$$(\overline{X}_{X,Y}) = \frac{(500 \times 186) + (600 \times 175)}{500 + 600} = \frac{1,98,000}{1,100} = ₹\,180$$

(ii) Income of village X = 500 × 186 = ₹ 93,000

Income of village Y = 600 × 175 = ₹ 1,05,000

Thus, Village Y has a larger income.

(iii) Coefficient of Variation of Village X $(C.V._X) = \frac{\sigma}{\overline{X}_x} \times 100 = \frac{9}{186} \times 100 = 4.84\%$

Coefficient of Variation of Village Y $(C.V._Y) = \frac{\sigma}{\overline{X}_y} \times 100 = \frac{10}{175} \times 100 = 5.71\%$

There is more variability in Village Y.

Ans. (i) Average income of the village X and Y taken together = ₹ 180;

(ii) Village Y has a larger income;

(iii) In village Y, variation in income is greater.

Example 48. For a group of 200 candidates, the mean and standard deviation were found to be 40 and 15. Later on it was discovered that the score 43 was misread as 53. Find the correct mean and standard deviations corresponding to the corrected figure.

Solution:

Calculation of Correct Mean

$$\overline{X} = \frac{\Sigma X}{N}$$

Or, $\Sigma X = \overline{X}N$

$\overline{X} = 40;\ N = 200$

$\Sigma X = 40 \times 200 = 8,000$

But 8,000 is a wrong value as one score was misread as 53 instead of 43

Correct ΣX = 8,000 – incorrect item + correct item = 8,000 – 53 + 43 = 7,990

$$\text{Correct } \overline{X} = \frac{\Sigma X}{N} = \frac{7,990}{200} = 39.95$$

Calculation of Correct Standard deviation

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

$$15 = \sqrt{\frac{\Sigma X^2}{200} - (40)^2}$$

$$15 = \sqrt{\frac{\Sigma X^2}{200} - 1,600}$$

Squaring both the sides

$$225 = \frac{\Sigma X^2}{200} - 1,600$$

$225 \times 200 = \Sigma X^2 - 1,600 \times 200$

$\Sigma X^2 = 3,20,000 + 45,000 = 3,65,000$

But, it is incorrect value

Correct ΣX^2 = Incorrect ΣX^2 – (Incorrect items)^2 + (Correct items)^2

Correct $\Sigma X^2 = 3,65,000 - (53)^2 + (43)^2 = 3,65,000 - 2,809 + 1,849 = 3,64,040$

$$\text{Correct } \sigma = \sqrt{\frac{\text{Correct } \Sigma X^2}{N} - (\text{Correct } \overline{X})^2}$$

$$\text{Correct } (\sigma) = \sqrt{\frac{3,64,040}{200} - (39.95)^2} = \sqrt{1,820.2 - 1,596} = \sqrt{224.2} = 14.97$$

Ans. Correct Mean = 39.95 marks; Correct Standard Deviation = 14.97 marks

Example 49. Calculate variance and coefficient of variation from the following data:

Values	2	6	10	14
Frequency	4	8	2	1

Solution:

Values (X)	Frequency (f)	fX	$x = X - \overline{X}$	x^2	fx^2
2	4	8	–4	16	64
6	8	48	0	0	0
10	2	20	+4	16	32
14	1	14	+8	64	64
	N = Σf = 15	ΣfX = 90			$\Sigma fx^2 = 160$

$$\text{Arithmetic Mean } (\overline{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{90}{15} = 6$$

$$(\sigma) = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{160}{15}} = 3.2659$$

Variance = $\sigma^2 = (3.2659)^2 = 10.66$

$$\text{Coefficient of Variation} = \frac{\sigma}{\overline{X}} \times 100 = \frac{3.2659}{6} \times 100 = 54.43\%$$

Ans. Variance = 10.66; Coefficient of Variation = 54.43%

Example 50. For the following data, calculate: (i) Standard Deviation; (ii) Variance; (iii) Coefficient of Standard Deviation; (iv) Coefficient of Variation.

Class	10–19	20–29	30–39	40–49	50–59
Frequency	4	5	6	2	3

Solution:

This is a case of inclusive class-intervals. So, it has to be converted into exclusive series.

Marks (X)	No. of students (f)	Mid-point (m)	d = m − A (A = 24.5)	$d' = \frac{m-A}{C}$ C = 10	fd'	d'²	fd'²
9.5–19.5	4	14.5	− 10	− 1	− 4	1	4
19.5–29.5	5	**24.5 (A)**	0	0	0	0	0
29.5–39.5	6	34.5	+ 10	+ 1	+ 6	1	6
39.5–49.5	2	44.5	+ 20	+ 2	+ 4	4	8
49.5–59.5	3	54.5	+ 30	+ 3	+ 9	9	27
	N = Σf = 20				**Σfd′ = 15**		**Σfd′² = 45**

(i) Standard deviation (σ) $= \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$

$\Sigma fd'^2 = 45; N = 20; \Sigma fd' = 15; C = 10$

$$\sigma = \sqrt{\frac{45}{20} - \left(\frac{15}{20}\right)^2} \times 10 = \sqrt{2.25 - .5625} \times 10 = \sqrt{1.6875} \times 10 = 12.99$$

