1. Meaning of Dispersion

Dispersion is a statistical measure which shows the scatteredness of the values of the items of a series. Dispersion is used in two senses. *In first sense,* it measures the extent to which individual items vary. Larger the variations in individual items, the greater will be the dispersion. *In second sense* it signifies the extent to which the individual measures differ on an average from the mean. In other

We observe that some deviations from the central value are positive, while others are negative. Also, some deviations (or variations) are large and others are small. We, therefore, require an overall **'summary measure'** of variation in all values from the central value. This summary measure is called the measure of dispersion. **Dispersion is, thus a type of average and is also known as average of the second order.**

2. Objectives of Dispersion

The study of dispersion is important because of the following reasons:

1. **To Assess the Reliability of an Average.** Measure of dispersion shows how far an average is representative of statistical series, If the dispersion is small, the average will closely represent the individual observations and it is reliable in the sense that it makes a good estimate of the universe. On the other hand, if the dispersion is large, the average will be quite unreliable.

2. **For Controlling the Variability.** Measures of variation help us to control the variability in quantitative data. For example, in industrial production, efficient operation requires control of quality variation, the causes of which are sought through quality control programmes. Similarly, variation in body temperature and blood pressure are the best guides to doctors to diagnose properly. In social sciences, the measurement of inequality in the distribution of income and wealth requires the measures of dispersion. The government can thus adopt suitable policy to reduce economic inequalities.

3. **For Comparing two or more series with regard to their Variability.** The extent of variability between two or more series can be compared with the help of measures of dispersion. A high degree of variation would mean less consistency or less uniformity as compared to the data having less variation.

4. **For Facilitating the use of other Statistical Measures.** Many other statistical measures such as correlation, regression, testing of hypothesis, etc. are based on the measures of variation in one kind or another.

3. Characteristics (or Essentials) of a Good Measure of Dispersion

A good measure of dispersion should possess the following properties:

(i) It should be simple to understand.
(ii) It should be rigidly defined.
(iii) It should be easy to calculate.
(iv) It should be based on all observations of the distribution.
(v) It should be suited to further algebraic treatment.
(vi) It should be least affected by sampling fluctuations.
(vii) It should not be unduly affected by extreme value of the items.

4. Types of Measures of Dispersion

Measures of dispersion may be of two types:

1. Absolute measures, and
2. Relative measures.

4.1 Absolute Measures.

The measures of dispersion which are expressed in terms of the original units of a series are called absolute measures. These measures of dispersion are expressed in the same statistical units, e.g., rupees, kilograms, tonns, years, centimetres, etc. These values are used to compare the variation in two distributions having expressed in the same units. Absolute measures are used when only one set of data are taken into account. If frequency distributions are expressed in different units, these measures do not help. Range, quartile deviation, mean deviation and standard deviation are absolute measures of dispersion.

4.2 Relative Measures.

As discussed above, the absolute measures of dispersion, are measured in the same units. But when comparison is to be made between two sets of values having different units of measurement, it creates difficulty. For example, it is difficult to compare the dispersion of incomes in India with those in USA. The former are in terms of rupees while the latter are in dollars. This difficulty can be overcome if we use a relative measure of dispersion, which does not depend on units of measurement. *Relative measure of dispersion are calculated as the percentage or the coefficient of the absolute measure of dispersion.* Therefore, they are called relative measure or coefficient of dispersion. The relative measures of dispersion are: coefficient of range, coefficient of quartile deviation, coefficient of mean deviation, coefficient of standard deviation.

Difference (i) Absolute measures are expressed in the same units, whereas relative measures express the variability of data in terms of some relative value or percentage. (ii) Absolute measures of dispersion are used when one set of statistical distribution is studied, whereas relative measures are used when two or more series are considered simultaneously.

5. Methods of Measuring Dispersion

Following are the main methods of measuring dispersion:

I. **Methods of limits**
   (A) Range
   (B) Inter-Quartile range
   (C) Percentile range.
II. **Methods of averaging deviations**
   (A) Quartile deviation (or Semi-inter-quartile range)
   (B) Mean deviation
   (C) Standard deviation.
III. **Graphic method Lorenz curve.**

Box 9.1 Characteristics of the Range

Following are the characteristics of range.

(i) It is rigidly defined.
(ii) It is easy to calculate and simple to understand.
(iii) It does not depend on all values of the variable.
(iv) It is unduly affected by the extreme values.

5.1. Methods of Limits

5.1.1. Range: Meaning and its Calculation. Range is the simplest method of measuring dispersion. *it is the difference between the largest value and the smallest value of a series.* To find out range of a series, firstly, the largest and the smallest values are ascertained.

