5

Measures of Variation

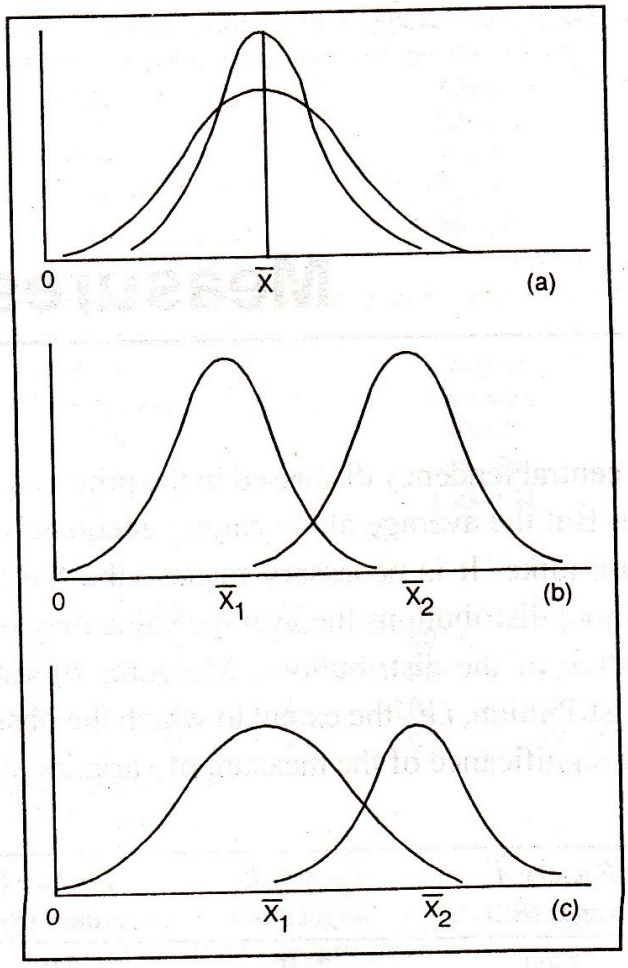
INTRODUCTION

The various measures of central tendency discussed in the previous chapter give us one single value that represents the entire data. But the average alone cannot adequately describe a set of observations, unless all the observations are alike. It is necessary to describe the *variability or dispersion of the observations.* Also in two or more distributions the average value may be the same but still there can be wide disparities in the formation of the distributions. Measures of variation help us in studying the important characteristic of a distribution, *i.e., the extent to which the observations vary from one another from some average value. The significance* of the measure of variation can best be appreciated from the following example :

	Factory A wages (Rs.)	*Factory B wages (Rs.)*	*Factory C wages (Rs.)*
	2300	2310	2380
	2300	2300	2210
	2300	2304	2220
	2300	2306	2200
	2300	2280	2490
Total :	11,500	11,500	11,500
$\bar{X}$:	2,300	2,300	2,300

The above data pertains to five workers each in three different factories. Since the average wage is the same in all factories, one is likely to conclude that the factories are alike in their wage structure, but a close examination shall reveal that the wage distribution in the three factories differs widely from one another. In factory *A,* each and every worker is perfectly represented by the arithmetic mean, *i.e.,* average wage or, in other words, none of the workers of factory *A* deviates from the arithmetic mean and hence there is no variation. In factory *B*, only one worker is perfectly represented by the arithmetic mean, the other workers vary from the mean but the variation is very small as compared to the workers of factory *C*. In factory *C*, the mean does not represent the workers as the individual wage figures differ widely from the mean. Thus we find there is no variation in the wages of workers in factory *A*, there is very little variation in factory *B* but the wages of workers of factory *C* differ most widely. For the student of social sciences, the mean wage is not so important as to know how these wages are distributed. Are there a large number receiving the mean wage or are there a few with enormous wages and millions with wages far below the mean? The following three diagrams (given on next page) represent frequency distribution with some of the characteristics we wish to emphasise :

The two curves in diagram (*a*) represent two distributions with the same mean $\bar{X}$, but with different variations. The two curves in (*b*) represent two distributions with the same variations but with unequal means, $\bar{X}_1$, and $\bar{X}_2$, Finally, (*c*) represents two distributions with unequal means and unequal variations.



The measures of central tendency are, therefore, insufficient. They must be supported and supplemented with other measures. In this chapter, we shall be especially concerned with the measures of variation (or spread, or dispersion). A measure of variation is designed to state the extent to which the individual measures* differ on an average from the mean. In measuring variation, we shall be interested in the amount of the variation or its *degree* but not in the *direction.*** For example, a measure of 6 centimetres below the mean has just as much variation as a measure of 6 centimetres above the mean.

Significance of Measuring Variation

Measures of variation are needed for four basic purposes :

(*i*) To determine the reliability of an average;

(*ii*) To serve as a basis for the control of the variability;

(*iii*) To compare two or more series with regard to their variability; and

(*iv*) To facilitate the use of other statistical measures.

A brief explanation of these points is given below :

(*i*) Measures of variation point out as to how far an average is representative of the entire data. When variation is small, the average is a typical value in the sense that it closely represents the individual value and it is reliable in the sense that it is good estimate of the average in the corresponding universe. On the other hand, when variation is large, the average is not so typical, and unless the sample is very large, the average may be quite unreliable.

(*ii*) Another purpose of measuring variation is to determine nature and cause of variation in order to control the variation itself. In matters of health, variation in body temperature, pulse beat and blood

*Generally from the mean, infrequently from other measures of central tendency.

**The question of the *direction* of the variation will be discussed later in the text.

pressure are the basic guides to diagnosis. Prescribed treatment is designed to control their variation. In industrial production, efficient operation requires control of quality variation, the causes of which are sought through inspection and quality control programmes. Thus measurement of variation is basic to the control of cause of variation. In engineering problems, measures of variation are often specially important. In social sciences, a special problem requiring the measurement of variability is the measurement of "inequality" of the distribution of income and wealth, etc.

(*iii*) Measures of variation enable comparison to be made of two or more series with regard to their variability. The study of variation may also be looked upon as a means of determining uniformity or consistency. A high degree of variation would mean little uniformity or consistency whereas a low degree of variation would mean greater uniformity or consistency.

(*iv*) Many powerful analytical tools in statistics such as correlation analysis, the testing of hypothesis, the analysis of fluctuations, techniques of production control, cost control, etc., are based on measures of variation of one kind or another.

Properties of a Good Measure of Variation

A good measure of variation should possess, as far as possible, the following properties :*

(*i*) It should be simple to understand.

(*ii*) It should be easy to compute.

(*iii*) It should be rigidly defined.

(*iv*) It should be based on each and every observation of the distribution.

(*v*) It should be amenable to further algebraic treatment.

(*vi*) It should have sampling stability.

(*vii*) It should not be unduly affected by extreme observations.

Methods of Studying Variation

The following are the important methods of studying variation :

I. The Range,

II. The Interquartile Range or Quartile Deviation,

III. The Average Deviation,

IV. The Standard Deviation, and

V. The Lorenz Curve.

Of these, the first four are mathematical methods and the last is a graphical one.

Absolute and Relative Measures of Variation

Measures of variation may be either absolute or relative. Absolute measures of variation are expressed in the same statistical unit in which the original data are given such as rupees, kilograms, tonnes, etc. These values may be used to compare the variation in two or more than two distributions provided the variables are expressed in the same units and have almost the same average value. In case the two sets of data are expressed in different units, such as quintals of sugar *versus* tonnes of sugarcane, or if the average value is very much different, such as manager's salary *versus* worker's salary, the absolute measures of variation are not comparable. In such cases measures of relative variation should be used.

*These properties are the same as those of a good mesure of central tendency. For details refer to the previous chapter.

A measure of relative variation is the ratio of a measure of absolute variation to an average. It is sometimes called a coefficient of variation, because "coefficient" means a pure number that is independent of the unit of measurement. It should be remembered that while computing the relative variation the average used as base should be the same one from which the absolute deviations were measured. This means that the arithmetic mean should be used with the standard deviation and either the arithmetic mean or median with the average deviation.

I. RANGE

Range is the simplest method of studying variation. It is defined as the difference between the value of the smallest observation and the value of the largest observation included in the distribution. Symbolically,

$$\text{Range} = L - S$$

L = Largest value, and

S = Smallest value

The relative measure corresponding to range, called the coefficient of range, is obtained by applying the following formula :

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

Illustration 1. The following are the prices of shares of a company from Monday to Saturday :

Day	*Price (Rs.)*	*Day*	*Price (Rs.)*
Monday	200	Thursday	160
Tuesday	210	Friday	220
Wednesday	208	Saturday	250

Calculate range and coefficient of range.

Solution.

$$\text{Range} = L - S$$

$L = 250$ and $S = 160$

$$\text{Range} = 250 - 160 = \text{Rs. } 90$$

$$\text{Coefficient of Range} = \frac{L - S}{L + S} = \frac{250 - 160}{250 + 160} = \frac{90}{410} = 0.219$$

In a frequency distribution, range is calculated by taking the difference between the lower limit of the lowest class and the upper limit of the highest class.

Illustration 2. Calculate coefficient of range from the following data :

Profits (Rs. lakhs)	*No. of Cos.*	*Profits (Rs. lakhs)*	*No. of Cos.*
10–20	8	40–50	8
20–30	10	50–60	4
30–40	12		

Solution. Coefficient of Range $= \frac{L - S}{L + S} = \frac{60 - 10}{60 + 10} = \frac{50}{70} = 0.714.$

Merits and Limitations of Range

Merits. Among all the methods of studying variation, range is the simplest to understand and the easiest to compute. It takes minimum time to calculate the value of range. Hence, if one is interested in getting a quick rather than a very accurate picture of variability, one may compute range.

Limitations. (*i*) Range is not based on each and every observation of the distribution. (*ii*) It is subject to fluctuations of considerable magnitude from sample to sample. (*iii*) Range cannot be computed in case of open-end distributions. (*iv*) Range cannot tell us anything about the character of the distribution within two extreme observations. For example, observe the following three series:

Series A :	6,	46,	46,	46,	46,	46,	46,	46
Series B :	6,	6,	6,	6,	46,	46,	46,	46
Series C :	6,	10,	15,	25,	30,	32,	40,	46

In all the three series range is the same (*i.e.*, $46 - 6 = 40$), but it does not mean that the distributions are alike. The range takes no account on the form of the distribution within the range. Range is, therefore, most unreliable as a guide to the variation of the values within a distribution.

Uses of Range

Despite serious limitations range is useful in the following cases :

(*i*) *Quality control.* The object of quality control is to keep a check on the quality of the product without 100% inspection. When statistical methods of quality control are used, control charts are prepared and in preparing these charts *range* plays a very important role. The idea basically is that if the range—the differences 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 of production have not followed their usual more consistent pattern.

(*ii*) *Fluctuation in the share prices.* Range is useful in studying the variations in the prices of stocks and shares and other commodities etc. They are very sensitive to price changes from one period to another.For example, by computing *range* we can get an idea about the range of variation of, say, gold prices. If the minimum price for 10 gm. during 2009–10 was Rs. 14,500 and the maximum price Rs. 16,700 this at once tells us about the range of variation, *i.e.*, $16,700 - 14,500 = 2,200$.

(*iii*) *Weather forecasts.* The meteorological department does make use of the *range* in determining the difference between the minimum temperature and maximum temperature. This information is of great concern to the general public because they know as to within what limits the temperature is likely to vary on a particular day.

II. THE INTERQUARTILE RANGE OR QUARTILE DEVIATION

The range as a measure of variation has certain limitations. It is based on two extreme observations and it falls to take account of the scatter within the range. From this there is reason to believe that if the variation of the extreme observations is discarded the limited range thus, established might be more instructive. For this purpose there has been developed a measure called the *interquartile range,* the range which includes the middle 50 per cent of the observations. That is, one quartile of the observations at the lower end and another quartile of the observations at the upper end of the distribution are excluded in computing the inter-quartile range. In other words, interquartile range represents the difference between the third quartile and the first quartile. Symbolically,

$$\text{Interquartile range} = Q_3 - Q_1$$

Very often the interquartile range is reduced to the form of the semi-interquartile range or quartile deviation by dividing it by 2. Symbolically,

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

where $Q.D.$ = Quartile deviation.

Quartile deviation gives the average amount by which the two quartiles differ from the median. In asymmetrical distribution, the two quartiles (Q_1 and Q_3) are equidistant from the median, *i.e.*, Med.–Q_1 = Q_3–Med. and as such the difference can be taken as a measure of variation. The median ±$Q.D.$ covers exactly 50 per cent of the observations.

In reality, however, one seldom finds a series in business and economic data that is perfectly symmetrical. Nearly all distributions of social series are not equidistant from the median. As a result an asymmetrical distribution includes only approximately 50 per cent of the observations.

When quartile deviation is very small it describes high uniformity or small variation of the central 50% observations, and a high quartile deviation means that the variation among the central observations is large.

Quartile deviation is an absolute measure of variation. The relative measure corresponding to this measure, called the coefficient of quartile deviation, is calculated as follows :

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

Coefficient of quartile deviation can be used to compare the degree of variation in different distributions.

Computation of Quartile Deviation

The process of computing quartile deviation is very simple since we have just to compute the values of the upper and lower quartiles. The following illustration would clarify the procedure.

Illustration 3. You are given the frequency distribution of 292 workers of a factory according to their average weekly wages. Calculate quartile deviation and its coefficient from the following data :

Weekly wages (Rs.)	*No. of workers*	*Weekly wages (Rs.)*	*No. of workers*
Below 1350	8	1450–1470	22
1350–1370	16	1470–1490	15
1370–1390	39	1490–1510	15
1390–1410	58	1510–1530	9
1410–1430	60	1530 & above	10
1430–1450	40		

Solution. CALCULATION OF QUARTILE DEVIATION

Weekly wages (Rs.)	*No. of workers* f	*c.f.*
Below 1350	8	8
1350–1370	16	24
1370–1390	39	63
1390–1410	58	121
1410–1430	60	181
1430–1450	40	221
1450–1470	22	243
1470–1490	15	258
1490–1510	15	273
1510–1530	9	282
1530 & above	10	292
	$N = 292$	

$$\text{Med.} = \text{Size of } \frac{N}{2} \text{ th observation} = \frac{292}{2} = 146\text{th observation}$$

Median lies in the class 1410 – 1430.