(ii) Variance $= \sigma^2 = (12.99)^2 = 168.74$

(iii) We know: Coefficient of Standard Deviation $= \frac{\sigma}{\bar{X}}$

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd'}{\Sigma f} \times C = 24.5 + \frac{15}{20} \times 10 = 32$$

$$\text{Coefficient of Standard Deviation} = \frac{12.99}{32} = 0.406$$

(iv) Coefficient of Variation (C.V.) $= \frac{\sigma}{\bar{X}} \times 100 = \frac{12.99}{32} \times 100 = 40.6\%$

Ans. (i) Standard Deviation = 12.99; (ii) Variance = 168.74; (iii) Coefficient of Standard Deviation = 0.406; (iv) Coefficient of Variation = 40.6%.

Example 51. From the following data, find out which group is more uniform.

Age Group (years)	No. of Persons	
	Group A	Group B
0–10	5	7
10–20	15	12
20–30	20	22
30–40	25	30
40–50	18	20
50–60	10	5
60–70	7	4

Solution:

In order to find which group is more uniform, we shall have to compare the coefficient of variation (C.V.) of the two groups.

Calculation of Coefficient of Variation (Group A)

Age Group (X)	No. of persons (f)	Mid-points (m)	d = m − A (A = 25)	$d' = \frac{m-A}{C}$ C = 10	fd'	d'²	fd'²
0–10	5	5	− 20	− 2	− 10	4	20
10–20	15	15	− 10	− 1	− 15	1	15
20–30	20	**25 (A)**	0	0	0	0	0
30–40	25	35	+ 10	+ 1	+ 25	1	25
40–50	18	45	+ 20	+ 2	+ 36	4	72
50–60	10	55	+ 30	+ 3	+ 30	9	90
60–70	7	65	+ 40	+ 4	+ 28	16	112
	N = Σf = 100				**Σfd′ = 94**		**Σfd′² = 334**

To calculate coefficient of variation, we will first calculate standard deviation and arithmetic mean.

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C = \sqrt{\frac{334}{100} - \left(\frac{94}{100}\right)^2} \times 10$$

$$\sigma = \sqrt{3.34 - 0.883} \times 10 = 15.67$$

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd'}{\Sigma f} \times C = 25 + \frac{94}{100} \times 10 = 34.4$$

$$\text{Coefficients of Variation (C.V.)} = \frac{\sigma}{\bar{X}} \times 100 = \frac{15.67}{34.4} \times 100 = 45.55\%$$

Calculation of Coefficient of Variation (Group B)

Age Group (X)	No. of persons (f)	Mid-points (m)	d = m − A (A = 25)	$d' = \frac{m-A}{C}$ C = 10	fd'	d'²	fd'²
0–10	7	5	− 20	− 2	− 14	4	28
10–20	12	15	− 10	− 1	− 12	1	12
20–30	22	**25 (A)**	0	0	0	0	0
30–40	30	35	+ 10	+ 1	+ 30	1	30
40–50	20	45	+ 20	+ 2	+ 40	4	80
50–60	5	55	+ 30	+ 3	+ 15	9	45
60–70	4	65	+ 40	+ 4	+ 16	16	64
	N = Σf = 100				**Σfd′ = 75**		**Σfd′² = 259**

To calculate coefficient of variation, we will first calculate standard deviation and arithmetic mean.

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C = \sqrt{\frac{259}{100} - \left(\frac{75}{100}\right)^2} \times 10$$

$$\sigma = \sqrt{2.59 - 0.5625} \times 10 = 14.24$$

$$\text{Mean } (\bar{X}) = A + \frac{\Sigma fd'}{\Sigma f} \times C = 25 + \frac{75}{100} \times 10 = 32.5$$

$$\text{Coefficient of Variation (C.V.)} = \frac{\sigma}{\bar{X}} \times 100 = \frac{14.24}{32.5} \times 100 = 43.82\%$$

Ans. Coefficient of variation of Group B (43.82%) is less than that of Group A (45.55%), so Group B is more uniform.

10.22 PROPERTIES OF STANDARD DEVIATION

1. ***The sum of the square of the deviations of the items from their arithmetic mean is the minimum.*** The sum is less than the sum of the square of the deviations of the items from any other value.

It is made clear with the following illustration:

X	$X-\bar{X}$ ($\bar{X}=7$)	$(X-\bar{X})^2$	$X-8$	$(X-8)^2$
3	−4	16	−5	25
5	−2	4	−3	9
8	+1	1	0	0
12	+5	25	+4	16
		$\Sigma(X-\bar{X})^2 = 46$		$\Sigma(X-8)^2 = 50$

It is clear from the above example that sum of the squares of deviations from mean (46) is less than the sum of squares of deviations (50) taken from assumed mean.

2. ***Standard deviation is independent of change of origin, i.e.*** value of standard deviation remains the same if in a series, a constant is added (or subtracted) to all observations.
3. ***Standard deviation is affected by change of scale, i.e.*** if all the observations are multiplied or divided by a constant, then the standard deviation also gets multiplied (or divided) by this constant.
4. ***Standard deviation of the combined series:*** Like the arithmetic mean, it is possible to compute combined standard deviations of two or more groups.