Symbolically,

$$R = L - S$$

Here R = Range, L = Largest value, S = Smallest value

Higher value of range implies higher dispersion and vice versa. Range will be zero, if all the observations in the series have the same value. For example, if the variable takes a constant value say 5, that is 5, 5, 5, 5, 5, 5 then largest value and smallest value is 5 and hence there is no dispersion in the series.

Coefficient of Range. *The relative measure of dispersion corresponding to range is called coefficient of range.* It is independent of the units of measurement. It is obtained by applying the following formula:

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

Individual Series:

Example 1. Find out range and coefficient of range from the following data:

Monthly	600	500	400	300	800
Salary	750	900	350	650	850
(in Rs.)	820	550	675	875	

Solution: Range = L — S = 900 — 300 = Rs. 600 **Ans.**

$$\text{Coefficientof Range} = \frac{L-S}{L+S} = \frac{900-300}{900+300}$$

$$= \frac{600}{1200} = 0.50 \textbf{ Ans.}$$

Discrete Series. In a discrete series, range is calculated by subtracting the lowest val

Merits and Demerits of Range

Merits

(i) Range is very easy to calculate as we need to know only two extreme values.

(ii) It is simple to understand.

(iii) It is helpful in statistical quality control and weather forecasting.

(iv) It is not affected by frequencies.

Demerits

(i) Range is not based on all observations.

(ii) It is unduly affected by extreme values.

(iii) It cannot be calculated in open end classes.

(iv) It does not give any indication of the nature of the distribution within the two extreme values.

5.1.2. Inter-Quartile Range. Range discussed above is based on two extreme values and it fails to take account of the scatteredness within range. This defect can be avoided by adopting the measure of inter-quartile range. *Inter-quartile range is the difference between upper quartile and lower quartile.* It divides the series into four equal parts and measures the dispersion between the third quartile and the first quartile. It ignores first quarter (25% observations) and the last quarter (25%) of the observations.

5.2. Methods of Averaging Deviations

5.2.1. Semi-inter-quartile Range (or Quartile Deviation). *Semi-inter-quartile Range is also called as quartile deviation. Half of the difference between upper quartile (Q_3) and lower quartile is called as quartile deviation..*

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

Unlike range, it is not affected by extreme values. In other respects, it has the same characteristics as those of range. It is an absolute measure of dispersion. **Since it does not take into account all the items of the series, it is called the method of limits.**

Coefficient of Quartile Deviation. Quartile deviation is an absolute measure of dispersion. *The relative measure corresponding to quartile deviation is called the coefficient of quartile deviation.* It is expressed symbolically as

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

Coefficient of quartile deviation is used for the purpose of comparing degree of dispersion in different series.

The following illustrations would make clear the computations of quartile deviation and its coefficient in different type of series, viz., individual series, discrete series and continuous series.

$2 \qquad 1043.75 + 915 \qquad 1958.75$

Merits and Demerits of Quartile Deviation

Merits

(i) Like range, quartile deviation is easy to calculate and simple to understand.

(ii) It is less affected by extreme values.

(iii) It can be computed in open end classes.

(iv) It is superior and more reliable than range.

(v) It is useful when dispersion of middle 50% is to be calculated.

Demerits

(i) It ignores 50% items i.e., the first 25% and the last 25% and hence is not based on all observations.

(ii) It is not suited to algebraic treatment.

(iii) It is very much affected by sampling fluctuations.

5.2.2. Mean Deviation or Average Deviation. The above methods, namely range and quartile deviation, are not measures of dispersion in true sense as they are based on two values of a series. In case of range, two extreme values are taken and in case of quartile deviation Q_1 and Q_3 are taken. Other values are ignored. Moreover, they do not show the scatteredness around an average or mean. Mean deviation (or average deviation) overcomes both these drawbacks.

Mean deviation of a series is the arithmetic average of the deviations of various items from their central tendency, i.e., mean, median or mode. Mode is not generally considered, as its value is sometimes indeterminate. Theoretically, median is preferred because the sum of the deviations of the items from median is minimum when ± signs are ignored. All deviations are treated as positive. A simple reason for ignoring ± signs is that if we do not ignore the signs of the deviations then mean deviation about mean is zero. It is because of the property of arithmetic mean, that is, *'Algebraic sum of the deviations of a given set of observations from their mean is always zero'*. Generally mean is used for the calculation of mean deviation. So, it is called mean deviation. Mean deviation is also called **average deviation**.

Computation of Mean Deviation

The following steps are involved in the computation of mean deviation:

(i) Compute mean or median of the series.

(ii) Take deviations d = X — A from either mean or median ignoring ± signs.

(iii) Obtain total of these deviations, i.e., Σ |D| where two parallel lines (||) indicate that absolute values are taken by ignoring + signs.

