – 7th term)}$$

$$= 33 + \frac{1}{2}(35 - 33) = 33 + \frac{1}{2} \times 2 = 34$$

2^{nd} Quartile (Q_2) = Value of $\frac{(9+1)}{2}$th term = 5th term = 26

$\therefore$ Interquartile Range = $Q_3 - Q_1 = 34 - 20 = 14.$

Example : 6

Calculate the quartile deviation and coefficient of Quartile deviation from the following data :

Age (in yrs)	10	15	20	25	30	35	40
No. of People	5	10	25	30	20	15	2

n = obs.
= 7 or

Solution :

We have the following Table :

Age in yrs (X)	No. of Members (f)	Cumulative Frequency (CF)
10	5	5
15	10	15
20	25	40
25	30	70
30	20	90
35	15	105
40	2	107

$\Sigma f = N$ Q_1 → 27

Q_3 → 81

Here N = 107

Q_1 = Value of $\left(\frac{N+1}{4}\right)^{th}$ term = $\frac{107+1}{4}$ = 27^{th} term = 20

Q_3 = Value of $3\left(\frac{N+1}{4}\right)^{th}$ term = $3\left(\frac{107+1}{4}\right)$ = 81^{th} term = 30

$\therefore$ Quartile Deviation = $\frac{Q_3 - Q_1}{2} = \frac{30-20}{2} = 5$

Coefficient of Q.D = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{30-20}{30+20} = \frac{10}{50} = 0.2$

Example : 7

Calculate the Quartile deviation and coefficient of quartile deviation for the following data :

Class	0–10	10–20	20–30	30–40	40–50	50–60
Frequency	3	7	15	12	8	5

Solution :

	Class	F	CF
	0–10	3	3
	10–20	7	10
Q1	20–30	15	25
	30–40	12	37
Q3	40–50	8	45
	50–60	5	50
		50	

For Q_1

$$\frac{N}{4} = \frac{50}{4} = 12.5^{th} \text{ term}$$

$\therefore Q_1$ lies in the class 20–30

$\therefore l_1 = 20,\ l_2 = 30.\ f = 15,\ CF = 10$

$$\therefore Q_1 = l_1 + \frac{\frac{N}{4} - CF}{F} \times (l_2 - l_1)$$

$$= 20 + \frac{12.5 - 10}{15} \times (30 - 20)$$

$$= 20 + \frac{2.5 \times 10}{15} = 21.67$$

$$\therefore \boxed{Q_1 = 21.67}$$

For Q_3

$$\frac{3N}{4} = \frac{3 \times 50}{4} = 37.5^{th} \text{ term}$$

$\therefore$ Q_3 lies in the class 40–50

$\therefore$ $l_1 = 40, l_2 = 50, f = 8, CF = 37$

$$\therefore Q_3 = l_1 + \frac{\frac{3N}{4} - CF}{F} \times (l_2 - l_1)$$

$$= 40 + \frac{37.5 - 37}{8} \times (50 - 40)$$

$$= 40 + \frac{0.5 \times 10}{8} = 40 + 0.625 = 40.625 \approx 40.63$$

$$\therefore \boxed{Q_3 = 40.63}$$

$$\therefore \text{Quartile Deviation (Q.D)} = \frac{Q_3 - Q_1}{2} = \frac{40.63 - 21.67}{2} = \frac{18.96}{2} = 9.48$$

$$\text{Coefficient of Q.D} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{40.63 - 21.67}{40.63 + 21.67} = \frac{18.96}{62.30} = 0.3044$$

Mean Deviation ungrouped data or raw data

While calculating the mean deviations, the algebraic sign of the deviation is always taken as **positive,** because the sum of deviations with their algebraic signs + and – from the arithmetic mean is always zero.