   {Combined Standard Deviation is discussed in detail in Section 10.23}
5. For a given set of observations, ***standard deviation is never less than mean deviation from mean,*** i.e., Standard Deviation > Mean Deviation from mean.

10.23 COMBINED STANDARD DEVIATION

As we can calculate mean of two or more than two series, we can also compute combined standard deviation of two or more than two series. The formula in case of two series:

$$\sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

Where

$\sigma_{1,2}$ = Combined standard deviation of two groups

σ_1 = Standard deviation of first group

σ_2 = Standard deviation of second group

$d_1^2 = (\bar{X}_1 - \bar{X}_{1,2})^2$

$d_2^2 = (\bar{X}_2 - \bar{X}_{1,2})^2$

$\bar{X}_{1,2}$ = Combined arithmetic mean of two groups

$\bar{X}_1$ = Arithmetic mean of first group

$\bar{X}_2$ = Arithmetic mean of second group

N_1 = Number of observations of first group

N_2 = Number of observations of second group

This formula can be extended upto N number of series. If there are three series, then the combined standard deviation is:

$$\sigma_{1,2,3} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

Where,

$d_1^2 = (\bar{X}_1 - \bar{X}_{1,2,3})^2$, $d_2^2 = (\bar{X}_2 - \bar{X}_{1,2,3})^2$, and $d_3^2 = (\bar{X}_3 - \bar{X}_{1,2,3})^2$

The concept of combined mean will be more clear from the following examples.

Example 52. Find the combined standard deviation from the following data:

	Boys	Girls
Number	30	20
Mean	20	30
Standard deviation	4	5

Solution:

To calculate combined standard deviation, we will have to first calculate combined mean:

$$\text{Combined Mean } (\bar{X}_{1,2}) = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$\bar{X}_1 = 20$, $\bar{X}_2 = 30$, $N_1 = 30$, $N_2 = 20$

$$\overline{X}_{1,2} = \frac{(30 \times 20) + (20 \times 30)}{30 + 20} = \frac{1,200}{50} = 24$$

Calculation of Combined Standard Deviation

$$\sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$d_1 = \overline{X}_1 - \overline{X}_{1,2} = 20 - 24 = -4$

$d_2 = \overline{X}_2 - \overline{X}_{1,2} = 30 - 24 = 6$

Now: $N_1 = 30, \sigma_1 = 4, N_2 = 20, \sigma_2 = 5, d_1 = -4, d_2 = 6$

$$\sigma_{1,2} = \sqrt{\frac{30(4)^2 + 20(5)^2 + 30(-4)^2 + 20(6)^2}{30 + 20}}$$

$$\sigma_{1,2} = \sqrt{\frac{480 + 500 + 480 + 720}{50}} = \sqrt{43.6} = 6.6$$

Ans. Combined Standard Deviation = 6.6

Example 53. Mean and standard deviations of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the standard deviation of all the 250 items taken together:

Solution:

First we will calculate combined mean:

$$\text{Combined Mean } (\overline{X}_{1,2}) = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

$\overline{X}_1 = 50, \overline{X}_2 = 40, N_1 = 100, N_2 = 150$

$$\overline{X}_{1,2} = \frac{(100 \times 50) + (150 \times 40)}{100 + 150} = \frac{11,000}{250} = 44$$

Calculation of Combined Standard Deviation

$$\sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$d_1 = \overline{X}_1 - \overline{X}_{1,2} = 50 - 44 = 6$

$d_2 = \overline{X}_2 - \overline{X}_{1,2} = 40 - 44 = -4$

Now: $N_1 = 100, \sigma_1 = 5, N_2 = 150, \sigma_2 = 6, d_1 = 6, d_2 = -4$

$$\sigma_{1,2} = \sqrt{\frac{100(5)^2 + 150(6)^2 + 100(6)^2 + 150(-4)^2}{100 + 150}}$$

$$\sigma_{1,2} = \sqrt{\frac{2,500 + 5,400 + 3,600 + 2,400}{250}} = \sqrt{55.6} = 7.456$$

Ans. Combined Standard Deviation = 7.456

10.24 MERITS, DEMERITS AND USES OF STANDARD DEVIATION

Merits of Standard Deviation

1. Based on all Values: Standard deviation considers every item of the series. So, a change in even one value affects the value of standard deviation.
2. Rigidly Defined: Standard deviation is by far the most important and widely used measure of dispersion. It is rigidly defined, i.e. it is a definite measure of dispersion.
3. Less effect of fluctuations in sampling: If several independent samples are drawn from the same population, it may be observed that standard deviation is least affected from sample to sample as compared with other measures of dispersions.
4. Algebraic Treatment: Standard Deviation is capable of further algebraic treatment. *For example,* if standard deviations of a number of groups are known, their combined standard deviation can be computed.
5. Better mathematical process: The squaring of the deviations removes the drawback of ignoring the signs of deviations (as done in case of mean deviation).

