• Introduction to Measures of Central Tendency

Welcome to the series of E-learning modules of Measures of Central Tendency. In this module we are going to cover the basic concept of measures of central tendency, their uses, their classifications and objectives, various types of central tendencies, their advantages and disadvantages.

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

- Explain concept of Central Tendency
- Explain the objectives of of Central Tendency
- Explain its uses
- Explain the different types of central tendency and compare their advantages and disadvantages.

Central Tendency: The measure of central tendency is also called as measures of location or Averages.

You would have noticed several examples of Averages in day today situations for example when student tells you that she has scored 80% in her examinations she is telling you that the average of the marks from all her subjects when converted to a percentage is 80%.

When a cricketer’s statistics are being discussed we say he has scored an average 52 runs in 40 matches.

The weather department measures the rain fall in a certain region over a period of time and then takes a average and reports it. These are few examples wich we come across in our everyday lives

Collecting data and organising them in a proper format is not easy. A large dump of statistical features normally creates confusion and makes it difficult to grasp any important information.

For this purpose a central value is of data is worked out to give a summarized information of data.

It is a single value within the range of data which represents a group of individual values in a simple and concise manner so that the mind can get a quick understanding of the general characteristics of the individual values in a group. Since the value lies in range of the data, it is known as a measure of central tendency.

Thus, a measure of central tendency is a single value that tries to describe a set of data by identifying the central position within that set of data.

Let us take some example, here a boy is thinking about his average marks which he got in his term test. Here the teacher is announcing in the class that the students scored an average of around 60 out of 100 in the test.

The teacher is using only one value 60 to represent a group of data, instead of describing a whole group of data.

Eminent statistician Professor Bowely has defined central tendency as a "statistical constant which enable us to comprehend in a single effort the significance of the whole."

According to Simpson and Kafka,

"A measure of central tendency is a typical value around which other figures congregate".

Averages are useful in several places. They are useful :

- Distribution is explained in a precise manner.
- Ccomparative study of distributions.
- Measuring other statistical measures like dispersion, skewness, kurtosis etc.

~~Now we will discuss the objectives of averages:~~

Objectives of Measures of Central Tendency

- Average is one value which may represent lakhs or millions of values. Thus, it helps in determining the living standards.
- The single value which we get through the measures of central value, is useful for making comparisons both as a point as well as over a period of time.
- Averages are helpful in the formation of policies.
- Averages are considered to be the replica of the universe as averaged depicts the fundamental characteristics of the population.
- Averages are also useful tools in tracing the mathematical relationship between groups, classes, or variables.

Characteristics of a good measure of central tendency are:

- It should be rigidly defined that is everyone finds the same result
- It should be easy to understand and calculate so that its meaning can be made clear even to the common people.
- It is based on all the observations of the data as an average will not be representative if some of the items of the series are excluded from the group.
- It should be suitable for further mathematical treatment.
- It should be least affected by fluctuation of sampling. If there is a difference in the value of average for different sample then it is called fluctuations of sampling.
- It should be least affected by the extreme value

Measures of the central tendency or averages are divided into three categories, they are: Mathematical Averages, Positional Averages, and Commercial Averages.

• Mathematical Averages

Let us first have a look Mathematical Averages. They are further to be classified into 3 categories

Arithmetic mean which is commonly denoted as (AM),

Geometric mean which is commonly denoted as (GM) and

Harmonic mean which is commonly denoted as (HM).

Arithmetic Mean is the average value of all the data in the set.

It is the most important, and most commonly used, measure of central tendency.

The mean is what **common people** call the “average”.

Whereas the **statistician** calls it as “arithmetic mean”.

The arithmetic mean is also known as average. When we talk about an average we usually refer to mean. The mean is simply the sum of the values divided by the total number of items in the set.

$$\bar{x} = \frac{\sum x}{n}$$

The formula for arithmetic mean is:

Let us take a look at an example and calculate the daily expenditure of a family.

The daily expenditure of the family over a week is shown in the table

Daily Expenditure of a family in Rs
150
120
130
160
170
180
130
Total = 1040

On adding the value of daily expenditure we will get a total of Rs 1040.

Here, N = 7 because 7 days of the week.

The total of the observation are 1040, so when we take a average it will be 1040/7 which is equal to 149.

$\bar{x}$ = 1040/7 which is equal to 149

Hence, the average daily expenditure is Rs. 149.

This is the simplest way of calculating the arithmetic mean.

Arithmetic mean is the most commonly used average in practice because of the following advantages:

- It is simple to understand and calculate
- It is affected by the value of each and every item in the series.
- It is relatively stable as it does not fluctuate much when repeated samples are taken
- It is useful for algebraic treatment and is better than geographic mean, harmonic mean, median and mode

However there are some disadvantages with Arithmetic mean some of them are as follows

- The value of arithmetic mean depends on each and every item of the series that is either very small or very large
- The arithmetic mean cannot be used for qualitative analysis as qualitative attributes like beauty, honesty etc. are not measurable in terms of numbers
- The value of arithmetic mean would be effective only if the distribution of the variable is normal
- The value of mean cannot be calculated without making an assumption regarding the size of the class interval of the open end classes.