$$\Sigma f |x_i - med|$$

$$n$$

$$\text{Mean deviation : } MD_{\bar{X}} = \frac{\sum_{i=1}^{n} F |X_i - \bar{X}|}{n}$$, where $\bar{X}$ is the A.M. Similarly, Mean Deviation from the median M_d for ungrouped data is denoted by MD_{med}.

$$\therefore \quad MD_{med} = \frac{\Sigma |X_i - M_d|}{n}$$

xf fx

Mean Deviation from the mode Z for ungrouped data is denoted by $MD_Z = \frac{|X_i - Z|}{n}$ x_

Example : 9

Calculate mean deviation about mean from the following data :

12, 20, 39, 46, 54, 61, 78, 90

Solution :

Here $n = 8$ and therefore

$$\text{Mean} = \frac{12+20+39+46+54+61+78+90}{8} = \frac{400}{8} = 50$$

$\bar{X}$ = mean

Now

X	$d = X - M = (X - 50)$	$\lvert d \rvert$
12	– 38	38
20	– 30	30
39	– 11	11
46	– 4	4
54	4	4
61	11	11
78	28	28
90	40	40
		$\Sigma d = 166$

$$\therefore \quad MD_{mean} = \frac{\Sigma d}{n} = \frac{166}{8} = 20.75$$

$$= \frac{\sum \lvert (X_i - \bar{X}) \rvert}{n}$$

Example : 10

Calculate the mean deviation from Mean, Median and Mode for the following data :

Age	10	11	12	13	14
No. of Boys	2	4	7	4	3

max.

Solution :

X	f	fx	$\lvert X-\bar{X}\rvert$	$f\lvert X-\bar{X}\rvert$	$\lvert X-M_d\rvert$	$f\lvert X-M_d\rvert$
10	2	20	2.1	4.2	2	4
11	4	44	1.1	4.4	1	4
12	7	84	0.1	0.7	0	0
13	4	52	0.9	3.6	1	4
14	3	42	1.9	5.7	2	6
	$\Sigma f=20$	$\Sigma fx=242$		$\Sigma f\lvert X-\bar{X}\rvert=18.6$		$\Sigma f\lvert X-M_d\rvert=18$

$$\text{Mean}\,(\overline{X}) = \frac{\Sigma fx}{\Sigma f} = \frac{242}{20} = 12.1$$

Mean

Mean Deviation from Mean

$$\boxed{\text{i.e. } MD_{mean} = \frac{\Sigma f|X-\overline{X}|}{\Sigma f} = \frac{18.6}{20} = 0.93}$$

Median = 12

Med

$$\therefore \text{ Mean Deviation from } \cancel{\text{Mean}} = \frac{\Sigma f|X-M_d|}{\Sigma f} = \frac{18}{20} = 0.9$$

Since, mode for the data is also 12, the mean deviation from mode is also 0.9.

Example : 11

Since, mode for the data is also 12, the mean deviation from mode is also 6.9.

Example : 11

Find the mean deviation from the mean for the following data :

C. I	0–10	10–20	20–30	30–40	40–50	50–60	60–70
f	8	12	10	8	3	2	7

Solution :

C.I	f	x	fx	$\lvert X-\overline{X}\rvert$	$f\lvert X-\overline{X}\rvert$
0–10	8	5	40	24	192
10–20	12	15	180	14	168
20–30	10	25	250	4	40
30–40	8	35	280	6	48
40–50	3	45	135	16	48
50–60	2	55	110	26	52
60–70	7	65	455	36	252
	$\Sigma f = 50$		$\Sigma fx = 1450$		$\Sigma f\lvert X-\overline{X}\rvert = 800$

$$\text{Mean }(\overline{X}) = \frac{\Sigma fx}{\Sigma f} = \frac{1450}{50} = 29$$

$$\text{Mean Deviation from Mean} = (MD_{mean}) = \frac{\Sigma f\lvert X-\overline{X}\rvert}{\Sigma f} = \frac{800}{50} = 16$$

$$\boxed{\text{Coefficient of } MD_{mean} = \frac{MD_{mean}}{\text{Mean}}} = \frac{16}{29} = 0.55$$

STANDARD DEVIATION

Standard deviation is the most important and commonly used measure of dispersion. It measures the absolute dispersion or variability of a distribution. A small standard deviation means a high degree of uniformity in the observations as well as homogeneity of the series.

