Measures of Dispersion

A measure of variation or dispersion is one that measures the extent to which there are differences between individual observation and some central or average value.

The various measures of central value give us one single figure that represents the entire data. But the average alone cannot adequately describe a set of observations, unless all the observations are the same. Measures of central tendency are not fully representatives of the data. It is necessary to describe the variability or dispersion of the observations. In two or more distributions the central value may be the same but still there can be wide disparities in the formation of distribution. Measures of dispersion help us in studying this important characteristic of a distribution.

Definitions of dispersion

1. "Dispersion is the measure of the variation of the items." -A.L. Bowley

2. "The degree to which numerical data tend to spread about an average value is called the variation of dispersion of the data." –Spiegel

3. Dispersion or spread is the degree of the scatter or variation of the variable about a central value." -- Brooks & Dick

4. "The measurement of the scatterness of the mass of figures in a series about an average is called measure of variation or dispersion." -Simpson & Kafka.

It is clear from above that dispersion measures the extent to which the items vary from some central value. Since measures of dispersion give an average of the differences of various items from an average, they are also called averages of the second order. An average is more meaningful when it is examined in the light of dispersion.

For example, if the average wage of the workers of factory A is Rs. 3885 and that of factory B Rs. 3900, we cannot necessarily conclude that the workers of factory B are better off because in factory B there may be much greater dispersion in the distribution of wages. The study of dispersion is of great significance in practice as could well be appreciated from the following example:

	Series A	Series B	Series C
	100	100	1
	100	105	489
	100	102	2
	100	103	3
	100	90	5
Total	500	500	500
$\bar{X}$ (Mean)	100	100	100

Since arithmetic mean is the same in all three series, one is likely to conclude that these series are alike in nature. But a close examination shall reveal that distributions differ widely from one another. In series A, each and every item is perfectly represented by the arithmetic mean or in other words none of the items of series A deviates from the arithmetic mean and hence there is no dispersion. In series B, only one item is perfectly represented by the arithmetic mean and the other items vary but the variation is very small as compared to series C. In series C. not a single item is represented by the arithmetic mean and the items vary widely from one another. In series C, dispersion is much greater compared to series B. Similarly, we may have two groups of labourers with the same mean salary and yet their distributions may differ widely. The mean salary may not be so important a characteristic as the variation of the items from the mean.

Significance of Measures of Dispersion

Measures of variation are needed for four basic purposes:

1. Measures of variation point out as to how far an average is representative of the mass. When dispersion 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 a good estimate of the average in the corresponding universe. On the other hand, when dispersion is large, the average is not so typical, and unless the sample is very large, the average may be quite unreliable.

2. Another purpose of measuring dispersion is to determine nature and cause of variation in order to control the variation itself. In matters of health variations in body temperature, pulse beat and blood 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 is basic to the control of causes of variation. In social sciences a special problem requiring the measurement of variability is the measurement of "inequality" of the distribution of income or wealth etc.

3. Measures of dispersion enable a comparison to be made of two or more series with regard to their variability. The study of variation may also be looked 60 upon as a means of determining uniformity of consistency. A high degree of variation would mean little uniformity or consistency whereas a low degree of variation would mean great uniformity or consistency.

4. Many powerful analytical tools in statistics such as correlation analysis. the testing of hypothesis, analysis of variance, the statistical quality control, regression analysis is based on measures of variation of one kind or another.

Properties of a good measure of dispersion

1. It should be simple to understand.

2. It should be easy to compute.

3. It should be rigidly defined.

4. It should be based on each and every item of the distribution.

5. It should be suitable for further algebraic treatment.

6. It should have sampling stability.

7. It should not be unduly affected by extreme values.

Various Measures of Dispersion

1. Range:

Range is the simplest measure of dispersion is, which is the difference between the maximum value and the minimum value of data.i.e.,

R = L-S , where, L= Largest value, and S= smallest value

e.g., Find the range for the following three sets of data:

Set 1: 05 15 15 05 15 05 15 15 15 15

Set 2: 8 7 15 11 12 5 13 11 15 9

Set 3: 5 5 5 5 5 5 5 5 5 5

In each of these three sets, the highest number is 15 and the lowest number is 5. Since the range is the difference between the maximum value and the minimum value of the data, it is 10 in each case. But the range fails to give any idea about the dispersal or spread of the series between the highest and the lowest value. This becomes evident from the above data.

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

e.g.,

Size of Items	Frequency
20-40	7
40-60	11
60-80	30
80-100	15
100-120	5

Here, the upper limit of the highest class is 120 and the lower limit of the lowest class is 20. Hence, the range is 120 - 20 = 100.

Coefficient of Range: It is a relative value of dispersion which is used to compare two or more than two sets of data. It is obtained as follow:

Coefficient of Range =L-S /L+S

Example: Calculate the coefficient of range separately for the two sets of data given below:

Set 1: 8 10 20 9 15 10 13 28

Set 2: 30 35 42 50 32 49 39 33

Limitations of range

There are some limitations of range, which are as follows:

1. It is based only on two items and does not cover all the items in a distribution.

2. It is subject to wide fluctuations from sample to sample based on the same population.

3. It fails to give any idea about the pattern of distribution.

4. Finally, in the case of open-ended distributions, it is not possible to compute the range.

Despite these limitations of the range, it is mainly used in situations where one wants to quickly have some idea of the variability or' a set of data. When the sample size is very small, the range is considered quite adequate measure of the variability. Thus, it is widely used in quality control where a continuous check on the variability of raw materials or finished products is needed. The range is also a suitable measure in weather forecast. The meteorological department uses the range by giving the maximum and the minimum

temperatures. This information is quite useful to the common man, as he can know the extent of possible variation in the temperature on a particular day.

2. Inter Quartile Range or Quartile Deviation (IQR or QD)

The interquartile range or the quartile deviation is a better measure of variation in a distribution than the range. Here, avoiding the 25 percent of the distribution at both the ends uses the middle 50 percent of the distribution. In other words, the interquartile range denotes the difference between the third quartile and the first quartile. Symbolically,

Interquartile range = $Q_3 - Q_1$

Many times the interquartile range is reduced in the form of semi-interquartile range or quartile deviation as shown below:

Semi interquartile range or Quartile deviation = $Q_3 - Q_1 / 2$

When quartile deviation is small, it means that there is a small deviation in the central 50 percent items. In contrast, if the quartile deviation is high, it shows that the central 50 percent items have a large variation.

In a symmetrical distribution, the two quartiles, that is, Q3 and QI are equidistant from the median. Symbolically, M-QI = Q3-M

However, this is seldom the case as most of the business and economic data are asymmetrical. But, one can assume that approximately 50 percent of the observations are contained in the interquartile range. It may be noted that interquartile range or the quartile deviation is an absolute measure of dispersion. It can be changed into a relative measure of dispersion as follows:

Coefficient of Q D = Q3 –Q1/ Q3 +Q1

Merits of Quartile Deviation

The following merits are entertained by quartile deviation:

1. As compared to range, it is considered a superior measure of dispersion.

2. In the case of open-ended distribution, it is quite suitable.

3. Since it is not influenced by the extreme values in a distribution, it is particularly suitable in highly skewed or erratic distributions.

Limitations of Quartile Deviation

1. Like the range, it fails to cover all the items in a distribution.

2. It is not amenable to mathematical manipulation.

3. It varies widely from sample to sample based on the same population.

4. Since it is a positional average, it is not considered as a measure of dispersion. It merely shows a distance on scale and not a scatter around an average.

In view of the above-mentioned limitations, the interquartile range or the quartile deviation has a limited practical utility.