(iv) Divide the above total, i.e., in step (iii) by the number of items, i.e.,

$$\text{M.D.} = \frac{\Sigma|D|}{N}$$

Coefficient of Mean Deviation. Mean deviation is an absolute measure of dispersion. *The relative measure corresponding to the mean deviation is called the coefficient of mean deviation.* It is obtained by dividing mean deviation by the particular average.

Symbolically,

$$\text{Coefficient of M.D.} = \frac{\text{Mean Deviation}}{\text{Median or Mean}}$$

If mean is used while calculating mean deviation, coefficient of mean deviation shall be obtained by dividing the sum total of mean deviation by mean. But, if deviations are taken from median, it will be divided by median.

Merits and Demerits of Mean Deviation

Merits

(i) Mean deviation is simple to understand and easy to compute.

(ii) It is rigidly defined and its value is precise and definite.

(iii) It is based on all observations of the series and not on limits like range and quartile deviation.

(iv) It is less affected by the extreme values than the standard deviation.

(v) Comparison about formation of different distributions can easily be made as deviations are taken from a central value.

Demerits

(i) Ignoring ± signs is bad from the viewpoint of mathematics.

(ii) Mean deviation is a non-algebraic method.

(iii) If mean, mode and median are in fractions, then calculation of mean deviation becomes difficult.

(iv) Mean deviation when calculated from mode is not reliable because in many cases mode is indeterminate.

5.2.3. Standard Deviation. Standard deviation measure of dispersion was first used by **Karl Pearson** in 1893. It is by far the best and widely used measure of dispersion. It is useful because it is free from those defects from which the earlier measures of dispersion suffer. It satisfies most of the properties of a good measure of dispersion.

Standard deviation is the square root of the arithmetic average of the squares of all the deviations taken from mean. That is why, it is shortly known as **root mean square deviation**. It is denoted by the small Greek letter σ (read as sigma).

Two Main Differences between Mean Deviation and Standard Deviation:

(i) While calculating mean deviation algebraic signs (i.e. $\pm$) are not taken into account whereas while calculating standard deviation these signs are taken into account.

(ii) Mean deviation is computed either from mean or median or mode whereas standard deviation is always computed from mean.

5.2.4. Other Measures based on Standard Deviation

1. **Coefficient of Variation (CV).** It is a relative measure of variation. It was developed by Karl **Pearson** and is most commonly used to measure relative variation of two or more series. It shows the relationship between the standard deviation (σ) and the arithmetic mean ($\overline{X}$) expressed in terms of percentage. To calculate coefficient of variation, standard deviation of the series is divided by mean of the series and multiplied by 100. It can be calculated with the help of the following formula :

$$\text{Coefficient of variation} = \frac{\sigma}{\overline{X}} \times 100$$

This measure is used to compare **uniformity, consistency** and **variability** in two different series. The series having greater coefficient of variation, it is called to be less uniform or less consistent.

2. **Variance.** Variance is another measure based on standard deviation. The term variance was first used by **R.A. Fisher** in 1918. Square of the standard deviation is known as variance, i.e.,

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

Mathematical Properties of Standard Deviation

1. **The sum of the squares of the deviations of items from the arithmetic mean is minimum.**

In other words, the sum of the squares of the deviations of items of a series from a value other than arithmetic average would always be greater. That is why, standard deviation is computed from the arithmetic mean. The following illustration makes this clear

Example 26. A= 4

It is clear from the above example that, sum of the squares of deviations from mean (40) is less than the sum of squares of deviations (60) taken from assumed mean.

2. Like the arithmetic mean, it is possible to compute combined standard deviations of two or more groups. The required formula is as follows:

$$\sigma_{12} = \sqrt{\frac{N_1\left(\sigma_1^2 + D_1^2\right) + N_2\left(\sigma_2^2 + D_2^2\right)}{N_1 + N_2}}$$

where σ_{12} = Combined standard deviation of two groups.

σ_1 = Standard deviation of first group.

σ_2 = Standard deviation of second group.

$D_1 = \overline{X}_1 - \overline{X}_{1,2}$

$D_2 = \overline{X}_2 - \overline{X}_{1,2}$

The above formula can be extended for three or more groups.

Example 27. Two samples of sizes 40 and 50 respectively have the same mean as 53, but different standard deviations as 19 and 8 respectively. Find the standard deviation of the combined sample of size 90.