Geometric Mean is defined at the 'n'th root of the product of 'n' observations of a distribution.

Formula for geometric mean is: G.M = $(x_1 x_2 \ldots\ldots x_n)^{1/n}$

Example: Let us take 6 numbers: 2,3,3,5,3 and 2.

Here, N is 6.

On applying the formula we will get $(2x3x3x5x3x2)^{1/6}$

$$\sqrt[6]{2\times3\times3\times5\times3\times2}$$

$$\sqrt[6]{540}$$

Thus, GM is equal to 2.856.

Advantages of Geometric Mean are:

- Geometric Mean is based on each and every observations in the data set
- It is rigidly defined
- As compared to the arithmetic mean it gives more weight to small values and less weight to large values. Sometimes it may be equal to the arithmetic mean
- It is capable of algebraic manipulation

Disadvantages are:

- As compared to arithmetic mean, it is more difficult to understand
- Both computation of the geometric mean and its and interpretation are difficult
- It cannot be calculated if there are any negative items or zero values

Harmonic Mean: Harmonic Mean is the reciprocal of the arithmetic average of the reciprocal of the values of its various items.

Let us take the same numbers: 2,3,3,5,3 and 2.

$$\frac{6}{\frac{15}{30}+\frac{10}{30}+\frac{10}{30}+\frac{6}{30}+\frac{10}{30}+\frac{15}{30}}$$ Here also N is 6.

On applying the formula we will get

$$\frac{30}{11} \quad \frac{1}{\frac{11}{30}} \quad \frac{6}{\frac{66}{30}}$$

HM = **=** **=**

= 2.73,

Thus HM is 2.73.

Advantages of Harmonic mean are:

- All observations in a series are taken into consideration
- It is amenable to algebraic treatment
- It is most appropriate when greater weight age needs to be given to the small observations and less weight age to the large ones.
- In problems involving time and rates, it provides better results than other averages

Disadvantages of Harmonic mean are:

- It is not easily understood
- Its computation is rather difficult
- It cannot be computed if there is both positive and negative values or when the item is zero
- As it gives the largest weight to the smallest item, it is not appropriate in the analysis of economic data
-

• Positional Averages

Let us now look at another concept Positional Averages

Positional averages determine the positions or place of the central value or variables in the series.

Positional Averages can be broadly classified into 2 categories Median and Mode.

The Median refers to the middle value of a set of data arranged in either ascending or descending order. If there is an even number of entries, the median is defined to be the mean of the two center entries.

It is the value which divides the series into two equal parts. 50% of observations are above the Median and 50% are below it. To find the median, the data points must first be sorted into either ascending or descending numerical order.

$\frac{N+1}{2}$ The *position* of the Median value can then be calculated using the following formula:

Example: In a company five workers earn different amount of wages.

Wages in Rs.				
220	280	190	300	150

$\frac{N+1}{2}$ Data for wages are Rs. 220, 280, 190, 300 and 150. Let us now calculate the median.

On applying the formula: , we will get,

$\frac{5+1}{2}$

=

= 3 or the 3rd item.

Now, on arranging the data from ascending to descending we will get this data 150, 190, 220, 280, and 300. Thus the median is Rs. 220, which is the 3rd item or the middle value of the data.

Thus, both the mean and median are examples of a statistic, which is simply a number that gives information about a sample. Sometimes the median gives a truer representation or typical element of the data than the mean.

Advantages of median are:

- It is easy to understand and calculate.
- It can be calculated even if the value of the extreme items is not known, but the number of items should be known.
- Like an ideal average median is rigidly defined.
- It can be used with non-numerical data.

Median suffers from the following defects:

- When there are large variations between the values of different items, then median is not a representative average of a series.

- Median is not suitable for further algebraic treatment. Example: we cannot find out the total values of the item, if we know their number and median. It is easily known through arithmetic mean.

There are also other positional measures such as: Quartiles, Deciles, and Percentiles. These are also called as Partition Values.

Let us look at each of them in more detail.

Quartiles: It divides a series into four equal parts. For any series, there will be three quartiles.

First or Lower Quartile (Q1): Q1 divides the distribution in such a way that one fourth of the total items fall below it and three fourth falls above it.

For Individual and Discrete Series:

Q1= Size of the (N+1/4)th item

In case of Discrete Series, for N, cumulative frequency is calculated.

For Continuous Series:

Q1 = Size of the {N/4}th item

It will determine the size of class interval where Q1 falls. For interpolation of the value of Q1, we use this formula:

$$Q_1 = L_1 + \frac{\frac{N}{4} - cf}{f} x C$$

Where,

Q1 = Lower Quartile or First Quartile.

L1 = Lower limit of class interval where Q1 lies.

N = Number of observations.

Cf = Cumulative frequency of the previous group.

f = simple frequency of Q1 group.

C = class interval.

The second quartile Q2 is a Median.