Demerits of Standard Deviation

1. Difficult to Compute: Standard deviation is more difficult to be measured as compared to other measures of dispersion.
2. More stress on extreme items: It gives more weightage to extreme values and less to those which are nearer to mean.
3. Depend upon units of measurement: It depends upon the units of measurement of the observations. So, it cannot be used to compare the dispersion of the distributions expressed in different units.

Uses of Standard Deviation

1. Standard deviation can be used to compare the dispersions of two or more distributions when their units of measurements and arithmetic means are same.
2. It is used to test the reliability of mean. Mean of a distribution with least standard deviation is said to be more reliable.

10.25 CHOICE OF A SUITABLE MEASURE OF DISPERSION

We have already studied the merits and demerits of the four measures of dispersion namely range, quartile deviation, mean deviation and standard deviation. *Let us now make a comparative study of all the four measures.* It would help in the selection of an appropriate measure of dispersion for a particular problem under study:

1. Rigidly defined: All the four measures of dispersion are rigidly defined and their values are definite.

2. Calculations: It is easy to calculate range and quartile deviation. But the calculations of mean deviation and standard deviation are a little more complicated.

3. Based on values: The range and quartile deviation do not depend on all values, whereas mean deviation and standard deviations are based on all values.

4. Algebraic Treatment: In terms of algebraic treatment, standard deviation is considered to be the best measure of dispersion.

Conclusion

1. Range: It is the simplest to calculate, but it is an unstable measure as it is considerably affected by the extreme values. This method is advisable only when the variation in the size of items is very little.

2. Quartile deviation: The quartile deviation is a better measure than range as it is not affected too much by the values of extreme items. It is easily calculated and is readily understood. However, quartile deviation has no algebraic properties and its interpretation is difficult.

3. Mean Deviation: Mean deviation is based on all the items, but it ignores the signs of deviation and cannot be used for further algebraic treatment.

4. Standard deviation: Standard deviation is rigidly defined and is based on all the observations. It is capable of algebraic treatment and is not affected very much by fluctuations of sampling. So, the standard deviation scores over all other measures of dispersion.

However, it should be kept in mind that standard deviation gives comparatively greater importance to extreme variations, which should usually be ignored.

Comparison Between Mean Deviation and Standard Deviation

S. No.	Mean Deviation	Standard Deviation
1.	It is based on simple average of the sum of absolute deviations	It is based on the square root of the average of the squared deviations.
2.	Mean deviation can be computed from mean, median or mode and its value differs in these cases (unless the distribution is normal).	The standard deviation is always calculated from the arithmetic mean.
3.	Mean Deviation does not take into account the algebraic signs (plus or minus) in its calculation which is illogical.	In calculation of standard deviation, the deviations are squared. So, the plus and minus signs need not be omitted.
4.	Mean Deviation is not capable of further algebraic treatment as it considers only the absolute values. So, it is not possible to compute combined mean deviation.	Standard deviation is capable of further algebraic treatment, i.e., we can find the combined standard deviation of two or more series.

Comparison Between Different Measures of Dispersion

Basis	Range	Quartile Deviation	Mean Deviation	Standard Deviation
Concept	It is the difference between the largest and the smallest item of a series.	It is half the difference between the third and the first quartile.	It is the arithmetic mean of the deviation of all the values taken from an average.	It is the square root of the arithmetic mean of the squares of deviation of items from their arithmetic mean.
Calculations	It is easiest to calculate and simple to understand.	It is easy to calculate and simple to understand.	It is difficult to calculate as compared to range and quartile deviation, but it is simple to understand.	It involves difficult calculations.
Based on items	It is based on only two items of the series, the largest and the smallest.	It covers only 50% items.	It is based on all the items of the series.	It is based on all the items of the series.
Effect of Extreme Items	It is highly affected by the extreme values.	It is not unduly affected by extreme values.	It is less affected by extreme values.	It is affected by extreme values.

10.26 LORENZ CURVE

The Lorenz Curve is a graphic method of studying dispersion. It was devised by Dr. Max O. Lorenz, a famous economics statistician. *This curve was used by him to measure the inequalities of income or wealth of a society.*

Now a days, the curve is also used to study the distribution of profit, wages, turnover, etc. However, still the most common use of this curve is to show inequality of income or wealth in a country.

Steps involved in Drawing a Lorenz Curve

Step 1. Calculate cumulative values of size of items (in case of discrete series) or mid-points (in case of continuous series).

Step 2. Calculate percentages for these cumulative values. For this, the last cumulative total is considered as equal to 100 and then percentages are obtained.

Step 3. Determine cumulative frequencies.

Step 4. Calculate the percentage for each cumulative frequency. For this, the last cumulative total is considered as equal to 100 and then percentages are obtained.

Step 5. On the X-axis, start from 0 to 100 and take the percentage of cumulative frequencies.

Step 6. On the Y-axis, start from 0 to 100 and take the percentage of variable.

Step 7. Draw a diagonal line joining 0 to 100. This line is known as *"Line of Equal Distribution"* or *"Equality line"*.

Step 8. Plot the various points corresponding to the values of the variables X and Y and then join these points with a smooth free hand curve. The curve so obtained shows the actual distribution. This curve is known as Lorenz Curve.

