1. Introduction

Welcome to the series of e-learning modules on Variation and Dispersion. In this module we are going to cover the definition, types, measures and the merits and demerits of dispersion.

By the end of this session, you will be able to understand:

- The concepts of Variation and Dispersion
- Various types and measures of Variation and Dispersion
- Significance of Variation and Dispersion

To begin with, we shall look at what Descriptive Statistics is.
Data collection is an easy and fun-filled activity, but conveying what is found in the data is a difficult job. So, we use statistics to describe the data.
This branch of statistics is called *Descriptive Statistics* as it helps in describing the data collected.

Now, while describing the data we can use single score numbers that represent the entire data or we can use scores that will tell us the spread of the data.
To represent the data in single scores we use the measures of central tendency and to understand the spread of the scores we use the measures of dispersion or variation.

The various measures of central tendency gives us one single figure that represents the entire data. The averages alone cannot adequately describe a set of observations, unless all the observations are alike. We also need to make a note that at times the data set may have the same central values but contain a lot of disparities in the formation of the distributions.

Let us take an example to understand the concept better. Here is a table containing three sets of data:
Set A contains the numbers 2, 5, 17, 17 and 44,
Set B – 17,17,17,17,17 and
Set C contains 13, 14, 17, 17 and 24.
Now note that, in Set B all observations are equal. In Set A and C the observations are dispersed and in all the three sets the mean and more are the same.

Definitely we need another measure, concept and technique which will account for such dispersion of data and which will help us learn to deal with the body of data in regard to their distribution over the range of variation of the variable.

Let us take another example which will help us appreciating the significance of the measures of variation. The data in the table here pertains to the wages of five workers in three different factories.

Figure 1

FACTORY A (Wages)	FACTORY B (Wages)	FACTORY C (Wages)
300	310	380
300	300	210
300	304	220
300	306	200
300	280	490
Total 1500	1500	1500
X 300	300	300

Wage of the workers in Factory A is uniform for all the workers – 300. In case of factory B, the wages given are 310, 300, 304, 306 and 280 while in factory C the wages are given as 380, 210, 220, 200 and 490.

In the table, observe that, Average wage is the same in all factories.

Figure 2

Factory A **No Variation**	**The wage of every worker is the same as the mean**

Factory B **Little Variation**	**Only one workers wage represents the mean, the rest all vary from the average with small variations**

Factory C **Wide Variation**	**Mean does not represent the workers wage and all individual wages differ from the mean**

Factory A there is no variation in wages. The wage of every worker is the same as the mean. The figures in Factory B shows a little variation where only one workers wage represents the mean, the rest all vary from the average with small variations.
And, in Factory C, we see a wide variation. The mean does not represent the workers wage and all individual wages differ from the mean.

Hence, in such data it is not important to look at the mean of the income but is important to know how this income is distributed. We need to know whether a large number of the workers are receiving the mean income or are there a few with enormous income and millions who are below the mean income.

2. Characteristics of Dispersion

Now, we shall look at the characteristics of dispersion.

- There could be two distributions with the same mean but different dispersions
- There could be two distributions with the same dispersion but with unequal means (X1 and X2)
- There could be two dispersions with unequal means and unequal dispersions

These characteristics of the data bring in a need for a measure to support and supplement the data so as to state the extent to which the individual measures differ on an average from the mean.

We will now look at a couple of definitions for dispersion.
“The measurement of the scatterness of the mass of figures in a series about an average is called the measure of variation of dispersion”- Simpson and Kafka.
According to Spiegel, "The degree to which numerical data tend to spread about an average value is called the variation or dispersion of the data."

Literally, dispersion means deviation. It is the difference or spread of certain values from their central value.
In relation to statistical series, it means, deviations of various items of the series from its central value.
The word ‘measure’ refers to a method of measuring certain values.

Thus, the phrase ‘measure of dispersion’ refers to various possible methods of measuring the dispersions of different values from an average value or any other extreme value.

Here, the concern is in the amount or degree of variation not its direction.

Now, why do we need the measure of variation?