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i = 1410 + \frac{146 - 121}{60} \times 20 = 1410 + 8.333 = 1418.333$$

Q_1 = Size of $\frac{N}{4}$ th observation = $\frac{292}{4}$ = 73rd observation

Q_1 lies in the class 1390 –1410.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i = 1390 + \frac{73 - 63}{58} \times 20 = 1390 + 3.448 = 1393.448$$

Q_3 = Size of $\frac{3N}{4}$ th observation = $\frac{3 \times 292}{4}$ = 219 observation. Q_3 lies in the class 1430 – 1450.

$$Q_3 = L + \frac{3N/4 - p.c.f.}{f} \times i = 1430 + \frac{219 - 181}{40} \times 20 = 1430 + 19 = 1449$$

$$\text{Coeff. of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{1449 - 1393.448}{1449 + 1393.448} = \frac{55.552}{2842.448} = 0.020.$$

Merits and Limitations of Quartile Deviation

Merits. (*i*) In certain respects it is superior to range as a measure of variation. (*ii*) It has a special utility in measuring variation in case of open-end distributions or one in which the data may be ranked but measured quantitatively. (*iii*) It is also useful in erratic or highly skewed distributions, where the other measures of variation would be warped by extreme value. The quartile deviation is not affected by the presence of extreme values.*

Limitations. (*i*) Quartile deviation ignores 50% items, *i.e.*, the first 25% and the last 25%. As the value of quartile deviation does not depend upon every observation it cannot be regarded as a good method of measuring variation. (*ii*) It is not capable of mathematical manipulation. (*iii*) Its value is very much affected by sampling fluctuations. (*iv*) It is in fact not a measure of variation as it really does not show the scatter around an average but rather a distance on a scale, *i.e.*, quartile deviation is not itself measured from an average, but it is a positional average.

Consequently, some statisticians speak of quartile deviation as measure of *partition* rather than as a measure of variation. If we really desire to measure variation in the sense of showing the scatter around an average, we must include the deviation of each and every observation from an average in the measurement.

Because of the above limitations, quartile deviation is rarely used in practice.

III. THE AVERAGE DEVIATION

The two methods of variation discussed above, namely, range and quartile deviation, are not measures of variation in the strict sense of the term because they do not show the scatterness around an average. However, to study the formation of a distribution we should take the deviations from an average. The two other measures, namely, the average deviation and the standard deviation, help us in achieving this goal.

Average deviation is obtained by calculating the absolute deviations of each observation from median (or mean), and then averaging these deviations by taking their arithmetic mean. The formula for average deviation may be written as :

$$\text{A.D.}_{(\text{Med.})} = \frac{\Sigma |X - \text{Med.}|}{N}$$

In case deviations are taken from mean the formula shall be written as :

$$\text{A.D.}_{(\bar{X})} = \frac{\Sigma |X - \bar{X}|}{N}$$

*The range and *Q.D.* are positional measures of variation as they are based on the position of certain items in a distribution.

The reason for taking absolute deviations, that is, deviations in which signs are ignored, is that it is the amount of the differences of observations from median rather than the direction of the differences which is of main interest.

While calculating average deviation, deviation of observations can be taken from any average. However, theoretically speaking, there is an advantage in taking the deviations from median because the sum of the deviations of observations from median is minimum when signs are ignored. In actual practice the arithmetic mean is more popularly used in calculating the value of mean deviation because of its wide usage as a measure of central tendency. In any case, the average used must be clearly stated in a given problem so that any possible confusion in meaning is avoided.

Computation of Average Deviation—Ungrouped Data

The formula for computing average deviation is

$$\text{A.D.}_{(\text{Med.})} = \frac{\Sigma |X - \text{Med.}|}{N}$$

If the distribution is symmetrical the average (mean or median) $\pm$ average deviation is the range that will include 57.5 per cent of the observations in the series. If it is moderately skewed, then we may expect approximately 57.5 per cent of the observations to fall within this range. Hence if average deviation is small, the distribution is highly compact or uniform, since more than half of the cases are concentrated within a small range around the mean.

The relative measure corresponding to the average deviation, called the coefficient of average deviation, is obtained, by dividing average deviation by the particular average used in computing average deviation. Thus, if average deviation has been computed from median, the coefficient of average deviation shall be obtained by dividing average deviation by the median.

$$\text{Coefficient of A.D.}_{(\text{Med.})} = \frac{\text{A.D.}}{\text{Median}}$$

If mean has been used while calculating the value of average deviation, in such a case coefficient of average deviation shall be obtained by dividing average deviation by the mean.

Illustration 4. Calculate the average deviation and coefficient of average deviation of the two income groups of five and seven workers working in two different branches of a firm :

Branch I *Income (Rs.)*	*Branch II* *Income (Rs.)*
4,000	3,000
4,200	4,000
4,400	4,200
4,600	4,400
4,800	4,600
	4,800
	5,800

Solution. CALCULATION OF AVERAGE DEVIATION

Branch I		*Branch II*	
Income (Rs.)	$\lvert X - \text{Med.}\rvert$ *Med. = 4,400*	*Income (Rs.)*	$\lvert X - \text{Med.}\rvert$ *Med. = 4,400*
4,000	400	3,000	1,400
4,200	200	4,000	400
4,400	0	4,200	200
4,600	200	4,400	0
4,800	400	4,600	200
		4,800	400
		5,800	1,400
$N = 5$	$\Sigma \lvert X - \text{Med.}\rvert = 1{,}200$	$N = 7$	$\Sigma \lvert X - \text{Med.}\rvert = 4000$

Branch I : $$\text{A.D.} = \frac{\Sigma|X - \text{Med.}|}{N} = \frac{1200}{5} = 240.$$

$$\text{Coeff. of A.D.} = \frac{\text{A.D.}}{\text{Median}} = \frac{240}{4{,}400} = 0.054$$

Branch II : $$\text{A.D.} = \frac{\Sigma|X - \text{Med.}|}{N} = \frac{4{,}000}{7} = 571.43$$

$$\text{Coeff. of A.D.} = \frac{571.43}{4{,}400} = 0.13.$$

Calculation of Average Deviation—Grouped Data

In case of grouped data, the formula for calculating average deviation is :

$$\text{A.D.}_{(\text{Med.})} = \frac{\Sigma f\,|X - \text{Med.}|}{N}$$

Illustration 5. Calculate average deviation from mean from the following data :

Sales (in thousand Rs.)	*No. of days*	*Sales (in thousand Rs.)*	*No. of days*
10–20	3	40–50	3
20–30	6	50–60	2
30–40	11		

Solution. CALCULATION OF AVERAGE DEVIATION

Sales (in thousand Rs.)	*m.p. X*	*f*	*(X – 35)/10*	*fd*	$\lvert X - \bar{X} \rvert$	$f\lvert X - \bar{X} \rvert$
10–20	15	3	–2	–6	18	54
20–30	25	6	–1	–6	8	48
30–40	35	11	0	0	2	22
40–50	45	3	+1	+3	12	36
50–60	55	2	+2	+4	22	44
		$N = 25$		$\Sigma fd = -5$		$\Sigma f\lvert X - \bar{X} \rvert = 204$

$$\text{A.D.} = \frac{\Sigma f\,|X - \bar{X}|}{N}$$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 35 - \frac{5}{25} \times 10 = 35 - 2 = 33$$

$$\text{A.D.} = \frac{204}{25} = 8.16$$

Thus the average sales are Rs. 33 thousand per day and the average deviation of sales is Rs. 8.16 thousand.

Merits and Limitations of Average Deviation

Merits. (*i*) The outstanding advantage of the average deviation is its relative simplicity. It is simple to understand and easy to compute. Any one familiar with the concept of the average can readily appreciate the meaning of the average deviation. If a situation requires a measure of variation that will be presented to the general public or any group not thoroughly grounded in statistics, the average deviation is very useful. (*ii*) It is based on each and every observation of the data. Consequently change in the value of any observation would change the value of average deviation. (*iii*) Average deviation is less affected by the values of extreme observation. (*iv*) Since deviations are taken from a central value, comparison about formation of different distributions can easily be made.

Limitations. (*i*) The greatest drawback of this method is that algebraic signs are ignored while taking the deviations of the items. If the signs of the deviations are not ignored, the net sum of the deviations will be zero if the reference point is the mean, or approximately zero if the reference point is

median. (*ii*) This method may not give us very accurate results. The reason is that average deviation gives us best results when deviations are taken from median. But median is not a satisfactory measure when the degree of variability in a series is very high. And if we compute average deviation from mean that is also not desirable because the sum of the deviations from mean (ignoring signs) is greater than the sum of the deviations from median (ignoring signs). If average deviation is computed from mode that also does not solve the problem because the value of mode cannot always be determined. (*iii*) It is not capable of further algebraic treatment. (*iv*) It is rarely used in sociological and business studies.

Because of these limitations its use is limited and it is overshadowed as a measure of variation by the superior standard deviation.

Usefulness of the Average Deviation. The serious drawbacks of the mean deviation should not blind us to its practical utility. Because of its simplicity in meaning and computation, it is especially effective in reports presented to the general public or to groups not familiar with statistical methods. This measure is useful for small samples with no elaborate analysis required. Incidentally it may be mentioned that the National Bureau of Economic Research has found in its work on forecasting business cycles, that the average deviation in the most practical measure of variation to use for this purpose.

IV. THE STANDARD DEVIATION

The standard deviation concept was introduced by Karl Pearson in 1893. It is by far the most important and widely used measure of studying variation. Its significance lies in the fact that it is free from those defects from which the earlier methods suffer and satisfies most of the properties of a good measure of variation. It is a measure of how much "spread" or "variability" is present in the sample. If all the numbers in the sample are very close to each other, the standard deviation is close to zero.If the numbers are well dispersed, the standard deviation will tend to be large. Standard deviation is also known as *root mean square deviation* for the reason that it is the square root of the means of square deviations from the arithmetic mean. Standard deviation is denoted by the small Greek letter σ (read as sigma) and is defined as :

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

If we square standard deviation, we get what is called Variance.

Hence Variance $= \sigma^2$ or $\sigma = \sqrt{\text{Variance}}$

The standard deviation measures the absolute variation of a distribution ; the greater the amount of variation, the greater the standard deviation, for the greater will be the magnitude of the deviation of the values from their mean. A small standard deviation means a high degree of uniformity of the observations as well as homogeneity of a series, a large standard deviation means just the opposite. Thus if we have two or more comparable series with identical or nearly identical means, it is the distribution with the smallest standard deviation that has the most representative mean. Hence standard deviation is extremely useful in judging the representativeness of the mean.

Calculation of Standard Deviation – Ungrouped Data

Standard deviation may be computed by applying any of the following two methods :

1. By taking deviations from the actual mean ; and
2. By taking deviations from an assumed mean.

1. Deviations taken from Actual Mean. When deviations are taken from the actual mean, the following formula is applied :

$$\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}} \quad ...(i)$$

If we calculate standard deviation without taking deviations, the above formula (*i*) after simplification (opening the brackets) can be used and is given by

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2} \quad \text{or} \quad \sigma = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2}$$

2. Deviations taken from Assumed Mean. When the actual mean is in fractions, say 87.297, it would be too cumbersome to take deviations from it and then find squares of these deviations. In such a case either the mean may be approximated or else the deviations be taken from an assumed mean and the necessary adjustment be made in the value of standard deviation. The former method of approximation is less accurate and therefore, invariably in such a case deviations are taken from assumed mean.

When deviations are taken from assumed mean the following formula* is applied :

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} \qquad ...(ii)$$

where $d = (X - A)$

Illustration 6. Find the standard deviation from the weekly wages of ten workers working in a factory :

Workers	*Weekly wages (Rs.)*	*Workers*	*Weekly wages (Rs.)*
A	1320	*F*	1340
B	1310	*G*	1325
C	1315	*H*	1321
D	1322	*I*	1320
E	1326	*J*	1331

Solution. CALCULATIONS OF STANDARD DEVIATION

Workers	*Weekly wages (Rs.)*	$(X - \bar{X})$	$(X - \bar{X})^2$
A	1320	–3	9
B	1310	–13	169
C	1315	–8	64
D	1322	–1	1
E	1326	+3	9

Contd.

*This formula has been derived from the original formula (*i*).

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

Since $d = X - A, \therefore X = A + d$ and $\bar{X} = A + \bar{d}$

Subtracting $\bar{X}$ from X, we get

$$(X - \bar{X}) = (d - \bar{d})$$

Substituting the value of $(X - \bar{X})$ in (*i*), we have

$$\sigma = \sqrt{\frac{\Sigma (d - \bar{d})^2}{N}} = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$$

F	1340	+17	289
G	1325	+2	4
H	1321	–2	4
I	1320	–3	9
J	1331	+8	64
$N = 10$	$\Sigma X = 13230$	$\Sigma (X - \bar{X}) = 0$	$\Sigma (X - \bar{X})^2 = 622$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{13230}{10} = \text{Rs. } 1323$$

$$\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}} = \sqrt{\frac{622}{10}} = 7.89$$

If, in the above question, deviations are taken from 1320 instead of the actual mean 1323, the assumed mean method will be applied and the calculations would be as follows :

CALCULATION OF STANDARD DEVIATION (ASSUMED MEAN METHOD)

Workers	*Weekly wages (Rs.)* X	$(X - A)$ $A = 1320$ d	d^2
A	1320	0	0
B	1310	–10	100
C	1315	–5	25
D	1322	+2	4
E	1326	+6	36
F	1340	+20	400
G	1325	+5	25
H	1321	+1	1
I	1320	0	0
J	1331	+11	121
$N = 10$		$\Sigma d = 30$	$\Sigma d^2 = 712$

$$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2} = \sqrt{\frac{712}{10} - \left(\frac{30}{10}\right)^2} = \sqrt{71.2 - 9} = \sqrt{62.2} = 7.89$$

Thus the answer remains the same by both the methods. It should be noted that when actual mean is not a whole number, assumed mean method should be preferred because it simplifies calculations.

Calculation of Standard Deviation—Grouped Data

In grouped frequency distribution, standard deviation can be calculated by applying any of the following two methods :

1. By taking deviations from actual mean; and
2. By taking deviations from assumed mean.

1. Deviations taken from Actual Mean. When deviations are taken from actual mean, the following formula* is used :

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

If we calculate standard deviation without taking deviations, then this formula after simplification (opening the brackets) can be used and is given by

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left(\frac{\Sigma fX}{N}\right)^2} \quad \text{or} \quad \sigma = \sqrt{\frac{\Sigma fX^2}{N} - (\bar{X})^2}$$

2. Deviations taken from Assumed Mean. When deviations are taken from the assumed mean, the following formula* is applied :

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i$$

Illustration 7. An analysis of production rejects resulted in the following figures :

No. of rejects per operator	*No. of operators*	*No. of rejects per operator*	*No. of operators*
21–25	5	41–45	15
26–30	15	46–50	12
31–35	28	51–55	3
36–40	42		

Calculate mean and standard deviation.

Solution. Converting the discrete data into the continuous data, we get the following table :

CALCULATION OF MEAN AND STANDARD DEVIATION

No. of rejects per operator	*m.p. X*	*No. of operators f*	*(X – 38)/5 d*	*fd*	*fd²*
20.5–25.5	23	5	–3	–15	45
25.5–30.5	28	15	–2	–30	60
30.5–35.5	33	28	–1	–28	28
35.5–40.5	38	42	0	0	0
40.5–45.5	43	15	+1	+15	15
45.5–50.5	48	12	+2	+24	48
50.5–55.5	53	3	+3	+9	27
		$N = 120$		$\Sigma fd = -25$	$\Sigma fd^2 = 223$

* $$\sigma = \frac{\Sigma f(X-\bar{X})^2}{N} \qquad ...(i)$$

Let $d = \frac{X-A}{i}$; then $X = A + id$ and $= \bar{X} = A + i\bar{d}$

Substituting the value of $(X - \bar{X})$ in (i), we have

$$\sigma = \sqrt{\frac{\Sigma f[i(d-\bar{d})]^2}{N}} = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i$$

$$\text{Mean : } \bar{X} = A + \frac{\Sigma fd}{N} \times i = 38 - \frac{25}{120} \times 5 = 38 - 1.04 = 36.96$$

$$\text{Standard deviation : } \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{223}{120} - \left(\frac{-25}{120}\right)^2} \times 5$$

$$= \sqrt{1.858 - .043} \times 5 = 1.347 \times 5 = 6.375.$$

Illustration 8. An association doing charity work decided to give old age pensions to people over sixty years of age.