Important Points of Interpretation of Lorenz Curve

- If the distribution is uniform, the Lorenz Curve will coincide with the line of equal distribution. *Generally, the Lorenz curve lies below the line of equal distribution.*
- The area between the line of equal distribution and the Lorenz curve gives the extent of inequality in the items. The larger the area, more is the inequality.
- If curves of various distributions are shown on the same Lorenz presentation, the curve that is farthest from the diagonal line represents greatest inequality.

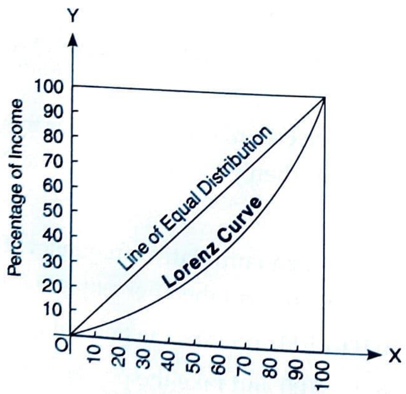
The following example will help us in understanding this phenomenon.

Example 54. From the following details of monthly income, draw a Lorenz curve.

Income (in '000 ₹)	0–10	10–20	20–30	30–40	40–50
No. of persons	4	5	7	5	4

Solution:

Monthly Income (₹)	Mid-Value (₹)	Cumulative Mid-values	% Cumulative Mid-values	No. of Persons (f)	Cumulative Frequency (c.f.)	% Cumulative Frequency
0–10	5	5	4	4	4	16
10–20	15	20	16	5	9	36
20–30	25	45	36	7	16	64
30–40	35	80	64	5	21	84
40–50	45	125	100	4	25	100



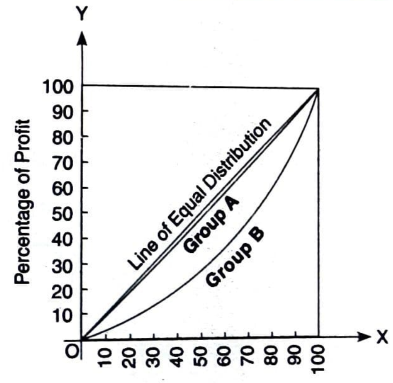
The curve drawn is farther from the line of equal distribution. So, there is inequality in distribution of income.

Example 55. From the following table, draw Lorenz curve for number of persons in Group A and B and interpret the result.

Profit Earned (₹ in '000)	20	30	40	50	60
No. of Persons (Group A)	6	8	10	12	14
No. of Persons (Group B)	15	10	9	11	5

Solution:

Profit Earned (₹ in '000)	Cumulative Profit (₹)	Cumulative (%)	Group A			Group B		
			No. of Persons	Cumulative No.	Cumulative (%)	No. of Persons	Cumulative No.	Cumulative (%)
20	20	10	6	6	12	15	15	30
30	50	25	8	14	28	10	25	50
40	90	45	10	24	48	9	34	68
50	140	70	12	36	72	11	45	90
60	200	100	14	50	100	5	50	100



The Lorenz curve of Group B is farthest from the line of equal distribution. So, Group B shows greater inequality as compared to Group A.

10.27 MERITS AND DEMERITS OF LORENZ CURVE

Merits of Lorenz Curve

1. Lorenz Curve is attractive and *it gives a rough idea of extent of dispersion.*
2. With the help of Lorenz curve, *it becomes easy to compare two or more series.*

Demerits of Lorenz Curve

1. Lorenz curve gives only a relative idea of the dispersion as compared with the line of equal distribution. It *does not provide us any numerical value of the variability for the given distribution.*
2. The *method of drawing Lorenz Curve is very difficult.*

FORMULAE AT A GLANCE

1. RANGE

1.1 Absolute Measure

Range = L − S

1.2 Relative Measure

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

Where, L = Largest item, and S = Smallest item

Note:

1. *In case of continuous series, L, i.e. largest item, will be the upper limit of the highest class and smallest item (S) will be the lower limit of the lowest class.*
2. *It both discrete and continuous series, the frequencies are immaterial (As range depends on two extreme observations).*

2. QUARTILE DEVIATION

2.1 Absolute Measure

Interquartile Range = $Q_3 - Q_1$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2}$$

2.2 Relative Measure

$$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Where,

Q_1 = Lower Quartile; Q_3 = Upper Quartile

Note: *In case of open-end series, Quartile Deviation is an appropriate measure of dispersion.*

3. MEAN DEVIATION

3.1 Absolute Measure

	Individual Series	*Discrete Series*	*Continuous Series*
Mean Deviation from Mean	$\frac{\Sigma\lvert X-\bar{X}\rvert}{N} = \frac{\Sigma\lvert D\rvert}{N}$	$\frac{\Sigma f\lvert X-\bar{X}\rvert}{N} = \frac{\Sigma f\lvert D\rvert}{N}$	$\frac{\Sigma f\lvert m-\bar{X}\rvert}{N} = \frac{\Sigma f\lvert D\rvert}{N}$
Mean Deviation from Median	$\frac{\Sigma\lvert X-Me\rvert}{N} = \frac{\Sigma\lvert D\rvert}{N}$	$\frac{\Sigma f\lvert X-Me\rvert}{N} = \frac{\Sigma f\lvert D\rvert}{N}$	$\frac{\Sigma f\lvert m-Me\rvert}{N} = \frac{\Sigma f\lvert D\rvert}{N}$