The Third or Upper Quartile (Q3) divides the distribution is such a way that three fourth of the total items fall below it and one fourth of the total items above it.

For Individual and Discrete Series:

In individual observations and discrete series, the third Quartile is determined by the formula: Q3 = Size of the 3(N+1/4)th item.

In case of discrete series, cumulative frequency has to be calculated.

For Continuous Series:

In continuous series, like median and first quartile, the actual value has to be interpolated from the class interval which we get through the formula:

Q3 = Size of the 3(N/4)th item.

This will determine the class in which Q3 falls.

In order to determine the actual value Q3, we have to apply the formula of interpolation.

$$Q = L + \frac{3\{N/4\} - cf}{f} xC$$

Deciles divide the series into 10 equal parts. For any series, there are

9 deciles as there are three quartiles for any series. Deciles range from D1 to D9.

For Individual and Discrete Series: For determining decile in individual observations and discrete series, the formula is:

D1 = Size of the (N+1/10)th item.

Where, D1 = First decile, N = number of items.

In case of discrete series for N, cumulative frequency has to be calculated.

For Continuous Series:

For determining the deciles in continuous series, we take only size of (N/10)th item.

Thus, in continuous series, before interpolating the value of deciles, we have to determine the class group in which deciles falls.

Percentile divides the series into 100 parts. For any series, there are 99 percentile. Percentile is denoted by P. It ranges from P1 to P99.

For Individual and Discrete Series: For different percentile:

P1 = Size of the (N+1/100)th item.

P50 = Size of the 50(N+1/100)th item

P80 = Size of the 80(N+1/100)th item

P99 = Size of the 99(N+1/100)th item

In case of discrete series cumulative frequency has to be calculated to determine percentile.

First Continuous Series: In continuous series percentile cannot be determined as it is determined in case of individual or discrete series. In continuous series, we have to interpolate the actual value of percentile with the help of the formula.

P1 = Size of the (N/100)th item.

The class where percentile 20 falls is determined as under:

P20 = Size of the 20 (N/100)th item.

$P_{20} = L + \frac{20\{N/100) - cf}{f} xC$ This will determine the group where percentile falls. For determining the actual value of percentile, the interpolation has to be done.

Thus, we can calculate quartiles, deciles and percentiles with the help of formulas.

The mode is simply the value of the relevant variable that occurs most often (that is, has the highest frequency) in the sample.

Note that if you have done a frequency histogram, you can often identify the mode simply by finding the value with the highest bar.

However, that will not work when grouping was performed prior to plotting the histogram. Although you can still use the histogram to identify the modal group, just not the modal value.

Modes in particular are probably best applied to nominal data.

Thus, the most frequently occurring variable of the series is known as Mode. The variable which is repeated maximum number of times in the series will be the Mode of the series.

Let us take an example of runs scored by a batsman is a series.

The number of runs scored in a series of cricket games is listed below. Let us see which score occurred most often.

Score	12	14	23	20	14	10	19

On arranging the scores from least to greatest, we will get:

Score	10	12	14	14	19	20	23

Here, the score which occurs most often is 14.

Thus, mode = 14.

Advantages of mode are:

- It is easy to calculate and understand. That means it can be determined without much mathematical calculations
- It is not affected by the values of the extreme items and it is preferred over mean
- It can be used with non-numerical data
- It can also be determined graphically

Disadvantages of mode are:

- The mode cannot be determined always.
- It is not capable of algebraic treatment, as we can do in case of mean.
- Mode is also not rigidly defined
- Mode value is not based on each and every item of the series
-

• Commercial Averages

Commercial Averages include: Moving average, Progressive average, and Composite average.

Measure the average price or exchange rate of a currency pair over a specific time frame. Moving average is generally plotted on a graph.

For example, if we take the closing prices of the last 10 days, add them together and divide the result by 10, we have created a 10-day simple moving average. Generally moving average is plotted on a graph.

Advantages of moving average are:

- Method is easy as there is no mathematical complexities
- Quite flexible
- Follows the general movement of data
- Very effective if the trend of the series is very irregular

Disadvantages of moving average are:

- Cannot compute trend values for all the years
- If the trend is not linear the moving average in that case would be either above or below the true sweep of the data

Progressive average is used by business houses particularly in early years with a view to compare the current profits with those of the past.

Composite Average: It is also calculated by the help of simple arithmetic average. It is a cumulative average and is different from the moving average.

In the calculation of this average, figures of all previous years are added and no figure is left out as in the case of moving average.

Thus, the progressive average of the second year would be equal to the arithmetic average of the figures of the first two years.

The progressive average of the third year would be equal to the arithmetic average of the figures of the first three years and so on.

Here's a summary of our learning in this session:

- Explain and Define Central Tendency
- Explain the uses, classifications and objectives of Average or Central Tendency
- Explain the different types of Averages and their advantages and disadvantages.
- Arithmetic Mean, Geometric Mean , Harmonic Mean, Median, Mode , Moving Average, Progressive Average, Composite Average and partition values like Quartiles, Deciles, and Percentiles.