The scales of pensions were fixed as follows :

Age group 60 to 65—Rs. 2500 per month.
Age group 65 to 70—Rs. 3000 per month.
Age group 70 to 75—Rs. 3500 per month.
Age group 75 to 80—Rs. 4000 per month.
Age group 80 to 85—Rs. 4500 per month.

The age of 25 persons who secured the pension benefits are given below :

75	62	84	72	83	72	81	64	71	63	61	60	61
67	74	64	79	73	75	76	69	78	66	67	68	

Calculate the monthly average pension payable and the standard deviation.

Solution. CLASSIFYING THE ABOVE DATA

Age Group	*Tally*	*Frequency*
60–65	𝍸 ||	7
65–70	𝍸	5
70–75	𝍸 |	6
75–80	||||	4
80–85	|||	3
		$N = 25$

CALCULATIONS OF MONTHLY AVERAGE PENSION PAYABLE AND THE STANDARD DEVIATION

Pension (Rs.) X	$(X - 3500)/500$ d	f	fd	fd^2
2500	−2	7	−14	28
3000	−1	5	−5	5
3500	0	6	0	0
4000	+1	4	+4	4
4500	+2	3	+6	12
		$N = 25$	$\Sigma fd = -9$	$\Sigma fd^2 = 49$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 3500 - \frac{9}{25} \times 500 = 3500 - 180 = \text{Rs. } 3320$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{49}{25} - \left(\frac{-9}{25}\right)^2} \times 500 = 1.353 \times 500 = \text{Rs. } 676.5$$

Thus the monthly average pension is Rs. 3320 and standard deviation Rs. 676.5.

Mathematical Properties of Standard Deviation

Standard deviation has some very important mathematical properties which considerably enhance its utility in statistical work.

1. *Combined Standard Deviation.* Just as it is possible to compute combined mean of two or more than two groups, similarly we can also compute combined standard deviation of two or more groups. Combined standard deviation of two groups is denoted by σ_{12} and is computed as follows :

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

where σ_{12} = combined standard deviation,

σ_1 = standard deviation of first group,

σ_2 = standard deviation of second group, and

$d_1 = |\bar{X}_1 - \bar{X}_{12}|$; $d_2 = |\bar{X}_2 - \bar{X}_{12}|$.

The above formula can be extended to find out the standard deviation of three or more groups. For example, combined standard deviation of three groups would be :

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

where $d_1 = |\bar{X}_1 - \bar{X}_{123}|$; $d_2 = |\bar{X}_2 - \bar{X}_{123}|$; $d_3 = |\bar{X}_3 - \bar{X}_{123}|$.

Illustration 9. The number of workers employed, the mean wage (in Rs.) per week and the standard deviation (in Rs.) in each branch of a company are given below. Calculate mean wages and standard deviation of all the workers taken together for company.

Branch	*No. of workers employed*	*Weekly mean wage (in Rs.)*	*Standard deviation (in Rs.)*
A	50	1413	60
B	60	1420	70
C	90	1415	80

Solution :

$$\bar{X}_{123} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3}{N_1 + N_2 + N_3}$$

$$= \frac{(50 \times 1413) + (60 \times 1420) + (90 \times 1415)}{50 + 60 + 90}$$

$$= \frac{70{,}650 + 85{,}200 + 1{,}27{,}350}{200} = \frac{2{,}83{,}200}{200} = \text{Rs. } 1{,}416$$

Combined standard deviation of three branches

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

$$d_1 = \left|\bar{X}_1 - \bar{X}_{123}\right| = |1413 - 1416| = 3$$

$$d_2 = \left|\bar{X}_2 - \bar{X}_{123}\right| = |1420 - 1416| = 4$$

$$d_3 = \left|\bar{X}_3 - \bar{X}_{123}\right| = |1415 - 1416| = 1$$

$$\sigma_{123} = \sqrt{\frac{50(60)^2 + 60(70)^2 + 90(80)^2 + 50(3)^2 + 60(4)^2 + 90(1)^2}{50 + 60 + 90}}$$

$$= \sqrt{\frac{1,80,000 + 2,94,000 + 5,76,000 + 450 + 960 + 90}{200}}$$

$$= \sqrt{\frac{1051500}{200}} = \text{Rs. } 72.51.$$

2. *Standard deviation of natural numbers.* The standard deviation of the first n natural numbers* can be obtained by the following formula :

$$\sigma = \sqrt{\frac{1}{12}(N^2 - 1)}$$

Thus the standard deviation of natural numbers 1 to 10 will be

$$\sigma = \sqrt{\frac{1}{12}(10^2 - 1)} = \sqrt{\frac{1}{12} \times 99} = \sqrt{8.25} = 2.87.$$

3. The sum of the squares of the deviations of all the observations from their arithmetic mean is minimum. In other words, the sum of the squares of the deviations of observations from a value other than the arithmetic mean would always be greater. This is the reason why standard deviation is always computed from the arithmetic mean.

4. Standard deviation is independent of change of origin but not scale.

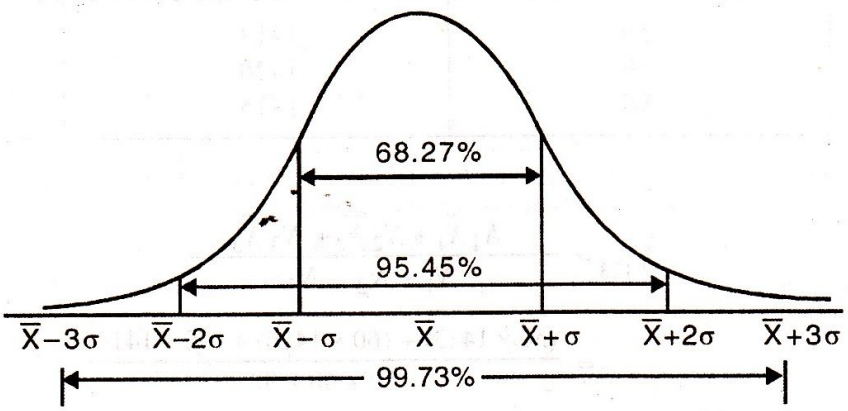
5. For a symmetrical distribution, the following area relationships hold good.

Mean $\pm$ 1 σ covers 68.27% observations.

Mean $\pm$ 2 σ covers 95.45% observations.

Mean $\pm$ 3 σ covers 99.73% observations.

This can be illustrated by the following diagram :



Relation between Measures of Variation

In a bell-shaped symmetrical distribution, there is a fixed relationship between the three most commonly used measures of variation. The quartile deviation is smallest, the mean deviation next and the standard deviation is largest in the following proportions :

$$Q.D. = \frac{2}{3}\sigma \text{ ; and } A.D. = \frac{4}{5}\sigma$$

These relationships can be easily memorized because of the sequence 2, 3, 4, 5. The same proportions tend to hold true for many distributions that are quite symmetrical. They are useful in estimating one measure of variation when another is known, or in checking roughly the accuracy of a calculated value. If the computed σ differs very widely from its value estimated from $Q.D.$ or $A.D.$ either an error has been made or the distribution differs considerably from symmetry.

*By natural number we mean only positive integers, *i.e.*, 1, 2, 3, 4, 5 ...

Another comparison may be made of the proportion of observations that are 'typically included within the range of one *Q.D.*, *A.D.* or *S.D.* measured both above and below the mean. In a normal distribution :

$\bar{X} \pm Q.D.$ includes 50 per cent of the observations.

$\bar{X} \pm A.D.$ includes 57.51 per cent of the observations.

$\bar{X} \pm \sigma$ includes 68.27 per cent or about two-thirds of the observations.

Illustration 10. The breaking strength of 80 'test pieces' of a certain alloy is given in the following table, the unit being given to the nearest thousand pounds per square inch.

Breaking strength	*No. of pieces*
44–46	3
46–48	24
48–50	27
50–52	21
52–54	5

Calculate the average breaking strength of the alloy and the standard deviation. Calculate the percentage of observations lying between mean ± 2σ. (*MBA, Kurukshetra Univ; MBA, Madras Univ., 2006*)

Solution : CALCULATION OF MEAN AND STANDARD DEVIATION

Breaking strength	*m.p.* X	f	$(X-49)/2$ d	fd	fd^2
44–46	45	3	−2	−6	12
46–48	47	24	−1	−24	24
48–50	49	27	0	0	0
50–52	51	21	+1	+21	21
52–54	53	5	+2	+10	20
		$N = 80$		$\Sigma fd = 1$	$\Sigma fd^2 = 77$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 49 + \frac{1}{81} \times 2 = 49.025$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{77}{80} - \left(\frac{1}{80}\right)^2} \times 2$$

$$= \sqrt{0.962 - 0.0001} \times 2 = 0.98 \times 2 = 1.96$$

Hence $\bar{X} = 49.025, \sigma = 1.96$

The value of $\bar{X} \pm 2\sigma = 49.025 \pm 2\,(1.96) = 49.025 \pm 3.92$

or $= 45.105, 52.945 = 45$ and 53 approx.

We have to calculate the percentage of items lying between 45 and 53. For this we make an assumption that the number of pieces are equally distributed within each class. Since between 44 and 46 there are 3 frequencies at 45, these would be 1.5. Similarly at 53 the frequency would be 2.5. Thus the total frequency between 45 and 53 = (1.5 + 24 + 27 + 21 + 2.5) = 76. The percentage is $\frac{76}{80} \times 100 = 95$. Thus there are 95 per cent observations lying within the limits mean ±2σ.

*For details please refer to Chapter on 'Probability Distribution'.

Merits and Limitations of Standard Deviation

Merits. (*i*) The standard deviation is the best measure of variation because of its mathematical characteristics. It is based on every item of the distribution. Also it is amenable to algebraic treatment and is less affected by fluctuations of sampling than most other measures of variation.

(*ii*) It is possible to calculate the combined standard deviation of two or more groups. This is not possible with any other measure.

(*iii*) For comparing the variability of two or more distributions coefficient of variation is considered to be most appropriate and this measure is based on mean and standard deviation.

(*iv*) Standard deviation is most prominently used in further statistical work. For example, in comparing skewness, correlation, etc., use is made of standard deviation. It is a key-note in sampling and provides a unit of measurement for the normal distribution.

Limitations. (*i*) As compared to other measures it is difficult to compute. However, it does not reduce the importance of this measure because of the high degree of accuracy of result it gives.

(*ii*) It gives more weight to extreme values and less to those which are near the mean. It is because of the fact that the squares of the deviations which are big in size would be proportionately greater than the squares of those deviations which are comparatively small. The deviations 2 and 8 are in ratio of 1 : 4 but their squares, *i.e.*, 4 and 64, would be in the ratio 1 : 16.

Correcting Incorrect Value of Standard Deviation

Mistakes in calculations are always possible. Sometimes it so happens that while calculating mean and standard deviation we unconsciously copy out wrong observations. For example, an observation 21 may be poised as 12. Similarly, one observation 127 may be taken as only 27. In such a case if the entire calculations are done again, it would become too difficult a task. By adopting a very simple procedure we can correct the incorrect values of mean and standard deviation. For obtaining correct mean we find out correct ΣX by deducting from the original ΣX the wrong observations and adding to it the correct observations. Similarly, for calculating correct standard deviation we obtain the value of correct ΣX^2. The following illustration shall clarify the procedure :

Illustration 11. The mean and standard deviation of a set of 100 observations were worked out as 40 and 5 respectively by a computer which by mistake took the value 50 in place of 40 for one of the observations. Find the correct mean and variance.

(M. Com., Jammu Univ. ; MBA, Lucknow Univ., 2002)

Solution : $\bar{X} = \dfrac{\Sigma X}{N}$, $N\bar{X} = \Sigma X, N = 100, \bar{X} = 40$

$\therefore \quad \Sigma X = 100 \times 40 = 4{,}000$

But this is not the correct ΣX because one observation has been taken as 50 instead of 40.

$\therefore \quad$ Correct $\Sigma X = 4{,}000 - 50 + 40 = 3{,}990$

$$\text{Correct Mean} = \frac{3{,}990}{100} = 39.9$$

$$\text{Variance} = \frac{\Sigma X^2}{N} - (\bar{X})^2$$

$$\text{Variance} = \sigma^2 = (5)^2 = 25, N = 100.$$

$$25 = \frac{\Sigma X^2}{100} - (40)^2 \quad \text{or} \quad 2{,}500 = \Sigma X^2 - 1{,}60{,}000$$

$$\Sigma X^2 = 1{,}60{,}000 + 2{,}500 = 1{,}62{,}500$$

$$\text{Correct } \Sigma X^2 = 1{,}62{,}500 - (50)^2 + (40)^2 = 1{,}62{,}500 - 2{,}500 + 1{,}600 = 1{,}61{,}600$$

$$\text{Correct variance} = \frac{\text{Correct } \Sigma X^2}{N} - (\textit{Correct } \bar{X})^2$$

$$= \frac{1{,}61{,}600}{100} - (39.9)^2 = 1{,}616 - 1{,}592.01 = 23.99$$

Thus correct mean = 39.9 and the correct variance = 23.99.

Coefficient of Variation

The standard deviation discussed so far is an absolute measure of variation. The corresponding relative measure is known as the *coefficient of variation.* This measure developed by Karl Pearson is the most commonly used measure of relative variation. It is used in such problems where we want to compare the variability of two or more than two series. That series (or group) for which the coefficient of variation is greater is said to be more variable or conversely less consistent, less uniform, less stable or less homogeneous. On the other hand, the series for which coefficient of variation is less is said to be less variable or more consistent, more uniform, more stable or more homogeneous. Coefficient of variation denoted by C.V. is obtained as follows :

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100$$

It may be pointed out that although any measure of variation can be used in conjunction with any average in computing relative variation, statisticians, in fact, almost always use the standard deviation as the measure of variation and the arithmetic mean as the average. Coefficient of variation is more useful when the two distributions are entirely different and the units of measurement are also different. When the relative variation is stated in terms of the arithmetic mean and the standard deviation, the resulting percentage is known as the coefficient of variation or coefficient of variability.

Illustration 12. Suppose that samples of polythene bags from two manufacturers *A* and *B* are tested by a prospective buyer for bursting pressure, with the following results :

Bursting pressure (*lbs*)	*Number of bags*	
	A	*B*
5.0–9.9	2	9
10.0–14.9	9	11
15.0–19.9	29	18
20.0–24.9	54	32
25.0–29.9	11	27
30.0–34.9	5	13
	110	110

Which set of bags has the highest average bursting pressure? Which has more uniform pressure? If prices are the same, which manufacturer's bags would be preferred by the buyer? Why?