Where, m = Mid-point

3.2 Relative Measure

$$\text{Coefficient of Mean Deviation from Mean } (\bar{X}) = \frac{MD_{\bar{X}}}{\bar{X}}$$

$$\text{Coefficient of Mean Deviation from Median (Me)} = \frac{MD_{Me}}{Me}$$

4. STANDARD DEVIATION

4.1 Absolute Measure

Standard Deviation

	Individual Series	*Discrete Series*	*Continuous Series*
Actual Mean Method	$\sigma = \sqrt{\frac{\Sigma x^2}{N}}$	$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$	$\sigma = \sqrt{\frac{\Sigma fx^2}{N}}$
Direct Method	$\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$	$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2}$	$\sigma = \sqrt{\frac{\Sigma fm^2}{N} - (\bar{X})^2}$
Short-Cut Method	$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$	$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$	$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$
Step Deviation Method		$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$	$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$

Variance = σ^2

4.2 Relative Measure

$$\text{Coefficient of Standard Deviation} = \frac{\sigma}{\bar{X}}$$

$$\text{Coefficient of Variation (C.V.)} = \frac{\sigma}{\bar{X}} \times 100$$

5. COMBINED STANDARD DEVIATION

$$\text{Two Related Groups: } \sigma_{1,2} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$\text{Three Related Groups: } \sigma_{1,2,3} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

ms from a measure of

e deviations measured

culate standard deviation

easuring it.

...tems; (iii) Affect of Extreme Items. ...iation and Standard Deviation on the basis of:

Q. 12. What is the meaning of Lorenz Curve? State the steps involved in drawing a Lorenz Curve.

Unsolved Practicals

Range

Individual Series

1. Calculate Range and coefficient of range of the following series, which gives the monthly expenditure (in ₹) of seven students: 22, 35, 32, 45, 42, 48, 39

{Range = ₹ 26; Coefficient of range = 0.37}

2. Find range and coefficient of range from the weekly wage (in ₹) of 10 workers of a factory: 310, 350, 420, 105, 115, 290, 245, 450, 300, 375.

{Range = ₹ 345; Coefficient of range = 0.62}

Discrete Series

3. Find the range and coefficient of range of the following:

Size of shoes	6	7	8	9	10	11	12	13
No. of shoes sold	8	12	14	20	15	8	7	10

{Range = 7; Coefficient of range = 0.37}

4. From the following data calculate range and coefficient of range:

Marks	10	20	30	40	50	60	70
No. of students	8	12	7	30	10	5	2

{Range = 60 marks; Coefficient of range = [illegible].75}

12. Calculate coefficient of quartile deviation from the following data:

X (Less than)	200	300	400	500	600
Frequency	8	20	40	46	50

{Coefficient of Quartile Deviation = 0.24}

13. Estimate an appropriate measure of dispersion for the following data:

Wages (₹)	Less than 25	25–30	30–35	35–40	Above 40
No. of workers	2	10	26	16	7

Hint: *It is an open-end series. So, it will be appropriate to find out quartile deviation and its coefficient.*

{Quartile Deviation = ₹3.40; Coefficient of Quartile Deviation = 0.1}

Mean Deviation

Individual Series

14. Calculate the mean deviation from median and its coefficient from the following data: 100, 150, 80, 90, 160, 200, 140

{Mean deviation from Median = 34.28; Coefficient of Mean deviation = 0.245}

15. Compute Mean deviation and its coefficient by mean from the data given below:

X	210	220	225	225	225	235	240	250	270	280

{Mean Deviation = 17.6; Coefficient of Mean Deviation = 0.074}

Discrete Series

16. Following are the marks of the students. Find mean deviation and coefficient mean deviation from mean.

Marks	5	10	15	20	25	30	35	40
No. of students	16	32	36	44	28	18	12	14

{Mean deviation = 7.65 marks; Coefficient of Mean deviation = 0.38}

17. Find out the mean deviation from the median and its coefficient.

Marks	10	11	12	13	14
No. of students	3	12	18	12	3

{Mean deviation = 0.75 marks}

Continuous Series

18. Compute the mean deviation from the median and from the mean for the following distribution of the scores of 50 college students:

Scores	140–150	150–160	160–170	170–180	180–190	190–200
Frequency	4	6	10	18	9	3

{Mean deviation from mean = 10.56 marks; Mean deviation from median = 10.24 marks}

19. Find the mean deviation from mean and its coefficient for the given data:

X	0–10	10–20	20–30	30–40	40–50	50–60
F	3	5	7	2	9	4

{Mean deviation from mean = 14.33; Coefficient of Mean deviation = 0.44}

20. Calculate mean deviation and its coefficient from the following figures:

Class-Interval	Frequency
Less than 10	5
Less than 20	12
Less than 30	20
Less than 40	35
Less than 50	54
Less than 60	60

Hint: *Mean deviation is calculated from median as it is a constant and representative value.*

{Mean deviation from median = 11.61; Coefficient of Mean deviation = 0.32}

Standard Deviation

Individual Series

21. Calculate the standard deviation from the following values: 8, 9, 15, 23, 5, 11, 19, 8, 10, 12.

{Standard deviation = 5.23}

22. Find the standard deviation for the following data: 3, 5, 6, 7, 10, 12, 15, 18.

{Standard deviation = 4.87}

23. Find out the standard deviation of the height of 10 men given below: 160, 160, 161, 162, 163, 163, 163, 164, 164, 170.