1. To determine the reliability of an average
2. To serve as a basis for the control of variability
3. To compare two or more series with regards to their variability
4. To facilitate the use of other statistical measures
5. To establish the trends in the time series

Let us understand the importance of these points in detail;

To determine the reliability of an average:

Here, we try to understand the extent of how an average is a representative of the data.

If the average is closely representing the individual value, then it is reliable in the sense that it is a good estimate of the average.

On the other hand when the dispersion of the data is large from the average then the average is not a quite reliable one.

To serve as a basis for the control of variability:

Here, we try to determine the nature and cause of the variation in order to control the variation itself.

For example, In health matters, we need to know the variation in body temperature, pulse or blood pressure to help in guiding the treatment to be given to control the variation.

To compare two or more series with regard to their variability:

Here two or more data is compared with regard to their variability.

It helps in determining the uniformity or consistency of the data.

A high degree of variation means a little uniformity or consistency, whereas, a low degree of variation means greater uniformity or consistency.

To facilitate the use of other statistical measures:

Here we understand that the measure of variation of a data is the base for using other statistical tools such as correlation analysis, testing of hypothesis, analysis of fluctuations, techniques of production control, cost control and so on.

To establish trends in the time Series:

In time series we remove cyclical, seasonal or random fluctuations, which we form after studying the central values.

After understanding the role of the measures of variation let us look at the properties needed

by the measure of variation, to be defined as a good measure of variation.

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

- It should be simple to understand
- It should be easy to compute
- It should be rigidly defined
- It should be based on each and every item of the distribution
- It should be amenable to further algebraic treatment
- It should have sampling stability
- It should not be unduly affected by the extreme items

3. Measure of Dispersion

For the study of dispersion, we need some measures which show whether the dispersion is small or large. There are two types of measure of dispersion which are:

Absolute Measure of Dispersion and
Relative Measure of Dispersion

An absolute measure of variation is expressed in the same statistical unit in which the original data is given. That is, When the observations are in kilograms, the absolute measure is also in kilograms.
These measures are used to compare the variation of any two sets of data that are expressed in the same units and have almost the same average size.

For example, Say, there are two sets of data - salary of managers versus wages of workers.

They are not of the same unit and hence the measure of dispersion is not comparable.

Absolute measures give us an idea about the amount of dispersion in a set of observations.

The absolute measures which are commonly used are: Range, Quartile, Inter Quartile Range, Mean Deviation, Standard Deviation and Lorenz Curve.

Absolute Measure: Range
Range is the simplest method of studying variation.
It is defined as the difference between the value of the smallest item and the value of the largest item in the distribution.
This is a very useful measure while handling quality control, fluctuations in share prices, weather forecasts and so on.

Absolute Measure: Quartile Deviation
The Quartile Deviation gives the average amount by which the two quartiles differ from the median. In a symmetrical distribution the two quartiles are equidistant from the median and as such the difference can be taken as a measure of variation.

Absolute Measure: Inter Quartile Deviation
The Inter Quartile Range is the range which includes the middle 50 percent of the distribution. It represents the difference between the third quartile and the first quartile.

Absolute Measure: Mean Deviation
The Mean Deviation is also known as the average deviation and helps in studying the formation and scatterness of the observation by taking the deviations from the average. The average deviation is obtained from the median and is taken as absolute measures.The signs are ignored as we are not concerned about the direction of the data.

Mean deviation treats every single observation with equal weight, in the form of arithmetic mean of deviations based on each observation. It is most popularly used in small samples

that do not need any elaborate analysis.

Absolute Measure: Standard Deviation and Variance

The Standard deviation and Variance is a concept introduced by Karl Pearson. It is the most widely used measure of variation. It also known as root mean square deviation it measures the absolute dispersion or variability of a distribution. Variance is a measure which suitably combines individual deviations from the mean, treating each observation with equal weight as in mean deviation.

For variance, however, measure of individual deviation is taken as the squared difference from the mean. Since it is more manageable to use the squared difference rather than absolute difference, particularly while doing formal mathematics, use of variance has become more popular

Absolute Measure: Lorenz Curve

Devised by Max O. Lorenz, a famous economic statistician, this is a graphic method of studying the dispersion.