Solution. For determining which set of bags has the highest average bursting pressure, calculate arithmetic mean and for finding out which has more uniform pressure compute coefficient of variations.

Manufacturer A :

CALCULATION OF MEAN AND STANDARD DEVIATION

Bursting pressure (*lbs*)	*m.p.* *X*	*f*	$(X-17.45)/5$ *d*	*fd*	fd^2
4.95–9.95	7.45	2	−2	−4	8
9.95–14.95	12.45	9	−1	−9	9
14.95–19.95	**17.45**	29	0	0	0
19.95–24.95	22.45	54	+1	+54	54
24.95–29.95	27.45	11	+2	+22	44
29.95–34.95	32.45	5	+3	+15	45
		$N = 110$		$\Sigma fd = 78$	$\Sigma fd^2 = 160$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 17.45 + \frac{78}{110} \times 5 = 17.45 + 3.55 = 21$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{160}{110} - \left(\frac{78}{110}\right)^2} \times 5$$

$$= \sqrt{1.455 - 0.503} \times 5 = \sqrt{0.952} \times 5 = 0.976 \times 5 = 4.879$$

$$C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{4.879}{21} \times 100 = 23.23\%.$$

Manufacturer B :

CALCULATION OF MEAN AND STANDARD DEVIATION

Bursting pressure (lbs.)	*m.p.* X	f	$(X-17.45)/5$ d	fd	fd^2
4.95–9.95	7.45	9	–2	–18	36
9.95–14.95	12.45	11	–1	–11	11
14.95–19.95	17.45	18	0	0	0
19.95–24.95	22.45	32	+1	+32	32
24.95–29.95	27.45	27	+2	+54	108
29.95–34.95	32.45	13	+3	+39	117
		$N = 110$		$\Sigma fd = 96$	$\Sigma fd^2 = 304$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 17.45 + \frac{96}{110} \times 5 = 21.81$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{304}{110} - \left(\frac{96}{110}\right)^2} \times 5$$

$$= \sqrt{2.764 - 0.762} \times 5 = 1.4149 \times 5 = 7.0745$$

$$C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{7.0745}{21.81} \times 100 = 32.44\%$$

Since the average bursting pressure is higher for manufacturer *B*, hence the bags of manufacturer *B* have a higher bursting pressure. The bags of manufacturer *A* have more uniform pressure since the coefficient of variation is less for manufacturer *A*. If prices are the same, the bags of manufacturer *A* should be preferred by the buyer because they have more uniform pressure.

Illustration 13. In two factories *A* and *B* engaged in the same industry, the average monthly wages and standard deviations are as follows :

Factory	*Average Monthly Wages (Rs.)*	*S.D. of Wages (Rs.)*	*No. of Wage Earners*
A	4600	500	100
B	4900	400	80

(*i*) Which factory *A* or *B* pays larger amount as monthly wages?

(*ii*) Which factory shows greater variability in the distribution of wages?

(*iii*) What is the mean and standard deviation of all the workers in two factories taken together?

Solution. (*i*) For finding out which factory *A* or *B* pays larger amount as monthly wages; we have to compare the total wage bill.

Factory A : Total wage bill = 4600 × 100 = Rs. 4,60,000

Factory B : Total wage bill = 4900 × 80 = Rs. 3,92,000

Hence factory *A* pays larger amount as monthly wages.

(*ii*) For determining which factory shows greater variation in the distribution of wages, we have to compare coefficient of variation.

$$C.V. \text{ (factory } A) = \frac{\sigma}{\bar{X}} \times 100 = \frac{500}{4600} \times 100 = 10.87$$

$$C.V. \text{ (factory } B) = \frac{\sigma}{\bar{X}} \times 100 = \frac{400}{4900} \times 100 = 8.16$$

Since coefficient of variation is higher in factory *A*, hence factory *A* shows greater variability in the distribution of wages.

(*iii*) ***Combined mean*** :

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$= \frac{(100 \times 4600) + (80 \times 4900)}{100 + 80} = \frac{460000 + 392000}{180}$$

$$= \frac{852000}{180} = \text{Rs. } 4{,}733.33$$

Combined Standard Deviation : $\sigma_{12} = \sqrt{\dfrac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$

$$N_1 = 100, \sigma_1 = 500, N_2 = 80, \sigma_2 = 400, d_1 = |\bar{X}_2 - \bar{X}_{12}|$$

$$|4600 - 4733.33| = 133.33, d_2 = |\bar{X}_2 - \bar{X}_{12}| = |4900 - 4733.33| = 166.67$$

$$\sigma_{12} = \sqrt{\frac{100\,(500)^2 + 80\,(400)^2 + 100\,(133.33)^2 + 80\,(166.67)^2}{100 + 80}}$$

$$= \sqrt{\frac{25000000 + 1200000 + 1777688.89 + 222231.11}{180}}$$

$$= \sqrt{\frac{42192486.335}{180}} = 484.15.$$

Hence combined standard deviation is Rs. 484.15.

Illustration 14. The number of employees, daily wages per employee and the variance of the wages per employee for two factories are given below :

	Factory A	*Factory B*
Number of employees	50	100
Average daily wages per employee (Rs.)	120	85
Variance of the daily wages per employee (Rs.)	9	16

(*a*) In which factory is there greater variation in the distribution of daily wages per employee?

(*b*) Suppose in factory *B*, the daily wages of an employee were wrongly noted as Rs. 120 instead of Rs. 100. What would be the correct variance for factory *B*? ***(Diploma in Mgt., IGNOU, 2002; MBA, UP Tech, Univ, 2003)***

Solution : (*a*) *Variation in the Distribution of Wages*

Factory A

$$C.V. = \frac{\sigma}{\bar{X}} \times 100$$

$$\sigma = \sqrt{9} = 3, \bar{X} = 120$$

$$\therefore \quad C.V. = \frac{3}{120} \times 100 = 2.5$$

Factory B

$$C.V. = \frac{\sigma}{\bar{X}} \times 100$$

$$\sigma = \sqrt{16} = 4, \bar{X} = 85$$

$$\therefore C.V. = \frac{4}{85} \times 100 = 4.7$$

The coefficient of variation is greater for factory *B*, hence there is greater variation in the distribution of wages per employee in factory *B*.

(*b*) *Correct Variance* : For finding correct variance we have first to find the correct mean.

$$\bar{X} = \frac{\Sigma X}{N}, N\bar{X} = \Sigma X$$

$$\Sigma X = 100 \times 85 = 8{,}500$$

$$\text{Correct } \Sigma X = 8{,}500 - 120 + 100 = 8{,}480$$

$$\text{Correct mean} = \frac{8480}{100} = 84.8$$

$$\text{Variance or } \sigma^2 = \frac{\Sigma X^2}{N} - (\bar{X})^2$$

Substituting the values of σ^2, $\bar{X}$, etc.,

$$16 = \frac{\Sigma X^2}{100} - (85)^2 \text{ or } 1600 = \Sigma X^2 - 7,22,500$$

$$\Sigma X^2 = 7,24,100$$

But in this total 100 has been taken as 120.

$$\therefore \quad \text{Correct } \Sigma X^2 = 7,24,100 - (120)^2 + (100)^2$$

$$= 7,24,100 - 14,400 + 10,000 = 7,19,700$$

$$\text{Correct variance} = \frac{\text{Correct } \Sigma X^2}{N} - (\text{correct } \bar{X})^2$$

$$= \frac{7,19,700}{100} - (84.8)^2 = 7,197 - 7,191.04 = 5.96$$

Hence the correct variance for factory *B* is 5.96.

V. LORENZ CURVE

The Lorenz Curve, devised by Max O. Lorenz, a famous economic statistician, is a graphic method of studying variation. This curve was used by him for the first time to measure the distribution of wealth and income. Now the curve is also used to study the distribution of profits, wages, turnover, etc. However, still the most common use of this curve is in the study of the degree of inequality in the distribution of income and wealth between countries or between different periods of time. It is a cumulative percentage curve in which the percentage of items is combined with the percentage of other things as wealth, profits, turnover, etc.

While drawing the Lorenz Curve the following procedure is adopted :

(*i*) The size of items and frequencies are both cumulated and then percentages are obtained for these various cumulative values.

(*ii*) On the *X*-axis, start from 0 to 100 and take the per cent of variable.

(*iii*) On the *Y*-axis, start from 0 to 100 and take the per cent of variable.

(*iv*) Draw a diagonal line joining 0 with 100. This is known as line of equal distribution. Any point on this line shows the same per cent on *X* as on *Y*.

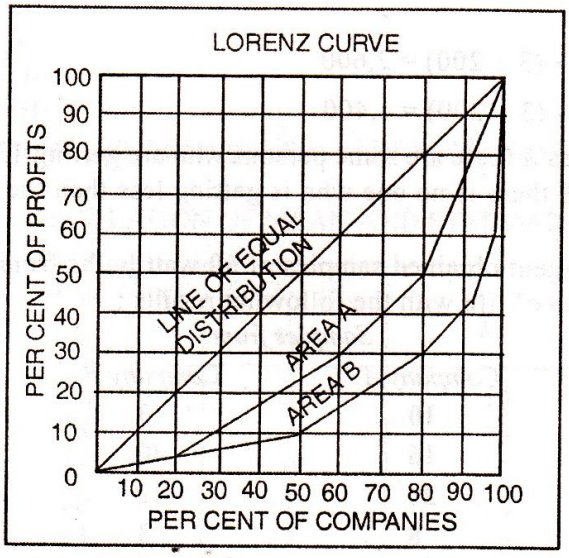
(*v*) Plot the various points corresponding to *X* and *Y* and join them. The distribution so obtained, unless it is exactly equal, will always curve below the diagonal line. If two curves of distribution are shown on the same Lorenz presentation, the curve that is farthest from the diagonal line represents the greater inequality. Clearly the line of actual distribution can never cross the line of equal distribution.

Illustration 15. In the following table is given the number of companies belonging to two areas *A* and *B* according to the amount of profits earned by them. Draw in the same diagram their Lorenz curves and interpret them.

Profits earned in Rs. '000	*No. of Companies*	
	Area A	*Area B*
6	6	2
25	11	38
60	13	52
84	14	28
105	15	38
150	17	26
170	10	12
400	14	4

Solution. CALCULATION FOR DRAWING THE LORENZ CURVE

	Profit		*Area A*			*Area B*		
Profits earned in Rs. '000	*Cumulative profits*	*Cumulative percentage*	*No. of Companies*	*Cumulative Number*	*Cumulative Percentage*	*No. of Companies*	*Cumulative Number*	*Cumulative Percentage*
6	6	0.6	6	6	6	2	2	1
25	31	3.1	11	17	17	38	40	20
60	91	9.1	13	30	30	52	92	46
84	175	17.5	14	44	44	28	120	60
105	280	28.0	15	59	59	38	158	79
150	430	43.0	17	76	76	26	184	92
170	600	60.0	10	86	86	12	196	98
400	1000	100.0	14	100	100	4	200	100



Since curve *B* is farther from the diagonal line, it represents greater inequality.

Which Measure of Variation to use ?

Unlike measures of central tendency, in case of measures of variation also the question arises which measure to use. The choice of a suitable measure depends on the following three factors :

1. *The type of data available.* If observations are few in numbers, or contain extreme values, avoid the standard deviation. If they are generally skewed, avoid the mean deviation as well. If they have gaps around the quartiles, the quartile deviation should be avoided. If there are open-end classes, the quartile measure of variation should be preferred.

2. *The purpose of investigation.* In an elementary treatment of statistical series in which a measure of variability is desired only for itself, any of the three measures, namely, range, quartile deviation and average deviation, would be acceptable. Probably the average deviation would be superior. However, in usual practice, the measure of variability is employed in further statistical analysis. For such a purpose, the standard deviation is by far the most popularly used. It is free from those defects with which other measures suffer. It lends itself to the analysis of variability in terms of normal curve of error. Practically, all advanced statistical methods deal with variability and centre around the standard deviation. Hence unless the circumstances warrant for the use of any other measure, we should make use of standard deviation for measuring variability.

MISCELLANEOUS ILLUSTRATIONS

Illustration 16. You are incharge of rationing in a State affected by food shortage. The following reports arrive from local investigators :

Daily caloric value of food available per adult during current period :

Area	*Mean*	*Standard Deviation*
A	2,500	400
B	2,000	200

The estimated requirement of an adult is taken at 2,800 calories daily and the absolute minimum is 1,350. Comment on the reported figures, and determine which area, in your opinion, need more urgent attention.

Solution. We know that $\bar{X} \pm 1\sigma$ covers 68.2% cases ; $\bar{X} \pm 2\sigma$ covers 95.45% cases and $\bar{X} \pm 3\sigma$ covers 93.73% cases. In the given problem, if we take into consideration 99.73%, *i.e.*, almost the whole of the population, the limits would be $\bar{X} \pm 3\sigma$.

For Area *A*, these limits are :

$$\bar{X} + 3\sigma = 2,500 + (3 \times 400) = 3,700$$

$$\bar{X} - 3\sigma = 2,500 - (3 \times 400) = 1,300$$

For Area *B*, these limits are :

$$\bar{X} + 3\sigma = 2,000 + (3 \times 200) = 2,600$$

$$\bar{X} - 3\sigma = 2,000 - (3 \times 200) = 1,400$$

It is clear from above that in Area *A* there are some persons who are getting 1300 calories, *i.e.*, below the minimum which is 1,350. But in case of area *B* there is no one who is getting less than the minimum. Hence area *A* needs more urgent attention.

Illustration 17. A purchasing agent obtained samples of 60 watt bulbs from two companies. He had the samples tested in his own laboratory for length of life with the following results :

Length of life (in hours)	*Samples from*	
	Company A	*Company B*
1,700 and under 1,900	10	3
1,900 " " 2,100	16	40
2,100 " " 2,300	20	12
2,300 " " 2,500	8	3
2,500 " " 2,700	6	2

(*a*) Which Company's bulbs do you think are better in terms of average life ?

(*b*) If prices of both types are the same, which company's bulbs would you buy and why ? *(MBA, DU, 2000)*

Solution. In order to answer these questions we have to calculate mean and coefficient of variation.

CALCULATION OF MEAN AND COEFFICIENT OF VARIATION

Length of life (in hours)	*m.p.* X	$(X-2200)/200$ d	*Samples from Co. A* f	fd	fd^2	*Samples from Co. B* f	fd	fd^2
1,700–1,900	1,800	–2	10	–20	40	3	–6	12
1,900–2,100	2,000	–1	16	–16	16	40	–40	40
2,100–2,300	**2,200**	0	20	0	0	12	0	0
2,300–2,500	2,400	+1	8	+8	8	3	+3	3
2,500–2,700	2,600	+2	6	+12	24	2	+4	8
			$N = 60$	$\Sigma fd = -16$	$\Sigma fd^2 = 88$	$N = 60$	$\Sigma fd = -39$	$\Sigma fd^2 = 63$

For Company A :

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 2,200 - \frac{16}{60} \times 200 = 2,146.67$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{88}{60} - \left(\frac{-16}{60}\right)^2} \times 200$$

$$= \sqrt{1.467 - 0.071} \times 200 = 1.182 \times 200 = 236.4$$

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{236.4}{2146.67} \times 100 = 11 \text{ per cent.}$$

For Company B :

$$\bar{X} = 2200 - \frac{39}{60} \times 200 = 2200 - 130 = 2070$$

$$\sigma = \sqrt{\frac{63}{60} - \left(\frac{-39}{60}\right)^2} \times 200 = \sqrt{1.05 - .42} \times 200 = .794 \times 200 = 158.8$$

$$\text{C.V.} = \frac{158.8}{2070} \times 100 = 7.67 \text{ per cent.}$$

(*a*) Since average length of life is greater in case of company *A*, hence bulbs of company *A* are better.