Definition : Standard deviation is the positive square root of the average of squared deviations taken from arithmetic mean. It is generally denoted by the Greek alphabet σ or by S.D standard deviation of 'n' observations is given by

Standard Deviation

For ungrouped data

$$\sigma = \sqrt{\frac{\Sigma (X - \overline{X})^2}{n}}, \text{ where}$$

$$\overline{X} = \frac{\Sigma X}{n} \text{ is the mean of these observations}$$

Alternatively

$$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{\Sigma X^2}{n} - (\overline{X})^2}$$

For a discrete data

$$\sigma = \sqrt{\frac{\Sigma f X^2}{\Sigma f} - \left(\frac{\Sigma f X}{\Sigma f}\right)^2}$$

Variance : The variance is the square of standard deviation and is denoted by σ^2

Variance : The variance is the square of standard deviation and is denoted by σ^2.

Formula for C.V. = $\frac{\sigma}{x} \times 100\%$

Where $\overline{X}$ is the arithmetic mean of the given series. It is a relative measure of standard deviation.

Example : 13

Calculate the standard deviation from the following data :

2, 3, 7, 8, 10

Solution :

X	2	3	7	8	10	$\Sigma X = 30$
X^2	4	9	49	64	100	$\Sigma X^2 = 226$

$\bar{X} = 5$

$$\sigma = \sqrt{\frac{\Sigma X^2}{n} - \left(\frac{\Sigma X}{n}\right)^2} = \sqrt{\frac{226}{5} - \left(\frac{30}{5}\right)^2} = \sqrt{45.2 - 36} = 3.033$$

Variance = 9.2 , SD (σ) = 3.033

Example : 14

Calculate the standard deviation, variance and coefficient of variation for the following data :

Marks	2	5	6	7	9
No. of Students	2	3	8	5	2

Solution :

Marks (X)	f	fx	fX^2 unbold
2	2	4	8
5	3	15	75
6	8	48	288
7	5	35	245
9	2	18	162
	$\Sigma f = 20$	$\Sigma fx = 120$	$\Sigma fX^2 = 778$

$$\text{S.D} = \sqrt{\frac{\Sigma fX^2}{\Sigma f} - \left(\frac{\Sigma fX}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{778}{20} - \left(\frac{120}{20}\right)^2}$$

$$= \sqrt{38.9 - 36}$$

$$= \sqrt{2.9} = 1.703$$

$\therefore$ variance = 2.9

$$\text{Coefficient of variation (C.V)} = \frac{1.703}{6}$$

$$= 0.28$$

$$di = \frac{x - A}{i}$$

Where A is assumed mean ; i is width of CI.

$$S.D = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd}{\Sigma f}\right)^2} \times i$$

$$= \sqrt{\frac{28}{20} - \left(\frac{4}{20}\right)^2} \times 5$$

$$= \sqrt{1.44 - 0.04} \times 5$$

$$= 1.166 \times 5$$

$$S.D = \mathbf{5.830}$$

Example : 16

Calculate the standard deviation for the frequency distribution of marks of 100 candidates given below. Also calculate coefficient of variance.

Marks	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
Frequency	5	12	32	40	11

Solution :

Let the assumed mean be 50.

