LECTURE 23

MEASURES OF DISPERSION

PROF. SHRUTI

Introduction

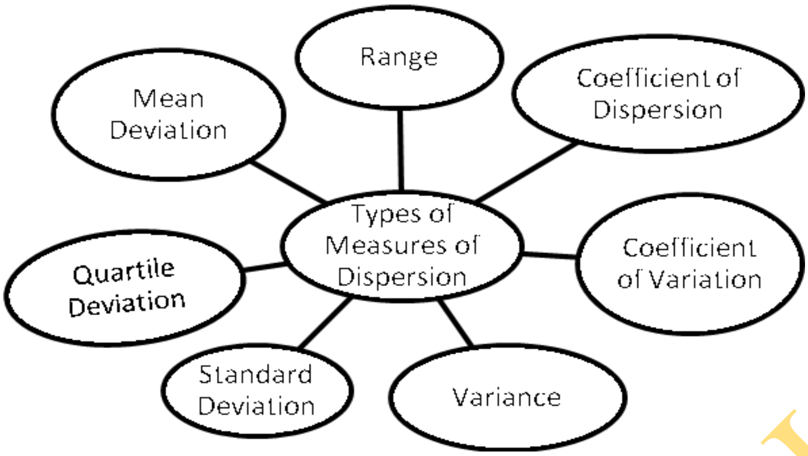
Averages fail to reveal the full details of the distribution. Two or three distributions may have the same average but still they may differ from each other in many ways. In such cases, rather statistical analysis of the data is necessary so that these differences between various series can be studied and accounted for such analysis will make our results more accurate and we shall be more confident of our conclusions.

Suppose, there are three series of nine items each as follows:

Series A	Series B	Series C
40	36	1
40	37	9
40	38	20
40	39	30
40	40	40
40	41	50
40	42	60
40	43	70
40	44	80
Total 360	360	360

Mean 40	40	40

In the first series, the mean is 40 and the value of all the items is identical. The items are not at all scattered, and the mean fully discloses the characteristics of this distribution. However, in the second case, though the mean is 40 yet all the items of the series have different values. But the items are not very much scattered as the minimum value of the series is 36 and the maximum is 44 in the range. In this case also, mean is a good representative of the series because the difference between the mean and other items is not very significant. In the third series also, the mean is 40 and the values of different items are also different, but here the values are very widely scattered and the mean is 40 times of the smallest value of the series and half of the maximum value. Though the mean is the same in all the three series, yet the series differ widely from each other in their formation. Obviously, the average does not satisfactorily represent the individual items in this group and to know about the series completely, further analysis is essential. The scatter among the items in the first case is nil, in the second case it varies within a small range, while in the third case the values range between a very big span and they are widely scattered. It is evident from the above, that a study of the extent of the scatter around average should also be made to throw more light on the composition of a series. **The name given to this scatter is dispersion.**



10.2. Definition

Some important definitions of dispersion are given below:

(i) "Dispersion or spread is the degree of the scatter or variation of the variable about a central value." – *Brooks and Dick*

(ii) "Dispersion is the measure of the variations of the items."

– *A.L. Bowley*

(iii) "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data." – *Spiegel*

From the above definitions, it is clear that in a general sense the term dispersion refers to the variability in the size of items. If the variation is substantial, dispersion is said to be considerable and if the variation is very little, dispersion is insignificant.

Usually, in a precise study of dispersion the deviations of size of items from a measure of central tendency are found out and then these deviations are averaged to give a single figure representing the dispersion of the series. This figure can be compared with similar figures representing other series. Such comparisons give a better idea about the formation of series than a mere comparison of their averages.

Averages of second order : For a precise study of dispersion, we have to average deviations of the values of the various items, from their average. We have seen earlier that arithmetic mean, median, mode, geometric mean and harmonic mean, etc., are all averages of the first order. Since in the calculation of measures of dispersion, the average values are derived by the use of the averages of the first order, the measures of dispersion are called averages of the second order.

12.3. Objectives of Measuring Dispersion

Measures of variations are calculated to serve the following purposes:

(i) To judge the reliability of measures of central tendency.

(ii) To make a comparative study of the variability of two series.

(iii) To identify the causes of variability with a view to control it.

Spur and **Bonimi** have very rightly observed that, "in matters of health, variations in body temperature, pulse beats and blood pressure are 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 programs."

In Social Sciences where we have to study problems relating to inequality in income and wealth, measures of dispersion are of immense help.

(iv) To serve as a basis for further statistical analysis.

10.4. Characteristics of A Good Measure of Dispersion

The properties of a good measure of dispersion are the same as the properties of a good measure of central tendency. Precisely, they are:

(i) It should be rigidly defined.

(ii) It should be based on all the observations of the series.

(iii) It should be capable of further algebraic treatment.

(iv) It should be easy to calculate and simple to follow.

(v) It should not be affected by fluctuations of sampling.

10.5. Different Measures of Dispersion

Absolute and relative "dispersion : Dispersion or variation can be expressed either in terms of the original units of a series or as an abstract figure like a ratio or percentage. If we calculate dispersion of a series relating to the income of a group of persons in absolute figures, it will have to be expressed in the unit in which the original data are, say, rupees. Thus, we can say that the income of a group of persons is Rs. 5000 per month and the dispersion is Rs. 500. This is called Absolute Dispersion. If, on the other hand, dispersion is measured as a percentage or ratio of the average, it is called Relative Dispersion. Since the relative dispersion is a ratio, it has no units. In the above case, the average income would be referred to as Rs. 5000 per month and the relative dispersion $\frac{500}{5000} = 0.1$ or 10%. In a comparison of the variability of two or more series, it is the relative dispersion that has to be taken into account as the absolute dispersion may be erroneous or unfit for comparison if the series are originally in different units.