(*b*) Coefficient of variation is less for company *B*. Hence if prices are same, we will prefer to buy company *B*'s bulbs because their burning hours are more uniform.

Illustration 18. You are given the data pertaining to kilowatt hours of electricity consumed by 100 persons in Delhi.

Consumption (K. Watt hours)	*No. of users*
0 but less than 10	6
10 " " " 20	25
20" " " 30	36
30 " " " 40	20
40 " " " 50	13

Calculate (*i*) the standard deviation, and (*iii*) the range within which middle 50% of the consumers fall.

Solution. CALCULATION OF MEAN AND STANDARD DEVIATION

Consumption K. watt hours	*m.p. X*	*No. of Users f*	*(X–25)/10 d*	*fd*	fd^2	*c.f.*
0–10	5	6	–2	–12	24	6
10–20	15	25	–1	–25	25	31
20–30	**25**	36	0	0	0	67
30–40	35	20	+1	+20	20	87
40–50	45	13	+2	+26	52	100
		N=100		Σfd=9	Σfd^2=121	

(*i*) $$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 25 + \frac{9}{100} \times 10 = 25.9 \text{ k. watt hours}$$

(*ii*) $$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{121}{100} - \left(\frac{9}{100}\right)^2} \times 10$$

$$= \sqrt{1.21 - .008} \times 10 = 1.096 \times 10 = 10.96$$

(*iii*) For calculating the range, we have to find Q_1 and Q_3.

Q_1 = size of $\frac{N}{4}$ th observation = $\frac{100}{4}$ = 25th observation,

Q_1 lies in the class 10 – 20.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i = 10 + \frac{25-6}{25} \times 10 = 10 + 7.6 = 17.6$$

Q_3 = size of $\frac{3N}{4}$ th observation = $\frac{3 \times 100}{4}$ = 75th observation

Q_3 lies in the class 30–40.

$$Q_3 = L + \frac{3N/4 - p.c.f.}{f} \times i = 30 + \frac{75-67}{20} \times 10 = 30 + 4 = 34$$

Range within which the middle 50% of the consumers fall

$$= Q_3 - Q_1 = 34 - 17.6 = 16.4.$$

Illustration 19. The value of the arithmetic mean and standard deviation of the following frequency distribution of a continuous variable derived from the use of working origin and scale are Rs. 107 and 13.1 respectively. Determine the actual classes.

d:	–3	–2	–1	0	+1	+2
f:	1	3	4	7	3	2

(MBA, BIT, Ranchi, 2000)

Solution. In order to determine the actual classes we should know two things: (*i*) the assumed mean, *i.e.*, working origin, and (*ii*) the common factor or the class-interval.

CALCULATING ASSUMED MEAN AND COMMON FACTOR

d	f	fd	fd^2
–3	1	–3	9
–2	3	–6	12
–1	4	–4	4
0	7	0	0
+1	3	+3	3
+2	2	+4	8
	$N = 20$	$\Sigma fd = -6$	$fd^2 = 36$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i$$

$$13.1 = \sqrt{\frac{36}{20} - \left(\frac{-6}{20}\right)^2} \times i = \sqrt{1.8 - 0.09} \times i$$

$$13.1 = 1.31 \times i \quad \text{or} \quad i = \frac{13.1}{1.31} = 10$$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i \quad \text{or} \quad 107 = A - \frac{6}{20} \times 10 \quad \text{or} \quad A = 110.$$

Thus the working origin is 110 and the class-interval 10. Hence the various classes would be :

X:	75–85	85–95	95–105	105–115	115–125	125–135
f:	1	3	4	7	3	2

Illustration 20. The following data give the number of passengers travelling by Boeing 747 from one city to another in one week.

320, 290, 265, 300, 270, 200, 315

Calculate the mean and standard deviation and determine the percentage of cases that lie between (*i*) $\bar{X} \pm 1\sigma$, (*ii*) $\bar{X} \pm 2\sigma$, (*iii*) $\bar{X} \pm 3\sigma$. What percentage of cases lie outside 1σ, 2σ and 3σ limits ?

Solution. CALCULATION OF MEAN AND STANDARD DEVIATION

X	$(X - \bar{X})$	$(X - \bar{X})^2$
320	+40	1,600
290	+10	100
265	–15	225
300	+20	400
270	–10	100
200	–80	6,400
315	+35	1,225
$\Sigma X = 1960$	$\Sigma (X - \bar{X}) = 0$	$\Sigma (X - \bar{X})^2 = 10,050$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1960}{7} = 280$$

$$\sigma = \sqrt{\frac{\Sigma (X - \bar{X})^2}{N}} = \sqrt{\frac{10050}{7}} = 37.89$$

(*i*) Cases lying between :

$$\bar{X} \pm 1\sigma = 280 \pm 37.89 = 242.11 \text{ to } 317.89.$$

There are 2 observations, *i.e.*, 200 and 320, that fall outside these limits. Hence the required percentage is 28.57.

(*ii*) Cases lying between :

$$\bar{X} \pm 2\sigma = 280 \pm 2 \times 37.89 = 204.22 \text{ to } 355.78.$$

There are only one observations that lies outside this limits. Hence the required percentage is 14.28.

(*iii*) Cases lying between :

$$\bar{X} \pm 3\sigma = 280 \pm 3 \times 37.89 = 166.33 \text{ to } 393.67.$$

Not a single observation lies outside this limit.

Illustration 21. A company has three establishments E_1, E_2 and E_3 in three cities. Analysis of the daily wages paid to the employees in the three establishments is given below :

	E_1	E_2	E_3
Number of employees	20	25	40
Average daily wage (Rs.)	305	300	340
Standard deviation (Rs.)	50	40	45

Find the average and the standard deviation of the wages of all the 85 employees in the company.

Solution.

$$\bar{X}_{123} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + N_3 \bar{X}_3}{N_1 + N_2 + N_3}$$

$$= \frac{(20 \times 305) + (25 \times 300) + (40 \times 340)}{20 + 25 + 40}$$

$$= \frac{6100 + 7500 + 13600}{85} = \frac{27200}{85} = 320$$

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

$$d_1 = |\bar{X}_1 - \bar{X}_{123}| = |305 - 320| = 15$$

$$d_2 = |\bar{X}_2 - \bar{X}_{123}| = |300 - 320| = 20$$

$$d_3 = |\bar{X}_3 - \bar{X}_{123}| = |340 - 320| = 20$$

$$\sigma_{123} = \sqrt{\frac{20(50)^2 + 25(40)^2 + 40(45)^2 + 20(15)^2 + 25(20)^2 + 40(20)^2}{20 + 25 + 40}}$$

$$= \sqrt{\frac{50000 + 40000 + 81000 + 4500 + 10000 + 16000}{85}}$$

$$= \sqrt{\frac{201500}{85}} = \sqrt{2370.59} = 48.69.$$

Thus the combined average wage is Rs. 320 and the combined standard deviation is Rs. 48.69.

Illustration 22. The mean and the standard deviation of a sample of size 10 were found to be 9.5 and 2.5 respectively. Later on, an additional observation became available. This was 15.0 and was included in the original sample. Find the mean and the standard deviation of the 11 observations.

Solution. $\bar{X} = \frac{\Sigma X}{N}$ or $9.5 \times 10 = \Sigma X$ or $\Sigma X = 95$

Adding 15, *i.e.*, 11th observation

$$\Sigma X = 95 + 15 = 110$$

$$\bar{X} = \frac{110}{11} = 10$$

$$\sigma^2 = \frac{\Sigma X^2}{N} - (\bar{X})^2 \quad \text{or} \quad (2.5)^2 = \frac{\Sigma X^2}{10} - (9.5)^2$$

or $6.25 \times 10 = \Sigma X^2 - 902.5$ or $\Sigma X^2 = 965$

On adding one more observation, *i.e.*, 15

$$\Sigma X^2 = 965 + (15)^2 = 1{,}190$$

$$\sigma^2 = \frac{\Sigma X^2}{N} - (\bar{X})^2 = \frac{1190}{11} - (10)^2$$

$$= 108.18 - 100 = 8.18 \quad \text{or} \quad \sigma = \sqrt{8.18} = 2.86.$$

Thus the mean and standard deviation of 11 observations are 10 and 2.86 respectively.

Illustration 23 A collar manufacturer is considering the production of a new style of collar to attract young men. The following statistics of neck circumference are available based on measurements of a typical group of the college students :

Mid-value (in inches) :	12.0	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0
Number of students :	2	16	36	60	76	37	18	3	2

Compute the standard deviation and use the criterion $\bar{X} \pm 3\sigma$ where σ is the standard deviation and $\bar{X}$ is the arithmetic mean to determine the largest and smallest size of the collar he should make in order to meet the needs of practically all the customers bearing in mind that collars are worn on average ½ inch longer than neck size. *(MBA, Delhi Univ, 2005)*

Solution. CALCULATION OF MEAN AND STANDARD DEVIATION

Mid-value in inches X	*No. of students* f	$(X-14)/10.5$ d	fd	fd^2
12.0	2	−4	−8	32
12.5	16	−3	−48	144
13.0	36	−2	−72	144
13.5	60	−1	−60	60
14.0	76	0	0	0
14.5	37	+1	+37	37
15.0	18	+2	+36	72
15.5	3	+3	+9	27
16.0	2	+4	+8	32
	$N = 250$		$\Sigma fd = -98$	$\Sigma fd^2 = 548$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 14 - \frac{98}{250} \times 0.5 = 13.8$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{548}{250} - \left(\frac{-98}{250}\right)^2} \times 0.5 = 1.43 \times 0.5 = 0.715$$

Largest and smallest neck size

$$= \bar{X} \pm 3\sigma = 13.8 \pm 3(0.715) = 11.655 \text{ and } 15.945.$$

Since collars are worn on an average 1/2 inch longer than the neck sizes, we should add 0.5 to these limits. Thus the smallest and largest sizes of collar should be

(11.655+0.5) and (15.945+0.5), *i.e.*,

12.155 or 12.2 inches and 16.445 or 16.4 inches.

Illustration 24. A study of the age of 100 persons grouped in intervals of 20–22, 22–24..... etc., revealed the mean age and standard deviation to be 32.02 and 13.18 respectively, While checking it was discovered that the observation 57 was misread as 27. Calculate the correct mean age and standard deviation. *(MBA, Delhi Univ., 1997)*

Solution. We are given : $N = 100$, $\bar{X} = 32.02$, $\sigma = 13.18$.

$$\bar{X} = \frac{\Sigma fX}{N} \text{ or } \Sigma fX = N\bar{X}$$

Uncorrected $\Sigma fX = 100 \times 32.02 = 3202$

$$\sigma^2 = \frac{1}{N}\Sigma fX^2 - (\bar{X})^2 \text{ or } \Sigma fX^2 = N[\sigma^2 + (\bar{X})^2]$$

Uncorrected $\Sigma fX^2 = 100\,[(13.18)^2 + (32.02)^2]$

$= 100(173.71 + 1025.28) = 100 \times 1198.99 = 119899$

Correct $\Sigma fX = 3202 - 27 + 57 = 3232$

Correct $\bar{X} = \frac{3232}{100} = 32.32$

Correct $\Sigma fX^2 = 119899 - (27) + (57)^2 = 119899 - 729 + 3249 = 122419$

$$\text{Correct } \sigma^2 = \frac{\text{Correct } \Sigma fX^2}{N} - (\text{Correct Mean})^2$$

$$\text{Correct } \sigma^2 = \frac{122419}{100} - (32.32)^2 = 1224.19 - 1044.58 = 179.61$$

Correct $\sigma = \sqrt{179.61} = 13.402$.

Illustration 25. For a group of 50 male workers, the mean and standard deviation of their daily wages are Rs. 63 and Rs. 9 respectively. For a group of 40 female workers, these are Rs. 54 and Rs. 6 respectively. Find the standard deviation of daily wages for the combined group of 90 workers. *(MBA, DU, 2002)*

Solution Combined Mean : $\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$

$N_1 = 50$, $\bar{X}_1 = 63$, $N_2 = 40$, $\bar{X}_2 = 54$

$$\bar{X}_{12} = \frac{(50 \times 63) + (40 \times 54)}{50 + 40} = \frac{3150 + 2160}{90} = \frac{5310}{90} = 59$$

Combined Standard Deviation

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

$\sigma_1 = 9$, $\sigma_2 = 6$, $d_1 = |\bar{X}_1 - \bar{X}_{12}| = |63 - 59| = 4$, $d_2 = |\bar{X}_2 - \bar{X}_{12}| = |54 - 59| = 5$

$$\sigma_{12} = \sqrt{\frac{50(9)^2 + 40(6)^2 + 50(4)^2 + 40(5)^2}{50 + 40}}$$

$$= \sqrt{\frac{4050 + 1440 + 800 + 1000}{90}} = \sqrt{\frac{7290}{90}} = \sqrt{81} = 9$$

Illustration 26. Calculate variance and coefficient of variation from the following data :

Profits (Rs. crores)	*No. of Companies*
Less than 10	8
Less than 20	20
Less than 30	40
Less than 40	70
Less than 50	90
Less than 60	100

(MBA, Jodhpur Univ., 2001)

Solution. CALCULATION OF VARIANCE AND COEFFICIENT OF VARIATION

Profits (Rs. crores)	*No. of companies* f	*m.p.* X	$(X-35)/10$ d	fd	fd^2
0–10	8	5	–3	–24	72
10–20	12	15	–2	–24	48
20–30	20	25	–1	–20	20
30–40	30	**35**	0	0	0
40–50	20	45	+1	+20	20
50–60	10	55	+2	+20	40
	$N=10$			$\Sigma fd=-28$	$\Sigma fd^2=200$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 35 - \frac{28}{100} \times 10 = 35 - 2.8 = 32.2$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{200}{100} - \left(\frac{-28}{100}\right)^2} \times 10$$

$$= \sqrt{2-.0784} \times 10 = 1.3862 \times 10 = 13.862$$

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{13.862}{32.2} \times 100 = 43.05$$

$$\text{Variance} = \sigma^2 = (13.862)^2 = 192.155$$

Illustration 27. The prices of a Tea Company shares in Mumbai and Kolkata markets during the last ten months are recorded below:

Month	Mumbai	*Kolkata*	*Month*	Mumbai	*Kolkata*
January	105	108	June	127	125
February	120	117	July	109	125
March	115	120	August	110	120
April	118	130	September	104	110
May	130	100	October	112	135

Determine the Arithmetic Mean and Standard Deviation of the prices of shares. In which market are the share prices stable?