Marks	X	f	fx	$d = X - A$	d^2	fd	fd^2
0 – 20	10	5	50	–40	1,600	–200	8,000
20 – 40	30	12	360	–20	400	–240	4,800
40 – 60	50	32	1,600	0	0	0	0
60 – 80	70	40	2,800	20	400	800	16,000
80 – 100	90	11	990	40	1,600	440	17,600
		$\Sigma f = N = 100$	$\Sigma fx = 5,800$			$\Sigma fd = 800$	$\Sigma fd^2 = 46,400$

$$S.D. = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd^2}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{46,400}{100} - \left(\frac{800}{100}\right)^2}$$

$$= \sqrt{464 - 64}$$

$$= \sqrt{400} = 20$$

$$\overline{X} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{5,800}{100} = 58$$

$$C.V. = \frac{S.D.}{\overline{X}} \times 100 = \frac{20}{58} \times 100$$

$$C.V. = 3448$$

Example : 17

The score of two batsmen 'A' and 'B' in ten innings during certain season are as under which of the batsmen is more consistent in scoring?

A	32	28	47	63	71	39	10	60	96	14
B	19	31	48	53	67	90	10	62	40	80

Example : 16

Calculate the standard deviation for the frequency distribution of marks of 100 candidates given below. Also calculate coefficient of variance.

Marks	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
Frequency	5	12	32	40	11

Solution :

Let the assumed mean be 50.

Marks	X	f	fx	$d = \text{X} - \text{A}$	d^2	fd	fd^2
0 – 20	10	5	50	–40	1,600	–200	8,000
20 – 40	30	12	360	–20	400	–240	4,800
40 – 60	50	32	1,600	0	0	0	0
60 – 80	70	40	2,800	20	400	800	16,000
80 – 100	90	11	990	40	1,600	440	17,600
		$\Sigma f = \text{N} = 100$	$\Sigma fx = 5,800$			$\Sigma fd = 800$	$\text{I}fd^2 = 46,400$

$$\text{S.D.} = \sqrt{\frac{\Sigma fd^2}{\Sigma f} - \left(\frac{\Sigma fd^2}{\Sigma f}\right)^2}$$

$$= \sqrt{\frac{46,400}{100} - \left(\frac{800}{100}\right)^2}$$

$$= \sqrt{464 - 64}$$

$$= \sqrt{400} = 20$$

$$\overline{\text{X}} = \frac{\Sigma fx}{\Sigma f}$$

$$= \frac{5,800}{100} = 58$$

$$\text{C.V.} = \frac{\text{S.D.}}{\overline{\text{X}}} \times 100 = \frac{20}{58} \times 100$$

$$\text{C.V.} = 3448$$

2. It is very much used in biological studies.

Combined Standard Deviation

It is possible to find out the standard deviation of a composite series from the standard deviations of its component parts. Let A_1 and A_2 be two groups having N_1 and N_2 items respectively. Let their means be $\overline{X}_1$ and $\overline{X}_2$ and their respective standard deviations be σ_1 and σ_2. Then the combined standard deviation σ or σ_{12} of two groups A_1 and A_2 is given by the formula :

$$\sigma_{12} = \sqrt{\frac{N_1(\sigma_1^2 + d_1^2) + N_2(\sigma_1^2 + d_2^2)}{N_1 + N_2}}$$

Where $d_1 = (\overline{X}_1 - \overline{X})$, $d_2 = (\overline{X}_2 - \overline{X})$

$$\overline{X} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N_1 + N_2}$$ is the combined mean.

Example : 18

The means of two sets of sizes 40 and 60 respectively are 15 and 16 and the standard deviations are 3 and 4. Obtain the mean and standard deviation of the composite set of 100 items when the two sets are pooled together.

Solution :

Here

$n_1 = 40$	$\overline{X}_1 = 15$	$\sigma_1 = 3$
$n_2 = 60$	$\overline{X}_2 = 16$	$\sigma_2 = 4$
$n_1 + n_2 = 100$	$\overline{X} = ?$	$\sigma = ?$