Range

Range is the simplest possible measure of dispersion. It is the difference between the values of the extreme items of a series. Thus, if in a series relating to the weight measurements of a group of students, the lightest student has a weight of 40 kg. and the heaviest, of 110 kg. The value of range would be 110 — 40 = 70 kg. This figure indicates the variability in the weight of students.

Symbolically,

$$\text{Range (R)} = L - S$$

where, L is the largest value and S the smallest value in a series.

Range as calculated above is an absolute measure of dispersion which is unfit for purposes of comparison if the distributions are in different units. For example, the range of the weights of students cannot be compared with the range of their height measurements as the range of weights would be in kg. and that of heights in centimetres. Sometimes, for purpose of comparison, a relative measure of range is calculated. If range is divided by the sum of the extreme items, the resulting figure is called "The Coefficient of the Range" or "The Coefficient of the Scatter."

Symbolically,

The Ratio of Range or the Coefficient of the scatter (or Range)

$$= \frac{\text{Max. value} - \text{Min. value}}{\text{Max. value} + \text{Min. value}} = \frac{L - S}{L + S}$$

$$= \frac{\text{Absolute range}}{\text{Sum of the extreme values}}$$

The following illustration would illustrates the use of the above formulae

Example 1. The profits of a company for the last 8 years are given below. Calculate the Range and its Coefficient:

Year	1975	1976	1977	1978	1979	1980	1981	1982
Profits (in '000 Rs.)	40	30	80	100	120	90	200	230

Solution.

Here, L = 230 and S = 30

Range = L – S = 230 – 30 = 200

Coefficient of Range = $\frac{L-S}{L+S}$ or $\frac{230-30}{230+30}$ or $\frac{200}{260}=0.77$

Example 2. Calculate Co-efficient of Range from the following data :

Weekly Wages (Rs.)	No. of Labourers
50–60	50
60–70	45
70–80	45
80–90	40
90–100	35
100–110	30
110–120	30

Solution.

Coefficient of Range (first method) = $\frac{L-S}{L+S}$

Here, L = 120 and S = 50. such the

Coefficient of Range = $\frac{120-50}{120+50}=\frac{70}{170}=0.41$

Coefficient of Range (second method) = $\frac{L-S}{L+S}$ Here, L = 115 and S = 55

Co-efficient of Range = $\frac{115-55}{115+55}=\frac{60}{70}=0.35$

Example 3. Find the Range and the Co-efficient of range for the following observations 65, 70, 59, 81, 76,57,60,55, and 50.

Solution. Highest value = 82

Lowest value = 50

Range = 82 – 50 = 32

$$\text{Coefficient of Range} = \frac{82-50}{82+50} = \frac{32}{132} = 0.2424$$

Merits and Demerits of Range

As has been pointed out that' a good measure of dispersion should be rigidly defined, easily calculated, readily understood, should be capable of further mathematical treatment and should not be much affected by fluctuations of sampling.

Out of these the only merit possessed by Range is that it is easily calculated and readily understood.

As against this, the Range as a measure of dispersion has the following demerits :

1. It is affected greatly by fluctuation of Sampling.
2. It is not based on all the observations of the series.
3. It cannot be used in case of open distributions.

Uses of Range

With all its limitations Range is commonly used in certain fields. For example:

1. Quality Control.
2. Variation in Money Sales, Share values, Exchange Rates and Gold Prices, etc.
3. Weather forecasting.

Inter-Quartile Range

Third quartile (Q_3)

$$IR = Q_3 - Q_1$$

Inter-quartile range   First quartile (Q_1)

Just as in a case of range the difference of extreme items is found, similarly, if the difference in the values of two quartiles is calculated, it would give us what is called the Inter-Quartile Range. Inter-Quartile range is also a measure of dispersion. It has an advantage over range, in as much as, it is not affected by the values of the extreme items. In fact, 50% of the values of a variable are between the two quartiles and as such the inter-quartile range gives a fair measure of variability. However, the inter-quartile range suffers from the same defects from which range suffers. It is also affected by fluctuations of sampling and IS not based on all the observations of a series.

SEMI-INTER-QUARTILE RANGE

Semi-inter-quartile range, as the name suggests is the midpoint of the inter-quartile range. In other words, it is one-half of the difference between the third quartile and the first quartile. Symbolically,

$$\text{Semi-inter-quartile range or quartile deviation} = \frac{Q_3 - Q_1}{2}$$

where Q_3 and Q_1 are the upper and lower quartiles respectively.

In a symmetrical series median lies halfway on the scale form Q_1 to Q_3. In a symmetrical distribution Q_3 – median = Median – Q_1 or median = $\frac{Q_3 - Q_1}{2} = Q_3 - \left(\frac{Q_3 - Q_1}{2}\right) = Q_1 + \left(\frac{Q_3 - Q_1}{2}\right)$. If, therefore, the value of the quartile deviation is added to the lower quartile or subtracted from the upper quartile in a symmetrical series, the resulting figure would be the value of the median. But, generally series are not symmetrical and in a moderately asymmetrical series Q_1 + quartile deviation or Q_3 – quartile deviation would not give the value of the median. There would be a difference between the two figures and the greater the difference, the greater would be the extent of departure from normality.

Quartile deviation is an absolute measure of dispersion. If it is divided by the average value of the two quartiles, a relative measure of dispersion is obtained. It is called the Co-efficient of Quartile deviation.

$$\text{Co-efficient of a quartile deviation} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$