The most common use of this curve is to study the inequality in the distribution of income and wealth between countries and between time periods.

4. Relative Measure

Relative Measure: Introduction
A ‘measure of relative variation’ is the ratio of a measure of absolute variation to an average. These measures are calculated for the comparison of dispersion in two or more than two sets of observations.
These measures are free of the units in which the original data is measured.

For instance, if the original data is in dollars or kilometers, we do not use these units with relative measure of dispersion.
These measures are a sort of ratio and are called coefficients. Each absolute measure of dispersion can be converted into its relative measure.

Thus, the relative measures of dispersion are:

- Coefficient of Range
- Coefficient of Quartile
- Coefficient of Mean Deviation
- Coefficient of Standard Deviation
- Coefficient of Variation

Relative Measure: Coefficient of Range or Coefficient of Dispersion
This is the relative measure corresponding to the absolute measure range.
It indicates that the distribution with a smaller range has less dispersion.
It is calculated by taking the differences of the range and dividing it by the sum of the range.

Relative Measure: Coefficient of Quartile Deviation or Quartile Coefficient of Dispersion
Coefficient of Quartile Deviation is the relative measure corresponding to the absolute measure of quartile deviation. It is used to compare the degree of variation in different distributions.
It is obtained by taking the difference of the upper quartile and the lower quartile and dividing it by the sum of the upper quartile and the lower quartile.

Relative Measure: Coefficient of Mean Deviation / Mean Deviation of Dispersion
The relative measure corresponding to mean deviation is obtained by dividing the mean deviation by the particular average used in computing the mean deviation.

Relative Measure: Coefficient of Standard Deviation or Standard Coefficient of Dispersion
The coefficient of Standard deviation represents the ratio of the standard deviation to the mean, and it is a useful statistic for comparing the degree of variation from one data series to another, even if the means are drastically different from each other.

Relative Measure: Coefficient of Variation - ASpecial case of Standard Coefficient of Dispersion:
Coefficient of Variation is the value of standard deviation when mean is assumed to be equal to 100. It is a pure number and the unit of observations is not mentioned with its value.
It is written in percentage form like 20% or 25%. When its value is 20%, it means that when the mean of the observations is assumed as equal to 100, their standard deviation will be 20.

The coefficient of variation is used to compare the dispersion in different sets of data, particularly the data which differ in their means or differ in their units of measurement.

5. Selecting the Measure of Dispersion

Selecting the Measure of Dispersion to Use

The appropriate selection of the measures of variation or dispersion for a given set of data depends on two parameters :

- Type of data available
- Purpose of investigation

Selecting the measure of Dispersion to use: Type of Data Available

While deciding the kind of measure of variation or dispersion we need to use under this parameter, we need to look at the following factors:

- If the data are few in number or contain extreme values, avoid the standard deviation
- If the data are generally skewed, avoid the mean deviation
- If the data has gaps around the quartiles, avoid quartile deviation
- If there are open ended classes in the given data, the quartile measure of dispersion should be preferred

Selecting the measure of Dispersion to use: The Purpose of Investigation

While considering the measure of variation or dispersion under this parameter we look at the data in the following factors:

- For an elementary treatment of statistical series in which the measure of variability is desired only for itself, any of the three measures namely range deviation or quartile deviation or mean deviation would be acceptable.
- For data undergoing further statistical analysis the measure of variability to be used would be the standard deviation.
- For data having analysis of variability in terms of normal curve of error, standard deviation is used.
- For data where further analysis is done using advanced statistical tools it would be better to use the standard deviation.

We shall now look into the merits of dispersion

- It gives us the picture of center of the data.
- In certain instances, it may be desirable to examine further, on how the other members of the set of data are dispersed about the measure of central tendency.

The demerits of dispersion includes:

- It does not determine modal group
- It does not determine the median of a large grouped distribution

Here is a summary of our learning in this session:

In this session

- We have understood the concept of Variation and Dispersion
- The measures of Variation and Dispersion and
- Significance of Variation and Dispersion