We have,

$$\overline{X} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2} = \frac{40 \times 15 + 60 \times 16}{100} = \frac{600 + 960}{100}$$

$$= \frac{1,560}{100} = 15.60$$

Now, $d_1 = \overline{X}_1 - \overline{X} = 15 - 15.60 = -0.60$

$d_2 = \overline{X}_2 - \overline{X} = 16 - 15.60 = -0.40$

$$\sigma^2 = \frac{n_1 (\sigma_1^2 + d_1^2) + n_2 (\sigma_2^2 + d_2^2)}{n_1 + n_2}$$

$$\frac{40 \times (9 + (-0.60)^2) + 60\ (16 + 0.4^2)}{40 + 60}$$

$$= \frac{40 \times 9.36 + 60 \times 16.16}{100}$$

$$\frac{374.4 + 969.6}{100} = 13.44$$

$\therefore$ S.D $(\sigma) = \sqrt{13.44} = 3.67$

Example : 19

The arithmetic means of two samples of sizes 60 and 90 are 52 and 48 respectively. The standard deviations are 9 and 12 respectively. Find the standard deviation of the combined sample of size 150.

Solution :

Here

$$n_1 = 60 \qquad \overline{X}_1 = 52 \qquad \sigma_1 = 9$$

$$n_2 = 90 \qquad \overline{X}_2 = 48 \qquad \sigma_2 = 12$$

$$n_1 + n_2 = 150 \qquad \overline{X} = ? \qquad \sigma = ?$$

$$\overline{X} = \frac{n_1\ \overline{X}_1 + n_2\ \overline{X}_2}{n_1 + n_2} = \frac{60 \times 52 + 90 \times 48}{150} = \frac{7{,}440}{150} = 49.6$$

$$d_1 = \overline{X}_1 - \overline{X} = 52 - 49.60 = 2.40 \Rightarrow d_1^2 = 5.76$$

$$d_2 = \overline{X}_2 - \overline{X} = 16 - 15.60 = -1.60 \Rightarrow d_2^2 = 2.56$$

$$\therefore \quad \sigma^2 = \frac{n_1\,(\sigma_1^2 + d_1^2) + n_2\,(\sigma_2^2 + d_2^2)}{n_1 + n_2} = \frac{60\ (81 + 86.76) + 90\ (144 + 2.56)}{90 + 60}$$

$$= \frac{60 \times 86.76 + 90 \times 146.56}{150}$$

$$= \frac{5205.6 + 13190.4}{150} = \frac{18{,}396}{150} = 122.64$$

$$\sigma^2 = 122.64 \qquad \therefore \quad \sigma = \sqrt{122.64}$$

$$\therefore \text{S.D}\ (\sigma) = 11.07$$

Find mean and S.D of the distribution formed by the two distributions taken together.

[**Ans. :** $\overline{X}$ = 50, S.D = 6.7]

18. The mean and S.D of 20 items is found to be 10 and 2 respectively. At the time of checking, it was found that one item 8 was incorrect. Calculate the mean and S.D if (a) the wrong item is omitted, (b) it is replaced by 12. [**Ans. :** (i) $\overline{X}$ = 10.9, SD = 2, (ii) $\overline{X}$ = 10.2, SD = 1.99]
19. The mean and standard deviation of 100 items are found to be 40 and 10. At the time of calculations, two items are wrongly taken as 30 and 72 instead of 3 and 27. Find the correct mean and S.D. [**Ans. :** $\overline{X}$ = 39.28, SD = 10.18]
20. The mean and standard deviation of a sample of 50 observations are found to be 30 and 2 respectively. Later on it was discovered that one observation 35 was wrongly recorded as 25. Recalculate the mean and S.D if
    i) Wrong observation is dropped
    ii) Wrong observation is replaced by the correct observation.