{Standard deviation = 2.72}

Discrete Series

24. Calculate standard deviation of the given below:

Size	3	4	5	6	7	8	9
Frequency	3	7	22	60	85	32	8

{Standard deviation = 1.149}

25. Find the value of standard deviation and coefficient of variation from the following:

Variables	10	20	30	40	50	60	70
Frequency	6	8	16	15	32	11	12

{Standard deviation = 16.43; Coefficient of SD = 37.34}

26. Measurements are made to the nearest cm. of the heights of 10 children. Calculate mean and standard deviation.

Height (cms)	60	61	62	63	64	65	66	67	68
No. of children	2	0	15	29	25	12	10	4	3

{Mean = 63.89 cms; Standard deviation = 1.6 cms}

Continuous Series

27. Calculate standard deviation for the given data:

Age (in yrs)	20–25	25–30	30–35	35–40	40–45	45–50	50–55
No. of workers	17	11	8	5	4	3	2

{Standard deviation = 8.79 years}

28. Calculate Standard deviation from the following series:

Class	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	2	4	6	8	6	4	2

{Standard deviation = 15.81}

29. From the following figures, find the standard deviation and the coefficient of variation:

Marks	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
No. of persons	5	10	20	40	30	20	10	4

{Standard deviation = 15.69 marks; Coefficient of variation = 39.83%}

Combined Standard Deviation

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

{Combined Mean = 51.57; Combined Standard deviation = 7.56}

32. For a group of 100 males, mean and standard deviation of their daily wages are ₹ 36 and ₹ 9 respectively. For a group of 50 females, it is ₹ 45 and ₹ 6. Find the standard deviation for the whole group.

{Combined Mean = 39; Combined Standard deviation = 9.16}

Lorenz Curve

33. The profit of two business concerns for 5 years are as given below. Draw Lorenz Curves to show the distribution.

Year	2001	2002	2003	2004	2005
Firm A	15	30	45	60	50
Firm B	20	30	45	60	45

34. The given table shows the daily income of workers of two factories. Draw the Lorenz Curves for both the factories.

Daily Income (₹)	0–100	100–200	200–300	300–400	400–500
Factory A	8	7	5	3	2
Factory B	15	6	2	1	1

Miscellaneous Questions

35. Find the mean deviation from mean and its coefficient for the given data:

Marks (more than)	0	10	20	30	40	50	60
No. of students	200	180	150	100	40	15	5

{Mean deviation from mean = 11.5 marks; Coefficient of Mean deviation = 0.39}

36. Calculate the Mean and Standard Deviation from the following distribution.

Age (years)	15–19	20–24	25–29	30–34	35–39	40–44
No. of Persons	4	20	38	24	10	4

Hint: Convert above series into exclusive class-intervals and, thereafter, calculate mean and standard deviation.

{Mean = 28.4 years; Standard deviation = 5.66 years}

37. The following table gives the weights of one hundred persons. Compute the coefficient of dispersion by the method of limits.

Class-interval	40–45	45–50	50–55	55–60	60–65	65–70	70–75	75–80	80–85	85–90
No. of persons	4	13	8	14	9	16	17	9	8	2

{Range = 50 kg. Coefficient of range = 0.384}

38. Calculate standard deviation of the following data:

Age in years (below)	10	20	30	40	50	60	70	80
No. of persons	15	30	53	75	100	110	115	125

Hint: Make above series as continuous and, thereafter, calculate standard deviation.

{Standard deviation = 19.76 years}

39. Price of a particular item in 10 years in two cities are given below, which city has more stable prices?

City A	55	54	52	53	56	58	52	50	51	49
City B	108	107	105	105	106	107	104	103	104	101

(C.V. (City A) = 4.99%; CV (City B) = 1.90%; City B has more stable prices)

40. For a distribution, the coefficient of variation is 22.5% and mean is 7.5. Calculate standard deviation.

(Standard deviation = 1.69)

41. Find out the arithmetic mean and standard deviation from the following data:

Variable	5–10	10–15	15–20	20–25	25–30	30–35
Frequency	2	9	29	54	11	5

{Arithmetic mean = 21.05; Standard deviation = 4.88}

42. The mean and standard deviation of a series of 20 items are 20 and 5 respectively. While calculating these measures, an item of 13 was wrongly read as 30. Find out the correct mean and standard deviation.

{Correct Mean = 19.15; Standard deviation = 4.66}

43. Following are the marks obtained by two students: Mollie and Isha, in 10 sets of examinations:

Marks of obtained by Mollie	44	80	76	48	52	72	68	56	60	54
Marks of obtained by Isha	48	75	54	60	63	69	72	51	57	66

Out of Mollie and Isha, who is more consistent?