(MBA, HPU, 2002)

Solution. For determining in which market prices of shares are more stable, we shall compare the coefficient of variation. Let prices in Mumbai and Kolkata be denoted by X and Y respectively :

CALCULATION OF COEFFICIENT OF VARIATION

X	$(X-\bar{X})$ x	x^2	Y	$(Y-\bar{Y})$ y	y^2
105	–10	100	108	–11	121
120	+5	25	117	–2	4
115	0	0	120	+1	1
118	+3	9	130	+11	121
130	+15	225	100	–19	361
127	+12	144	125	+6	36
109	–6	36	125	+6	36
110	–5	25	120	+1	1
104	–11	121	110	–9	81
112	–3	9	135	+16	256
$\Sigma X = 1150$	$\Sigma x = 0$	$\Sigma x^2 = 694$	$\Sigma Y = 1190$	$\Sigma y = 0$	$\Sigma y^2 = 1018$

Mumbai : C.V. $= \frac{\sigma}{\bar{X}} \times 100$ Kolkata : C.V. $= \frac{\sigma}{\bar{Y}} \times 100$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1150}{10} = 115 \qquad \bar{Y} = \frac{\Sigma Y}{N} = \frac{1190}{10} = 119$$

$$\sigma = \sqrt{\frac{\Sigma x^2}{N}} = \sqrt{\frac{694}{10}} = 8.33 \qquad \sigma = \sqrt{\frac{\Sigma y^2}{N}} = \sqrt{\frac{1018}{10}} = 10.09$$

$$\text{C.V.} = \frac{8.33}{115} \times 100 = 7.24 \qquad \text{C.V.} = \frac{10.09}{119} \times 100 = 8.48$$

Since the coefficient of variation is less in Mumbai, hence the share price in Mumbai market shows greater stability.

Illustration 28. The following data give the number of finished articles turned out per day by different number of workers in a factory :

No. of articles :	18	19	20	21	22	23	24	25	26	27
No. of workers:	3	7	11	14	18	17	13	8	5	4

Find the mean, standard deviation and coefficient of variation of daily output of finished articles.

(MBA, Kurukshetra Univ.,2001)

Solution

CALCULATION OF $\bar{X}$, S.D. AND C.V.

No. of articles X	$(X-22)$ d	f	fd	fd^2
18	–4	3	–12	48
19	–3	7	–21	63
20	–2	11	–22	44
21	–1	14	–14	14
22	0	18	0	0
23	+1	17	+17	17
24	+2	13	+26	52
25	+3	8	+24	72
26	+4	5	+20	80
27	+5	4	+20	100
		$N = 100$	$\Sigma fd = 38$	$\Sigma fd^2 = 490$

$$\bar{X} = A + \frac{\Sigma fd}{N} = 22 + \frac{38}{100} = 22 + .38 = 22.38$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{490}{100} - \left(\frac{38}{100}\right)^2} = \sqrt{4.9 - .1444} = 2.18$$

$$\text{C.V.} = \frac{2.18}{22.38} \times 100 = 9.74$$

Illustration 29. Based on the frequency distribution given below, compute the following statistical measures to characterise the distribution.

(*i*) Coefficient of variation, (*ii*) Inter-quartile range, (*iii*) Modal value.

Annual Tax Paid (Rs. Thousand)	*No. of Managers*	*Annual Tax Paid (Rs. Thousand)*	*No. of Managers*
5–10	18	25–30	20
10–15	30	30–35	12
15–20	46	35–40	6
20–25	28		

(M.M.S, Bombay Univ., 2002)

Solution :

CALCULATION OF COEFFICIENT OF VARIATION, SEMI-INTER QUARTILE RANGE AND MODE

Annual Tax Paid (Rs. thousand)	*m.p. X*	*f*	*(X—22.5)/5 d*	*fd*	fd^2	*c.f.*
5–10	7.5	18	–3	–54	162	18
10–15	12.5	30	–2	–60	120	48
15–20	17.5	46	–1	–46	46	94
20–25	**22.5**	28	0	0	0	122
25–30	27.5	20	+1	+20	20	142
30–35	32.5	12	+2	+24	48	154
35–40	37.5	6	+3	+18	54	160
		$N=160$		$\Sigma fd = -98$	$\Sigma fd^2 = 450$	

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 22.5 - \frac{98}{160} \times 5 = 22.5 - 3.0625 = 19.4375$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{450}{160} - \left(\frac{-98}{160}\right)^2} \times 5$$

$$= \sqrt{2.8125 - 0.3752} \times 5 = 1.5612 \times 5 = 7.806$$

$$\text{Coefficient of Variation : C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{7.806}{19.4375} \times 100 = 40.1594$$

Semi-Inter Quartile Range :

$$\text{Semi-Inter Quartile Range} = Q_3 - Q_1$$

$$Q_1 = \text{Size of } \frac{N}{4} \text{ th observation} = \frac{160}{4} = 40\text{th observation}$$

Q_1 lies in the class 10–15.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i = 10 + \frac{40-18}{30} \times 5 = 10 + 3.67 = 13.67$$

$$Q_3 = \text{Size of } \frac{3N}{4} \text{ th observation} = \frac{3 \times 160}{4} = 120\text{th observation}$$

Q_3 lies in the class 20–25.

$$Q_3 = L + \frac{3N/4 - p.c.f.}{f} \times i = 20 + \frac{120-94}{28} \times 5 = 20 + 4.64 = 24.64$$

$$\text{Semi-Inter Quartile Range} = Q_3 - Q_1 = 24.64 - 13.67 = 10.97$$

Mode: Since the highest frequency is 46, the mode lies in the class 15–20,

$$\text{Mo} = L + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i = 15 + \frac{16}{16+18} \times 5 = 15 + 2.35 = 17.35$$

Illustration 30. Calculate coefficient of variation from the following data :

Profits (Rs. crores):	10–20	20–30	30–40	40–50	50–60
No. of Cos. :	8	12	20	6	4

(MBA, Osmania Univ., 2002)

Solution. CALCULATION OF COEFFICIENT OF VARIATION

Profits (Rs. Crores)	*m.p. X*	*f*	*(X—35)/10 d*	*fd*	fd^2
10–20	15	8	–2	–16	32
20–30	25	12	–1	–12	12
30–40	35	20	0	0	0
40–50	45	6	+1	+6	6
50–60	55	4	+2	+8	16
		$N = 50$		$\Sigma fd = -14$	$\Sigma fd^2 = 66$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 35 - \frac{14}{50} \times 10 = 35 - 2.8 = 32.2$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{66}{50} - \left(\frac{-14}{50}\right)^2} \times 10$$

$$= \sqrt{1.32 - .0784} \times 10 = 11.14$$

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{11.14}{32.2} \times 100 = 34.6 \text{ per cent}$$

Illustration 31. Find the missing information from the following :

	Group I	*Group II*	*Group III*	*Combined*
Number	50	?	90	200
Standard Deviation	6	7	?	7.746
Mean	113	?	115	116

(MBA, HPU; MBA, Osmania Univ., 1997)

Solution. *Finding the number of observations in the second group.*

Let N_1, N_2, N_3 denote the number of observations in the 1st, 2nd and 3rd group respectively.

We are given $N_1 + N_2 + N_3 = 200$

$N_1 = 50$, $N_3 = 90$, $\therefore N_1 + N_3 = 140$

$N_2 = 200 - 140 = 60$

Finding Mean of the Second Group

$$\bar{X}_{123} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2 + N_3 \bar{X}_3}{N_1 + N_2 + N_3}$$

$\bar{X}_{123} = 116$, $N_1 + N_2 + N_3 = 200$, $\bar{X}_1 = 113$, $\bar{X}_3 = 115$

We have to find $\bar{X}_2$.

Substituting the given values

$$116 = \frac{50(113) + 60(\bar{X}_2) + 90(115)}{200}$$

$$116 \times 200 = 5650 + 60\ \bar{X}_2 + 10350$$

$$60\ \bar{X}_2 = 23200 - 16000 = 7200 \quad \text{or} \quad \bar{X}_2 = \frac{7200}{60} = 120$$

Finding S.D. of third group

$$\sigma_{123} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1d_1^2 + N_2d_2^2 + N_3d_3^2}{N_1 + N_2 + N_3}}$$

$\sigma_{123} = 7.746, N_1 = 50, \sigma_1 = 6, N_2 = 60, \sigma_2 = 7, N_3 = 90$

$d_1 = |\bar{X}_1 - \bar{X}_{123}| = 113 - 116 = 3$

$d_2 = |\bar{X}_2 - \bar{X}_{123}| = 120 - 116 = 4$

$d_3 = |\bar{X}_3 - \bar{X}_{123}| = 115 - 116 = 1$

Substituting the values

$$7.746 = \sqrt{\frac{50(6)^2 + 60(7)^2 + 90\sigma_3^2 + 50(3)^2 + 60(4)^2 + 90(1)^2}{50+60+90}}$$

$$= \sqrt{\frac{1800 + 2940 + 90\sigma_3^2 + 450 + 960 + 90}{200}} = \sqrt{\frac{6240 + 90\sigma_3^2}{200}}$$

Squaring on both sides, we get

or $$(7.746)^2 = \frac{6240 + 90\sigma_3^2}{200}$$

or $$12{,}000 = 6240 + 90\sigma_3^2 \text{ or } 90\sigma_3^2 = 12{,}000 - 6240$$

$$\sigma_3^2 = \frac{5760}{90} = 64 \quad \text{or} \quad \sigma_3 = \sqrt{64} = 8$$

Thus the missing values are :

$N_2 = 60, \ \bar{X}_2 = 120, \sigma_3 = 8.$

Illustration 32. Given below are the daily wages, in rupees, of 60 workers in a factory manufacturing plastic products :

23	48	51	64	72	82	56	33	50	42
35	88	77	65	39	52	48	64	49	57
41	73	62	49	32	54	67	46	55	50
82	44	75	56	51	63	59	69	53	42
75	85	68	55	52	45	42	57	20	57
46	51	20	16	62	46	54	40	55	71

(*a*) Form a frequency distribution, taking the lowest class-interval as 10–20.

(*b*) Calculate the Standard Deviation and Coefficient of Variation of this distribution.

(MBA, HPU,2002)

Solution. FORMATION OF FREQUENCY DISTRIBUTION
CALCULATION OF COEFFICIENT OF VARIATION

Wages (Rs.)	*Tally*	*Frequency f*	*m.p. X*	*(X—45)/10 d*	*fd*	fd^2
10–20	|	1	15	–3	–3	9
20–30	||	2	25	–2	–4	8
30–40	||||	4	35	–1	–4	4
40–50	𝍸 𝍸 |||	13	45	0	0	0
50–60	𝍸 𝍸 𝍸 𝍸 |	21	55	+1	+21	21
60–70	𝍸 ||||	9	65	+2	+18	36
70–80	|||| |	6	75	+3	+18	54
80–90	||||	4	85	+4	+16	64
		$N = 60$			$\Sigma fd = 62$	$\Sigma fd^2 = 196$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 45 + \frac{62}{60} \times 10 = 45 + 10.33 = 55.33$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{196}{60} - \left(\frac{62}{60}\right)^2} \times 10$$

$$= \sqrt{3.267 - 1.068} \times 10 = 1.483 \times 10 = 14.83$$

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{14.83}{55.33} \times 100 = 26.8 \text{ per cent.}$$

Illustration 33. Particulars regarding the income of two villages are given below :

	Village X	*Village Y*
Number of employees	600	500
Average income (in Rs.)	1750	1860
S.D. of income (in Rs.)	100	81

(*i*) In which village is the variation in income greater ?

(*ii*) What is the combined standard deviation of the village *X* and village *Y* put together ?

Solution. (*i*) Compare coefficient of variation to find out in which village there is greater variation in income.

$$\text{Village } X : \text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{100}{1750} \times 100 = 5.71$$

$$\text{Village } Y : \text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{81}{1860} \times 100 = 4.35$$

Since coefficient of variation is more in case of village *X*, hence in this village there is greater variation in the income of employees.

(*ii*) For finding combined variation we have to find first the combined mean.

$$\bar{X}_{12} = \frac{N_1 \bar{X}_1 + N_2 \bar{X}_2}{N_1 + N_2} = \frac{(600 \times 1750) + (500 \times 1860)}{600 + 500} = \frac{1050000 + 930000}{1100} = 1800$$

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$d_1 = |\bar{X}_1 - \bar{X}_{12}| = |1750 - 1800| = 50$$

$$d_2 = |\bar{X}_2 - \bar{X}_{12}| = |1860 - 1800| = 60$$

$$\sigma_{12} = \sqrt{\frac{600(100)^2 + 500(81)^2 + 600(50)^2 + 500(60)^2}{600 + 500}}$$

$$= \sqrt{\frac{6000000 + 3280500 + 1500000 + 1800000}{1100}}$$

$$= \sqrt{\frac{12580500}{1100}} = \sqrt{11436.82} = 106.94.$$

Illustration 34. The profits (in Rs. lakhs) earned by 100 companies during 2009-10 are shown below :

Profits (Rs. lakhs)	*No. of Companies*	*Profits (Rs. lakhs)*	*No. of Companies*
20–30	4	60–70	15
30–40	8	70–80	10
40–50	18	80–90	8
50–60	30	90–100	7

Compute (*a*) Mean, (*b*) Median and (*c*) Standard Deviation.

Solution. COMPUTATION OF MEAN, MEDIAN AND STANDARD DEVIATION

Profits (Rs. lakhs)	*No. of cos. f*	*m.p. X*	*(X—55)/10 d*	*fd*	fd^2	*c.f.*
20–30	4	25	–3	–12	36	4
30–40	8	35	–2	–16	32	12
40–50	18	45	–1	–18	18	30
50–60	30	**55**	0	0	0	60
60–70	15	65	+1	+15	15	75
70–80	10	75	+2	+20	40	85
80–90	8	85	+3	+24	72	93
90–100	7	95	+4	+28	112	100
	$N = 100$			$\Sigma fd = 41$	$\Sigma fd^2 = 325$	

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 55 + \frac{41}{100} \times 10 = 55 + 4.1 = 59.1$$

$$\text{Med.} = \text{Size of } \frac{N}{2} \text{ th observation} = \frac{100}{2} = 50\text{th observation}$$

Median lies in the class 50–60.