    [**Ans. :** (i) $\overline{X}$ = 30.1, SD = 1.92, (ii) $\overline{X}$ = 30.2, SD = 1.99]
21. State whether the following statements are True/False. In case of false statement, give the correct statement.
    i) Algebraic sum of deviations from mean is minimum.
    ii) Mean deviation is least when calculated from median.
    iii) Variance is always non-negative.
    iv) Mean, standard deviation and coefficient of variation have same unit.
    v) Mean and standard deviation are independent of change or origin.
    vi) Relative measure of dispersion are independent of units of measurement.
    vii) Variance is the square of standard deviation.
    viii) Standard deviation is independent of change or origin and scale.
    ix) Variance is the minimum value of mean square deviation.
    x) Mean deviation can never be negative.
    xi) If each value in a distribution of 5 observations is 10, then its mean is 10 and variance is 1.
    xii) If mean and S.D of a distribution are 20 and 4 respectively, then C.V = 15%.
    xiii) The mean and S.D of 100 observations are 50 and 10 respectively. If 2 is added to each observation then the new S.D is 20.
    xiv) Range is the best measure of deviation.
    xv) Standard deviation can be calculated from any average.

    [**Ans. : True :** i, ii, iii, iv, v, vi, vii, viii, x, xii ; **False :** ix, xi, xiii, xiv, xv]
22. Fill in the blanks :
    i) Standard deviation is always _______ than range.
    ii) All relative measure of dispersion are ______ from units of measurement.
    iii) Variance is ______ of standard deviation.
    iv) When mean is 79 and variance is 64, then C.V is ______.
    v) The S.D of a distribution is 7. If each item is increased by 3, the new S.D is ______.
    vi) If Q_1 = 10, Q_3 = 40, the coefficient of quartile deviation is ______.
    vii) If in a series C.V is 20 and mean is 40, then S.D is ______.

    [**Ans. :** (i) better ; (ii) free ; (iii) square ; (iv) 10.1% ; (v) 7 ; (vi) 0.6 ; (vii) 80]

7 ELEMENTARY PROBABILITY THEORY

7.1 Introduction

In life, an action performed by us may yield to different results. Consider the simple example of throwing a six-faced cubic die. The uppermost face can show any one of the six numbers 1, 2, 3, 4, 5, 6. No one can say with certainty as to which number will show up at a particular time. Hence we have to estimate the chance that a particular number will come up. If the die is not a crooked die, then obviously each number has an equal chance of turning up. Since there are six possible results, we can say that each number has '1 chance in 6' of turning up. Such a chance is called "probability". In this chapter, we will learn various methods to calculate the probability for a given experiment. We will also make extensive use of Permutations and Combinations , so students are requested to revise those concepts before starting with this chapter.

7.2 Probability

The word probability is derived from the word 'probable'.

When the outcome of any event is **not certain**, the theory of probability enters the scene.

For example , when a coin is tossed , whether it will show Head or Tail , is not certain , and we say that "probably it may show a Head or maybe a Tail" !

The theory of probability calculates the chances of occurrence of an outcome.

Consider the example of a coin , for a fair toss , it is anybody's guess that the chances of occurrence of a Head or a Tail are equal , i.e., 50 : 50. However , this is just a common sense. The same common sense , supported by theoretic principles , enfolds the fascinating world of probability theory in front of us. So let us learn and enjoy the concept called Probability !

We begin with Basic Terminology.

7.21 Basic Terminology

1. Random Experiment :

Any action , whose result is uncertain , not pre-decided is called a random experiment or **trial.** Each result of the experiment is called an **outcome** of the experiment.

2. Sample Space :

The set of all possible outcomes of an experiment or trial is called the sample space of that experiment and is denoted by S.

This will be **Universal Set** for the said experiment.

We shall be considering only those experiments whose sample spaces are finite. Hence , n (S) , the number of elements of S is a natural number.

Let us study S for some experiments and find n (S).

i) Consider the simple experiment of tossing a coin. Since the coin can come up heads or tails , the sample space S is given by

$$S = \{H, T\}.$$

$$\therefore \quad n(S) = 2.$$

ii) If 3 coins are tossed , each can give one of the two results : H and T. Hence , S contains $2 \times 2 \times 2 = 8$ outcomes given by

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

$$\therefore \quad n(S) = 8.$$