{Isha is more consistent in securing the marks as her C.V(14.01%) is less than that of Mollie's C.V. (19.18%)}

44. Calculate coefficient of variation from the following data:

Marks (more than)	0	10	20	30	40	50	60	70
No. of students	100	90	75	50	20	10	5	0

{Coefficient of variation = 50%}

45. In two Towns A and B, daily pocket money and the standard deviations are given below:

Town	Average Daily Pocket	Standard Deviation	No. of Teenagers
A	34.5	5.0	476
B	28.5	4.5	524

(i) Which town, A or B, pays out the larger amount of daily pocket money?
(ii) What is the average daily pocket money of all teenagers taken together?
(iii) Calculate coefficient of variation of each town. Which town is more variable in terms of pocket money?

{(i) Pocket Money: Town A = ₹ 16,422; Town B = ₹ 14,934. Town A pays larger amount of daily pocket money; (ii) Combined Pocket Money = ₹ 31.36; (iii) C.V. (Town A) = 14.49%; C.V. (Town B) = 15.79%. Town B is more variable in terms of pocket money as its C.V. is higher}

46. The prices of share of Company X and Company Y are given below. State, which company is more stable?

Company X	25	50	45	30	70	42	36	48	34	60
Company Y	10	70	50	20	95	55	42	60	48	80

{C.V. of Prices of Share of X Co. = 29.72%; C.V. of Prices of Share of Y Co. = 45.94%; Prices of Share of X Co. is more stable}

47. A student obtained the mean and standard deviation of 100 observations as 40 and 5.1 respectively. It was later found that one observation was wrongly copied as 50 instead of 40. Find the correct mean and standard deviation.

{Correct Mean = 39.9; Standard deviation = 5}

48. From the following data, calculate standard deviation of the two groups A and B. Which group is more consistent?

Class-Interval	Group A	Group B
5–10	2	9
10–15	9	11
15–20	29	18
20–25	54	32
25–30	11	27
30–35	5	13

{$\sigma_A = 4.88$; $\sigma_B = 7.07$; Group A is more consistent as C.V. of Group A (23.18%) is less than C.V. of Group B (32.36%)}

49. From the following data of marks, calculate standard deviation. What will be the value of standard deviation, if marks obtained by each student is increased by one?

Marks Obtained	1	2	3	4	5	6	7	8	9
No. of students	32	41	57	98	123	83	46	17	3

Hint: *Standard deviation is independent of the change of origin, i.e. it is not affected if each value of the series is increased or decreased by a constant quantity. So, the standard deviation will remain the same even if marks of each student is increased by one.*

(Standard deviation = 1.7741 marks)

50. From the following data of two workers, identify who is a more consistent worker?

	Worker	
	A	B
Average Time in completing a job	40	42
Standard Deviation	8	6

(Worker B is more consistent as his C.V. (14.29%) is less than that of worker A (20%))

51. Find the standard deviation and coefficient of standard deviation:

X (less than)	10	20	30	40	50	60	70	80
Frequency	12	30	65	107	157	202	222	230

(Standard deviation = 17.26; Coefficient of Standard deviation = 0.43)

52. Find mean, standard deviation and coefficient of variation.

Class-interval	0–4	4–8	8–12	12–16	16–20	20–24
Frequency	10	15	20	25	20	10

(Mean = 12.4; Standard deviation = 5.85; C.V. = 47.19%)

53. The following table shows the marks obtained by 60 students. Calculate mean and standard deviation.

Marks (more than)	70	60	50	40	30	20
No. of students	7	18	40	40	55	60

Hint: *First, convert the given series as ordinary continuous series and then use the step deviation method.*

(Mean = 51.67 marks; Standard deviation = 15.13 marks)

54. Calculate standard deviation and coefficient of dispersion from the data below:

Mid-Points	5	15	25	35	45	55	65	75
Frequency	5	8	7	12	28	20	10	10

Hint: *For coefficient of dispersion, calculate coefficient of variation as it is a better measure in comparison to coefficient of standard deviation.*

(Standard deviation = 18.49; C.V. = 41.09%)

55. The following data shows the expected life of two models of T.V.: A and B:

Life (No. of years)	0–2	2–4	4–6	6–8	8–10	10–12
Model A	5	16	13	7	5	4
Model B	2	7	12	19	9	1

Which modal has greater uniformity?

(C.V. of Model A = 54.91%; C.V. of Model B = 36.21%. Model B has greater uniformity)

11

MEASURES OF CORRELATION

LEARNING OBJECTIVES

11.1 INTRODUCTION

In the earlier chapters, we have studied the statistical problems and distributions relating to one variable. We discussed various measures of central tendency and dispersion, which are confined to a single variable. This kind of statistical analysis involving one variable is known as *univariate distribution*.

But, we may come across a number of situations with distributions having two variables. For example, we may have data relating to income and expenditure, price and demand, height and weight, etc. The distribution involving two variables is called is called *bivariate distribution*.

In a bivariate distribution, we may be interested to find if there is any relationship between the two variables under study. In day-to-day life, we observe that there exists certain relationship between two variables, like between income and expenditure, price and demand and so on. **Correlation** *is a statistical tool which studies the relationship between two variables.*