Solution:

Given: $N_1 = 40; \overline{X}_1 = 53; \sigma_1 = 19$

$N_2 = 50; \overline{X}_2 = 53; \sigma_2 = 8$

Now **Combined Mean** $= \dfrac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$

$$= \frac{40 \times 53 + 50 \times 53}{40 + 50} = \frac{2120 + 2650}{90} = \frac{4770}{90} = 53$$

$$\sigma_{1,2} = \sqrt{\frac{N_1(\sigma_1^2 + D_1^2) + N_2(\sigma_2^2 + D_2^2)}{N_1 + N_2}} = \sqrt{\frac{40(19^2 + 0^2) + 50(8^2 + 0^2)}{40 + 50}}$$

$$= \sqrt{\frac{14440 + 3200}{90}} = \sqrt{\frac{17640}{90}} = \sqrt{196} = 14 \quad \textbf{Ans.}$$

3. Standard deviation is independent of change of origin, thus the value of standard deviation remains the same if a constant is either added or subtracted in a series to all observation.

4. Standard deviation is affected by change of scale – Thus f all the observations in a series are multipied or divided by a constant then the standard deviation also gets multiplied or divided by this same constant.

Merits and Demerits of Standard Deviation

Merits

(i) Standard deviation is based on all observations. Therefore, a change in even one value affects the value of standard deviation.

(ii) It is rigidly defined.

(iii) It is amenable to further algebraic treatment and possesses many mathematical problems.

(iv) It is less affected by fluctuations of sampling than most other measures of dispersion.

(v) For comparing variability of two or more series, coefficient of variation is considered an ideal measure which is based on standard deviation and mean.

Demerits

(i) Standard deviation is not easy to understand and simple to calculate as compared to other measures of dispersion.

(ii) It gives more importance to extreme values and less to those which are nearer to mean. It is because of this fact that squares of big deviations would be proportionately greater than the squares of small deviations.

5.3. Graphic Method or Loreng Curve

Dispersion can also be measured with the help of Lorenz Curve. It was devised by a famous economic statistician **Max O. Lorenz**. He used this curve to measure the distribution of income and wealth. Now a days, this curve is also used to study the distribution of profit, wage, turnover, production, population etc.

Lorenz Curve is a cumulative percentage curve in which the periods of time is combined with the percentage of other things such as wealth, income, turnover, etc. While drawing it, the following procedure is adopted:

(i) Firstly, size of the items and frequencies are both cumulated.

(ii) Then percentages are obtained for these cumulative values.

(iii) These percentages are to be plotted on a graph. The percentages of cumulated frequencies are plotted on x-axis and the percentages of the cumulated values are plotted on y-axis.

(iv) On both axis, we start from 0 to 100.

(v) Draw the diagonal line $y = x$ joining the origin O (0, 0) with the point P (100, 100) as shown in the diagram. The OP line will make an angle of 45° with the x-axis, which shows equal distribution. Hence, this line is called the **line of equal distribution.**

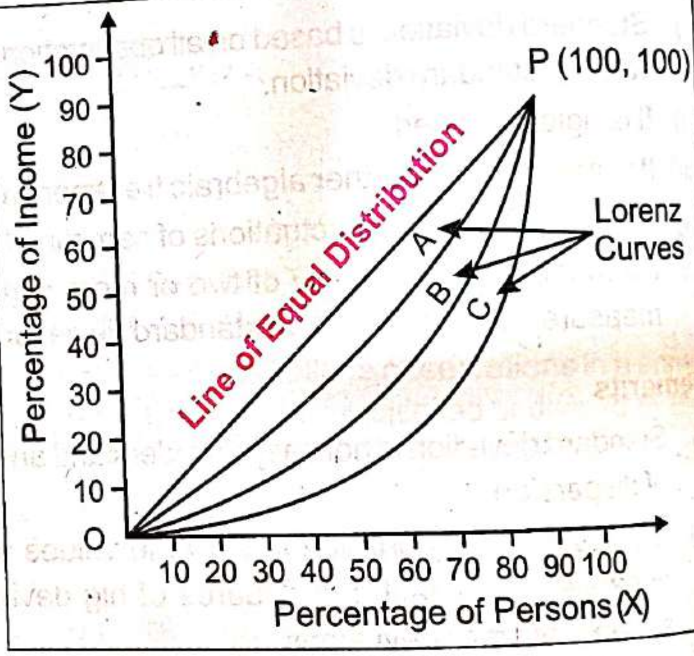

Fig. 9.2

(vi) Now, plot the percentages of the cumulated values of the variable (y) against the percentages of the corresponding cumulated frequencies (x) for the given distribution and join these points with a smooth free hand curve. The curve so obtained shows actual distribution and therefore, is called Lorenz Curve.

The Lorenz curve will always lie below line of equal distribution OP, unless the distribution is uniform. If the distribution is uniform, the Lorenz Curve will coincide with the line of equal distribution. **The more Lorenz Curve is away from the line of equal distribution, the greater the dispersion.** (See Figure 9.2.)

In the figure, OAP indicates a less-degree of variability as compared to points lying on OBP. Variability is still greater when the points lie on the curve OCP. Thus, the measure of dispersion is provided by the distance of the Lorenz Curve from the line of equal distribution.