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i = 50 + \frac{50-30}{30} \times 10 = 56.67$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{325}{100} - \left(\frac{41}{100}\right)^2} \times 10$$

$$= \sqrt{3.25 - .1681} \times 10 = 1.756 \times 10 = 17.56$$

Illustration 35. Find the mean, median and standard deviation of the weight of billet in 8 gms as given in the following table :

(Wt. in gms)	*Frequency*	*(Wt. in gms)*	*Frequency*
210–215	8	230–235	14
215–220	13	235–240	10
220–225	16	240–245	7
225–230	29	245–250	3

Solution : CALCULATION OF $\bar{X}$, MED. AND STANDARD DEVIATION

Wt. in gms	*f*	*m.p. X*	*(X–227.5/5) d*	*fd*	fd^2	*c.f.*
210–215	8	212.5	–3	–24	72	8
215–220	13	217.5	–2	–26	52	21
220–225	16	222.5	–1	–16	16	37
225–230	29	**227.5**	0	0	0	66
230–235	14	232.5	+1	+14	14	80
235–240	10	237.5	+2	+20	40	90
240–245	7	242.5	+3	+21	63	97
245–250	3	247.5	+4	+12	48	100
	N=100			Σfd= +1	Σfd^2=305	

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 227.5 + \frac{1}{100} \times 5 = 227.55$$

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i$$

$$\text{Med.} = \text{Size of } \frac{N}{2}\text{th observation} = \frac{100}{2} = 50\text{th observation}$$

Median lies in the class 225–230.

$$\text{Med.} = L + \frac{N/2 - p.c.f.}{f} \times i = 227.5 + \frac{50-37}{29} \times 5 = 227.5 + 2.24 = 229.74$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{3.05 - .0001} \times 5 = 1.7464 \times 5 = 8.732$$

Illustration 36. The following distribution shows the turnover of the branches of a group of multiple-shops in March 2010:

Turnover (in Rs. lakhs)	*No. of Shops*
5 and under 10	8
10 and under 15	18
15 and under 20	42
20 and under 25	62
25 and under 30	30
30 and under 35	10
35 and over	4

Using assumed mean of Rs. 22,500, calculate (*i*) Mean, (*ii*) Standard deviation, and (*iii*) Coefficient of variation.

Solution. CALCULATION OF MEAN, STANDARD DEVIATION AND COEFFICIENT OF VARIATION

Turnover (in Rs. lakhs)	*m.p.* X	f	$(X-22.5)/5$ d	fd	fd^2
5–10	7.5	8	–3	–24	72
10–15	12.5	18	–2	–36	72
15–20	17.5	42	–1	–42	42
20–25	**22.5**	62	0	0	0
25–30	27.5	30	+1	+30	30
30–35	32.5	10	+2	+20	40
35–40	37.5	4	+3	+12	36
		$N = 174$		$\Sigma fd = -40$	$\Sigma fd^2 = 292$

Mean: $$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 22.5 - \frac{40}{174} \times 5 = 22.5 - 1.15 = 21.35$$

Standard Deviation : $$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{292}{174} - \left(\frac{40}{174}\right)^2} \times 5$$

$$= \sqrt{1.678 - .053} \times 5 = 1.275 \times 5 = 6.375$$

Coefficient of Variation : $\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 = \frac{6.375}{21.35} \times 100 = 29.86$ per cent.

Illustration 37. For two firms *A* and *B* belonging to same industry, the following details are available :

	Firm *A*	Firm *B*
Number of Employees :	100	200
Average monthly wage :	Rs. 4,800	Rs. 5,100
Standard deviation :	Rs. 600	Rs. 540

Find (*i*) Which firm pays out larger amount as wages ?

(*ii*) Which firm shows greater variability in the distribution of wages ?

(*iii*) Find average monthly wage and the standard deviation of the wages of all employees in both the firms.

Solution. (*i*) For finding out which firm pays larger amount, we have to find out ΣX. (*MBA, Delhi Univ., 2006*)

$$\bar{X} = \frac{\Sigma X}{N} \quad \text{or} \quad \Sigma X = N\bar{X}$$

Firm *A* : $N = 100$, $\bar{X} = 4800$, $\therefore \Sigma X = 100 \times 4800 = 4,80,000$

Firm *B* : $N = 200$, $\bar{X} = 5100$, $\therefore \Sigma X = 200 \times 5100 = 10,20,000$

Hence firm *B* pays larger amount as monthly wages.

(*ii*) For finding out which firm shows greater variability in the distribution of wages, we have to calculate coefficient of variation.

$$\text{Firm } A: \ C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{600}{4800} \times 100 = 12.50.$$

$$\text{Firm } B: \ C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{540}{5700} \times 100 = 10.59.$$

Since coefficient of variation is greater in case of firm *A*, hence it shows greater variability in the distribution of wages.

(*iii*) Combined average monthly wage :

$$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

$$N_1 = 100, \ \bar{X}_1 = 4800, N_2 = 200, \ \bar{X}_2 = 5100$$

$$\text{Hence } \bar{X}_{12} = \frac{(100)(4800) + (200)(5100)}{100 + 200} = \frac{480000 + 1020000}{300} = \text{Rs. } 5{,}000$$

Combined Standard Deviation

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$$

$$N_1 = 100, \sigma_1 = 600, N_2 = 200, \sigma_2 = 540$$

$$d_1 = |\bar{X}_1 - \bar{X}_{12}| = |4800 - 5000| = 200$$

$$d_2 = |\bar{X}_2 - \bar{X}_{12}| = |5100 - 5000| = 100$$

$$\sigma_{12} = \sqrt{\frac{100(600)^2 + 200(540)^2 + 100(200)^2 + 200(100)^2}{100 + 200}}$$

$$= \sqrt{\frac{36000000 + 58320000 + 4000000 + 2000000}{300}}$$

$$= \sqrt{\frac{100{,}320{,}000}{300}} = 578.27$$

Hence the combined standard deviation is Rs. 578.27.

Illustration 38. From the following frequency distribution of heights of 360 boys in the age-group 15–20 years, calculate the: (*i*) arithmetic mean ; (*ii*) coefficient of variation ; and (*iii*) quartile deviation.

Heights (in cms)	*No. of boys*	*Height (in cms)*	*No. of boys*
126–130	31	146–150	60
131–135	44	151–155	55
136–140	48	156–160	43
141–145	51	161–165	28

Solution. CALCULATION OF $\bar{X}$, Q.D., AND C.V.

Height (in cms)	*m.p. X*	*f*	*(X–143)/5 d*	*fd*	fd^2	*c.f.*
126–130	128	31	–3	–93	279	31
131–135	133	44	–2	–88	176	75
136–140	138	48	–1	–48	48	123
141–145	143	51	0	0	0	174
146–150	148	60	+1	+60	60	234
151–155	153	55	+2	+110	220	289
156–160	158	43	+3	+129	387	332
161–165	163	28	+4	+112	448	360
		N=360		Σfd= 182	Σfd^2=1618	

(*i*) $$\bar{X} = A + \frac{\Sigma fd}{N} \times i$$

$$= 143 + \frac{182}{360} \times 5 = 143 + 2.53 = 145.53$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times i = \sqrt{\frac{1618}{360} - \left(\frac{182}{360}\right)^2} \times 5$$

$$= \sqrt{4.494 - 0.256} \times 5 = 2.059 \times 5 = 10.295$$

(*ii*) $$C.V. = \frac{\sigma}{\bar{X}} \times 100$$

$$= \frac{10.295}{145.53} \times 100 = 7.074 \text{ per cent}$$

(*iii*) $$Q.D. = \frac{Q_3 - Q_1}{2}$$

Q_1 = Size of $\frac{N}{4}$ th observation = $\frac{360}{4}$ = 90th observation

Q_1 lies in the class 136–140. But the real limit of this class is 135.5–140.5.

$$Q_1 = L + \frac{N/4 - p.c.f.}{f} \times i = 135.5 + \frac{90 - 75}{48} \times 5 = 135.5 + 1.56 = 137.06$$

Q_3 = Size of $\frac{3N}{4}$ th observation = $3 \times \frac{360}{4}$ = 270th observation

Q_3 lies in the class 151–155. But the real limit of this class is 150.5–155.5.

$$Q_3 = L + \frac{3N/4 - p.c.f.}{f} \times i = 150.5 + \frac{270 - 234}{55} \times 5 = 150.5 + 3.27 = 153.77$$

$$Q.D. = \frac{153.77 - 137.06}{2} = 8.355.$$

Illustration 39. A welfare organisation introduced on education scholarship scheme for the school going children of a backward village. The rates of scholarship were fixed as given below :

Age Group (in yrs.)	*Amount of Scholarship per month (Rs.)*
5–7	300
8–10	400
11–13	500
14–16	600
17–19	700

The ages (years) of 30 school going children are noted as 11, 8, 10, 5, 7, 12, 7, 17, 5, 13, 9, 8, 10, 15, 7, 12, 6, 7, 8, 11, 14, 18, 6, 13, 9, 10, 6, 15, 3, 5 years respectively. Calculate mean and standard deviation of monthly scholarship. Find out the total monthly scholarship amount being paid to the students. (*MBA, IGNOU, 2002*)

Solution : Let us first classify the given data in various class intervals.

Age group (yrs)	*Tally bars*	*Freq.*
5–7	𝍸 𝍸	10
8–10	𝍸 |||	8
11–13	𝍸 ||	7
14–16	|||	3
17–19	||	2

CALCULATION OF MEAN AND STANDARD DEVIATION

Age group (yrs.)	*m.p.* m	f	$(m-12)/3$ d	fd	fd^2
5–7	6	10	–2	–20	40
8–10	9	8	–1	–8	8
11–13	12	7	0	0	0
14–16	15	3	+1	+3	3
17–19	18	2	+2	+4	8
		$N = 30$		$\Sigma fd = -21$	$\Sigma fd^2 = 59$

$$\bar{X} = A + \frac{\Sigma fd}{N} \times i = 12 - \frac{21}{30} \times 3 = 12 - 2.1 = 9.9$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{n} - \left(\frac{\Sigma fd}{n}\right)^2} \times i = \sqrt{\frac{59}{30} - \left(\frac{-21}{30}\right)^2} \times 3$$

$$= \sqrt{1.967 - 0.49 \times 3} = 1.2153 \times 3 = 3.65$$

CALCULATION OF TOTAL MONTHLY SCHOLARSHIP

No. of students	*Amount of scholarship per month (Rs.)*	*Total Monthly Scholarship (Rs.)*
10	300	3000
8	400	3200
7	500	3500
3	600	1800
2	700	1400
		Rs. 12900

Illustration 40. The performance of two teams is summarized below :

	Mean	*S.D.*
Team '*A*'	10	2
Team '*B*'	15	4

Which is more consistent team? *(MBA, G.G.S.I.P. Univ., 2009)*

Solution : For finding out which is more consistent team, we compare the coefficient of variation of the two teams.

Team *A* $\quad C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{2}{10} \times 100 = 20$

Team *B* $\quad C.V. = \frac{\sigma}{\bar{X}} \times 100 = \frac{4}{15} \times 100 = 26.67$

Since coefficient of variation is less for Team '*A*', hence Team '*A*' is more consistent.

PROBLEMS

1-A : Answer the following questions, each question carries **one** mark :

(*i*) What is the formula for coefficient of variation ?

(*ii*) What is range ?

(*iii*) How quartile deviation is calculated ?

(*iv*) What is interquartile range ?

(*v*) State the formula for standard deviation. *(MBA., Madurai-Kamaraj Univ., 2003)*

(*vi*) What is Lorenz Curve ?

(*vii*) Is standard deviation independent of change of scale and origin .

(*viii*) Is the sum of deviations from mean is always least ?

(*ix*) Is variance the square of standard deviation ?

(*x*) Give the formula for combined standard deviation of two sets of data.

1-B : Answer the following questions, each question carries **four** marks :

(*i*) What are the various methods of measuring variation ?
(*ii*) Distinguish between mean deviation and standard deviation.
(*iii*) What are the properties of a good measure of variation ?
(*iv*) Distinguish between absolute and relative measures of dispersion.
(*v*) Why standard deviation is most widely used as a measure of variations.
(*vi*) What are the uses of Lorenz Curve ?

2. (*a*) Explain the term variation. What purpose does a measure of variation serve ? In the light of these, comment on some of the well-known measures of variation along with their respective merits and demerits.
(*b*) Point out the difference between absolute and relative variation.
(*c*) Under what circumstances range is more meaningful than any other measure of variation ? *(MBA, HPU, 2009)*

3. (*a*) What are the requisites of a good measure of dispersion ? Why is the standard deviation usually chosen as a measure of variation ?
(*b*) Explain how measure of central tendency and measure of variation complement each other in describing mass of data.

4. What is coefficient of variation ? What purpose does it serve ? Also distinguish between 'variance' and 'coefficient of variation'.

5. What do you understand by standard deviation? Explain its usefulness. Highlight its important properties.

6. With an example show that mean is dependent on both origin and scale while standard deviation is dependent on scale but not on origin.

7. (*a*) State the different measures of central tendency and variation.
(*b*) What do you understand by "coefficient of variation" ? Discuss its importance in business problems.

8. What is Lorenz Curve ? How is it drawn ? In what way does it help in studying variations of two or more distributions ? Illustrate with the help of an example.

9. (*a*) Explain with suitable examples the term 'variation'. Mention some common measures of variation and describe the one which you think is the most important.
(*b*) Critically examine the different methods of measuring variation. Which of these do you consider as the best and why ? *(MBA, KU, 2002)*

10. Explain and illustrate how the measures of variation afford a supplement to the information about frequency distribution furnished by averages. *(MBA, Delhi Univ., 2001)*

11. Explain briefly what is meant by 'Variation' of data. State the requisites of a satisfactory measure of dispersion and examine in their light any two measures of dispersion.

12. What are the uses of coefficient of variation in statistical analysis ?

13. Describe the various methods of measuring variation used in business decisions pointing out their limitations, if any. *(MBA, Delhi Univ., 2004)*

14. (*a*) Describe briefly the various methods of measuring variation in health statistics.
(*b*) What are the uses of coefficient of variation in statistical analysis ?

15. (*a*) Briefly describe the characteristics of a standard deviation as a measure of variation.
(*b*) Comment on the following statements :

(*i*) If the mean and standard deviation of n observations $x_1, x_2 \ldots\ldots x_n$ be $\overline{X}$ and σ respectively then the mean and the standard deviation of $-x_1, -x_2 \ldots\ldots -x_n$ will be $\overline{X}$ and $-\sigma$ respectively.

(*ii*) "After settlement the average daily wage in a factory had increased from Rs. 85 to Rs. 90 and the standard deviation had increased from 10 to 12.5. After settlement the wage has become higher and more uniform."

16. (*a*) Ten observations have mean 20 and standard deviation 5. If each of these 10 observations is doubled then the standard deviation of new observations will be...
(*b*) State whether the following statement is *true* or *false*:
Range is a measure of variation which gives us information about scatter of values about the measure of a central tendency.

17. Calculate mean deviation for the following frequency distribution :

No. of colds experienced in 12 months	*No. of persons*	*No. of colds experienced in 12 months*	*No. of persons*
0	15	5	95
1	46	6	82
2	91	7	26
3	162	8	13
4	110	9	2

[M.D.=1.47]

18. Calculate the appropriate measure of variation from the following data:

Daily Wages (in Rs.)	*No. of wage earners*	*Daily Wages (in Rs.)*	*No. of wage earners*
Less than 85	14	91–93	18
85–87	62	Over 93	7
88–90	99		

19. Compute coefficient of variation from the following table :

Weekly Income (in Rs.)	*No. of employees*	*Weekly Income (in Rs.)*	*No. of employees*
1300–1399	30	1700–1799	60
1400–1499	46	1800–1899	50
1500–1599	58	1900–1999	20
1600–1699	76		

20. The following table gives the fluctuations in the prices of shares of two companies *A* and *B*. Find out which of them shows greater variability. Comment on the result.

Price (in Rs.) Share A	*Price (in Rs.) Share B*	*Price (in Rs.) Share A*	*Price (in Rs.) Share B*
318	2542	324	2545
322	2522	315	2530
325	2534	308	2556
312	2532	319	2530

[Share *A* : C.V. = 1.75 ; Share *B* : C.V. = 0.42]

21. From an analysis of monthly wages paid to workers in two organisations *C* and *D*, the following results were obtained :

	C	*D*
No. of workers	550	600
Average monthly wages (Rs.)	2260	2348
Variance of the distribution of wages	100	144

Obtain the average monthly wages and the variability in individual wages of all the workers in the organisations taken together.

22. From the following table giving data regarding income of workers in two factories draw a graph (Lorenz curve) to show which factory has greater inequalities of income :

Income (Rs.)	*Factory A*	*Factory B*
Below 4500	6,000	5,000
4,500—5,000	4,250	4,500
5,000—6,000	3,600	4,800
6,000—7,000	1,500	2,200
7,000—8,000	650	1,500

[Factory *A*]

23. A distribution consists of three parts, characterised as follows :

Parts	*Number of items*	*Arithmetic average*	*S.D.*
1	200	5	3
2	250	10	4
3	300	15	5

Show that the arithmetic average of the whole distribution is 10.67 and its standard deviation is 5.83 approximately.

24. Lives of two models of refrigerators in a recent survey are :

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

What is the average life of each of these refrigerators ? Which model has greater uniformity ?
(MBA, Bharthidasan Univ., 2001; IAS, 2002; C.S.E., 2002)

[Model A: $\bar{X} = 5.12$, C.V. = 54.9; Model B: $\bar{X} = 6.16$, C.V. = 36.2]

25. The mean of two samples of size 50 and 100 respectively is 54.1 and 50.3 and the standard deviations are 8 and 7. Obtain the standard deviation of the sample of size 150 obtained by combining the two samples.
[7.55]

26. The mean and standard deviation 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 standard deviation if item 8 was incorrect. Calculate the mean and standard deviation if (*i*) the wrong item is omitted, and (*ii*) it is replaced by12. [10.1, 2.26, 10.2,2]

27. The mean and standard deviation of 1,000 observations of a frequency distribution (grouped in intervals 0–10, 10–20, etc.) were found to be 35 and 8. Later it was discovered that in calculating these values the errata which was supplied with the data was not considered. The errata read as follows :

Group 0–10 for frequency 25 read frequency 52
Group 10–20 for frequency 75 read frequency 57
Group 20–30 for frequency 121 read frequency 21
Group 30–40 for frequency 137 read frequency 73
Group 40–50 for frequency 59 read frequency 95.

Calculate the correct mean and standard deviation.

28. An analysis of the weekly wages paid to workers in two firms *A* and *B* belonging to the same industry, gives the following results:

	Firm A	*Firm B*
Number of wage-earners	550	650
Average daily wages	100	95
Standard deviation	$\sqrt{90}$	$\sqrt{120}$

(*a*) Which firm *A* or *B* pays out large amount as daily wages ?
(*b*) In which firm *A* or *B* is there greater variability in individual wages ?
(*c*) What are the measures of (*i*) average daily wages and (*ii*) standard deviation in the distribution of individual wages of all workers in the two firms taken together ? *(MBA, M.D. Univ.; Diploma in Mgt., AIMA, 1999)*

29. A factory produces two types of electric lamps *A* and *B*. In an experiment relating to their life, the following results were obtained :

Length of life (in hours)	No. of lamps *A*	No. of Lamps *B*
500–700	5	4
700–900	11	30
900–1,100	26	12
1,100–1,300	10	8
1,300–1,500	8	6

Compare the variability of the life of the two types using coefficient of variation.
$[CV_A = 21.64; CV_B = 23.41]$ *(AIMA, 2005)*

30. In a small town, a survey was conducted in respect of profits made by retail shops. The following results were obtained:

Profit or Loss (in '000 Rs.)	*No. of shops*	*Profit or Loss (in '000 Rs.)*	*No. of shops*
–4 to –3	4	1 to 2	56
–3 to –2	10	2 to 3	40
–2 to –1	22	3 to 4	24
–1 to 0	28	4 to 5	18
0 to 1	38	5 to 6	10

Calculate (*i*) the average profit made by a retail shop.
(*ii*) total profit made by all shops, and (*iii*) the coefficient of variation of earnings.
[(*i*) 1348; (*ii*) 3,37,000; (*iii*) 152.8)]

31. The following is the distribution of amounts spent for research and development and for marketing for the year 2010 by 10 drug firms and cosmetic firms :

Expenditure (Lakhs of rupees)	*Drug Cos.* *R and D*	*Marketing*	*R and D*	*Cosmetic Cos.* *Marketing*
10–20	2	3	5	0
20–30	2	2	3	2
30–40	3	3	2	4
40–50	3	2	0	4

(*a*) Compute the arithmetic mean for each type of company, for each type of expenditure.

(*b*) Compute the standard deviation for each type of company, for each type of expenditure.

(*c*) What conclusion do you draw from the results ? *(MBA, Kurukshetra Univ.)*

[$\overline{X}$ = 32, 29, 22, 37; σ = 11, 11.14, 7.81, 7.48]

32. The life of two types of tyres in a sample survey is given below :

Life (in km)	*Type A*	*Type B*
5,000–10,000	18	15
10,000–15,000	22	24
15,000–20,000	26	30
20,000–25,000	25	18
25,000–30,000	9	13

(*a*) Which of the two types of tyre give a higher average life?

(*b*) If prices are same for both the types, which type would you prefer and why ? *(MBA, Delhi Univ., 1999)*

[(*a*) type *B*, (*b*) Type *B*]

33. A collar manufacturer is considering the production of new style of collar to attract young men. The following statistics of neck circumference are available based on measurement of a typical group:

Mid-value (in inches)	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0
No. of Students	4	19	30	63	66	29	18	1

Compute the arithmetic mean and standard deviation and comment on the results.

[14.22, 0.70]

34. Calculate the arithmetic mean, median and standard deviation for the following distribution :

Height (inches)	*No. of persons*	*Height (inches)*	*No. of persons*
60 less than 63	4	69 less than 72	33
63 less than 66	14	72 less than 75	8
66 less than 69	59	75 less than 78	2

[68.32; 68.14; 2.79]

35. The following data relate to the number of bonds applied for number of applicants and number of bonds allotted to each applicant by the Rayon Silk Mfg. (Wvg.) Co. Ltd. in Jan. 2004:

No. of Bonds applied for	*No. of Applicants*	*No. of bonds allotted to each applicant*
5–15	61,685	5
20–45	18,879	10
50–105	7,230	15
110–185	647	20
190–300	177	25
305–325	79	30

Calculate the average number of bonds allotted to each applicant and the standard deviation.

36. The standard deviation of a distribution of 100 values was Rs. 2. If the sum of the squares of the actual values was Rs. 3,600, what was the mean of this distribution ?

[5.66]

37. 32 trials of a process to finish a certain job revealed the following information :

Mean time taken to complete the job = 1 hr. 20 mts.

Standard deviation =16mts.

Another set of 8 trials gave mean time as 100 minutes and standard deviation 25 minutes.

Find the combined mean and standard deviation.

[$\overline{X}_{12}$ = 84, σ_{12} =19.84]

38. The following table relates to the profits and losses of 100 firms. Calculate the average profits and the standard deviation of profits :

Profits & Loss	*No. of firms*	*Profits & Loss*	*No. of firms*
5,000–6,000	8	0–1,000	6
4,000–5,000	12	(–)1,000–0	5
3,000–4,000	30	(–)2,000–(–)1,000	8
2,000–3,000	10	(–)3,000–(–)2,000	9
1,000–2,000	5	(–)4,000–(–)3,000	7

[$\bar{X} = 531$, $\sigma = 214.8$, C.V.= 40.5%]

39. A study of 241 authors revealed the following data on the distribution of age :

Age (years)	*Number of Authors*
up to 30	20
up to 40	73
up to 50	80
up to 60	44
up to 70	22
up to 80	2

Compute the mean and coefficient of variation of the distribution.

[$\bar{X} = 44.21$, $\sigma = 11.18$, C.V. = 25.29]

40. A survey of domestic consumption of electricity gave the following distribution of the no. of units consumed:

Number of units	*Number of consumers*	*Number of units*	*Number of consumers*
0–200	9	800–1000	45
200–400	18	1000–1500	38
400–600	27	1500–2000	20
600–800	32	2000 and above	2

Use a graphical method to calculate as accurately as possible the two quartiles and hence find the quartile deviation.

[$Q_1 = 570$, $Q_3 = 1250$, $Q.D. = 340$]

41. The Shareholder Research Bureau of India conducted recently a research study on the price behaviour of three leading industrial shares *A*, *B*, *C* for the period 2005 to 2008, the results of which are published as following in the quarterly journal :

Share	*Average Price* (Rs.)	*Standard Deviation* (Rs.)	*Current selling price* (Rs.)
A	18.00	5.40	36.00
B	22.50	4.50	34.75
C	24.00	6.00	39.00

(*i*) Which share in your opinion appears to be more stable in value ?

(*ii*) If you are the holder of all three shares which one would you like to dispose of at present, and why ?

[C.V. : (A) = 30, (B) = 20, (C) = 25]

42. Find the missing value from the following table :

Sub-group	*N*	$\bar{X}$	*Variance*
A	?	25	9
B	250	?	96
C	300	15	?
	750	16	51.733

[$N_1 = 200$, $\bar{X}_2 = 10$, $\sigma^2_3 = 25$]

43. The number examined, the mean weight and the standard deviation in each group of examination and two medical examiners are given below. Find the weight and standard deviation of both groups taken together.

Medical Examiner	*Number Examined*	*Mean Weight*	*Standard Deviation*
A	50	113 pounds	6.5 pounds
B	60	120 pounds	8.2 pounds

44. Blood serum cholesterol levels of 10 persons are as under :

240, 260, 290, 245, 255, 288, 272, 263, 277, 250.

Calculate standard deviation with the help of assumed mean.

[$\sigma = 16.48$]

45. Mean and standard deviation of the following continuous series are 31 and 15.94 respectively. The distribution after step deviations is as follows :

d :	–3	–2	–1	0	1	2	3
f :	10	15	25	25	10	11	5

Determine the actual class-intervals. *(MBA, Vikram Univ., 2002)*

[$i = 10$, $A = 35$, 0–10, 10–20, 20–30, 30–40 etc.]

46. The mean of 5 observations is 15 and the variance is 9. If two more observation having values –3 and 10 are combined with these 5 observations, what will be the mean and variance of 7 observations ?

47. Daily rated employees in SLM-Devilal, an engineering firm, earn Rs. 66 per day. The workers estimate that on the average they turn out 30 pieces per day with a standard deviation of six pieces per day. Under a suggested piece rate plan how much will they ask per piece if they wish to earn more than their present daily income 90 per cent of the time ?

48. The coefficients of variation of wages of male workers and female workers are 55 per cent and 70 per cent respectively, while the standard deviations are 220 and 15.4 respectively. Calculate the overall average wage of all workers given that 80 per cent of the workers are male.

49. Calculate the standard deviation from the following data :

Temperature (C)	*No. of days*	*Temperature (C)*	*No. of days*
–40 to –30	10	0 to 10	65
–30 to –20	24	10 to 20	180
–20 to –10	30	20 to 30	14
–10 to –0	42		

(MBA, Jodhpur Univ., 1996)

50. Calculate coefficient of variation from the following data :

No. of days absent:	0–5	5–10	10–15	15–20	20–25	25–30
No. of Students :	29	140	250	108	52	21

[C.V. = 41.32%]

51. The index number of prices of Cotton and Coal shares in April 2008 were as under :

Month	*Jan.*	*Feb.*	*March*	*April*	*May*	*June*	*July*	*Aug.*	*Sept.*	*Oct.*
Cotton	188	178	173	164	172	184	184	185	211	217
Coal	131	130	130	129	129	120	127	127	130	137

Which of the shares you consider more variable in price?

52. (*a*) The mean and standard deviation of 17 observations were found to be 25 and 5 respectively. Later on it was found that two values 51 and 31 were wrongly read as 35 and 13 respectively. Find the correct mean and standard deviation.

[Correct : $\bar{X} = 27$, $\sigma = 6.96$]

(*b*) In a survey, data on daily wages paid to workers of two factories *A* and *B* are as follows :

(No. of workers)

Daily wages (Rs.) :	*20 – 30*	*30 – 40*	*40 – 50*	*50 – 60*	*60 – 70*	*70 – 80*	*80 – 90*
Factory *A* :	15	30	44	60	30	14	7
Factory *B* :	25	40	60	35	20	15	5

Find out :

(*a*) Which factory pays higher average wages ? By how much ?

(*b*) Wages of which have greater variability.

(*c*) Monthly wage bill of both Factories (month = 25 days).

53. The number of employees, wages per employee and the variance of the wages per employee for two factories is given below:

	Factory A	*Factory B*
No. of employees	100	150
Average wage per employee per month (Rs.)	3,200	2,800
Variance of the wages per employee per month (Rs.)	625	729

(*a*) In which factory is there greater variation in the distribution of wages per employee ?

(*b*) Suppose in factory *B*, the wages of an employee were wrongly noted as Rs. 3050 instead of 3650, what would be the correct variance for factory *B* ?

(*MBA, Kumaun Univ., 2003*)

54. Name the various measures of dispersion. How would you compare the performance of two companies which reported profits for last five years as follows :

Company I	:	4.0	4.1	4.3	4.0	4.1
Company II	:	7.3	–3.7	8.4	–2.5	11.0

(*MBA, M.D. Univ., 2000*)

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

Profits (*Rs. crore*)	*No. of Cos.*	*Profits* (*Rs. crore*)	*No. of Cos.*
Less than 10	8	Less than 40	70
" " 20	20	" " 50	90
" " 30	40	" " 68	100

(*MBA, Guru Jameshwar Univ., 2003*)

56. Assume that your pathology lab. provides the following details of blood test carried out on 100 patients for diabetes :

Blood Sugar	:	90 –100	100 –110	110 –120	120 –130	130 –140	140 –150	150 –160
No. of Patients	:	16	20	22	25	10	4	3

Calculate coefficient of variation. What inference do you draw.

(*MBA, HCA, 2002*)

($\bar{X}$ = 116.7, σ = 15.17, C.V. = 13%)

57. For two firms A and B belonging to the same industry, the following details are available.

	Firm A	*Firm B*
No. of Employees	100	200
Average monthly wage	Rs. 2400	Rs. 1800
S.D.	Rs. 60	Rs. 80

(*i*) Which firm pays out larger amount as wages ?

(*ii*) Which firm shows greater variability in the distribution of wages ?

(*iii*) Find average monthly wage and standard deviation of all the employees in both the firms.

(*MBA, D.U. 2006*)

*****