1. Introduction

Welcome to the series of E-learning modules on Median and Mode. In this module, we are going to cover the definition, calculation of Median and Mode for the various statistical series, its application merits and limitations.

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

- Explain the concept of Median and Mode
- Calculate the Median and Mode for statistical series
- Discuss the Application of Median and Mode
- List the Merits and Limitations of Median and Mode

Large humps of statistical figures generally, create confusion in mind and make it difficult to grasp out any useful information. It as such becomes essential to present the collected data in some condensed form so as to make them understandable, comparable and worthy of scientific treatment.

For this purpose, a central value, which represents the whole mass of data, is worked out. This value is called the central tendency or central value or average.
The various measures of central tendency may be grouped in the following manner.

The various measures of central tendency may be grouped in the following manner. Averages are broadly classified into three types Mathematical, positional and Commercial which is further bifurcated into Arithmetic mean, geometric mean, harmonic mean, positional into median and mode and commercial into moving, progressive and composite averages.

In this module, we will focus on the positional averages as we have already discussed the mathematical average in our previous modules of E-learning.

A Positional Average is an average, which determines the position or place of central value or variable in the series.

Positional Averages have nothing to do with the sum of the values of the variables like the mathematical averages but are concerned about the position they hold. The extreme items of the data set do not affect them.
Median and mode are the two positional averages we will discuss in this module.

Median is a positional average, which is widely used, in statistical analysis. It refers to the middle value of a distribution when the series are arranged in ascending or descending order. In other words, median is a value, which divides the series into two equal parts. Median is always determined by arranging the series in the ascending or descending order.

According to Socarist, median is defined as “median of a series is the value of the item, actual or estimated when series is arranged in order of magnitude which divides the distribution into two parts”.

It is denoted with the alphabet ‘M’.

Like the other measures of central tendency, we also calculate the positional average Median of a statistical distribution for all the statistical series that is the individual series, discrete frequency distribution and continuous frequency distribution.

2. Calculation of Median for Individual and Discrete Series

Let us now, discuss the calculation of median for individual series.
An individual series is a set of raw data. In this series, the following steps are to be followed to find the median.
Arrange the data in a series in ascending or descending order.
There are two types of series odd series and even series.

In odd series, that is if the data set has value numbering to 1,3, 5, 7, 9…. then we add one to the total number and get a value that is divisible by 2. In this case, the median will actually lie in the series.

Median = size of (N+1)/2 th item

In even series, that is of the data has value numbering to 2, 4, 6, 8…… we add one to the total number and get a value which is not exactly divisible by 2. Hence, we need to go in detail to find the median value and this value will not actually fall in the series.
Median = size of (N+1)/2 th item and then the average of the middle two values is taken for identifying the Median.

Let us take an example each for an odd series and even series to find out how we calculate the median.

Let us first consider the Odd series.
Calculate the value of median for the following figures:
101, 85, 145, 75, 175, 42, 210, 300 and 250

First, we arrange the data in ascending order.

Then, as the number of data in this equal to 9 we have an odd series so we use the following formula median is equal to the size of number of data ‘N’ plus 1 divided by 2.
Median is equal to size of N plus 1divided by2 th item when substituted we get
Median is equal to size of 9 plus 1 divided by 2th item is equal to 10 divided by2 is equal to 5th item in the data which is equal to 145 in the data arranged in the order of magnitude.

Now, let us first consider the even series.

Calculate the value of the median for the following figures:
360, 220, 280, 320, 260, 200, 180 and 400

First, we arrange the data in ascending order.

Then as the number of data in this equal to 8 we have an even series so we use the following formula median is equal to the size of number of data ‘N’ plus 1 divided by 2

Median is equal to size of N plus 1divided by 2th item when substituted we get
Median is equal to size of 8 plus 1divided by 2th item is equal to 9 divided by 2 is equal to 4.5th item in the data that is the average of the 4th item and the 5th item which is equal to 260 plus 280 divided by 2 is equal to 540 divided by 2 is equal to 270 which is not visible in the data arranged in the order of magnitude.

Let us now calculate the median for a discrete frequency distribution.
A discrete frequency distribution is a set of raw data grouped according to the variable and its frequency.
Here, also we use the same formulae for calculating the median.
Median is equal to size of N plus 1 divided by 2th item.
Here, we use the total of the frequency for N.

The following steps are used to calculate the median in a discrete series:

- Arrange the data in ascending or in descending order
- Calculate the cumulative frequencies (CF)
- Find the value of the middle item by applying the formulae
- Find the total in the cumulative frequency column which is either equal to or next higher that the value
- Locate the value of the variable corresponding to the cumulative frequency, which is the median of the data

This can be made clear with the help of the following example.
Calculate the median for the following data:

The data here is already arranged in ascending order. The next step is to calculate the cumulative frequency that is the frequency of the first variable in the data is taken as it is then, for the second variable, we add the frequency of that variable with the preceding value. We continue this method until we get all the values of the cumulative frequencies.

In this data, we have taken the first CF as 2, the second CF is 2+3 =5, the third value is 2+3+4 = 9 and so on we get 15, 25, 30 a 32.

Now, let us apply the formula for this data.
Median is equal to size of N plus 1 divided by 2th item. Here, N is equal to the total of the frequency, which is 32 so we get.
Median is equal to size of 32 plus 1divided by 2 is equal to 33 divided by 2th item is equal to 16.5th item.
Here, the cumulative frequency that is equal to or greater than 16.5 is the cumulative frequency 25 and the corresponding variable is 125. Hence, the median of the given data is equal to 125.

3. Calculation of Median for Continuous Frequency Distribution

Now, let us calculate the median of continuous frequency distribution.

The following steps are used to calculate the median in a continuous series:

- Arrange the data in ascending order
- Calculate the cumulative frequency
- Apply the formula Median = size of (N)/2 th item
- Identify the class interval by examining the cumulative frequency column

Once the class interval is determined apply the formula
Median is equal to l one plus N by two minus cf, whole divided by f into c.
Where,
l1 = lower limit of the median class.
cf = cumulative frequency of the class preceding the median class.
f= simple frequency of the median class.
c=magnitude of the median class.
Note when we use the upper limit (L2) of the class interval we need to use minus instead of plus in the formulae.

Let us take an example to calculate the median for a continuous frequency distribution.
Compute the median from the following series:

Arrange the data in ascending order: In this case, the data is already arranged in ascending order. Next, we calculate the cumulative frequency that the cumulative frequency of the first class interval is as same as its frequency that is 44. The second class interval we add the frequency of the previous class interval to the frequency of this class interval that is 44+60 which is equal to 104 and so on as shown in the table.

Now, we calculate the median class using the formulae size of (N)/2 th item, which is equal to 192/2, which is equal to 96th item. Here the cumulative frequency that is equal to or greater than 96th item is the cumulative frequency 104 and the corresponding class interval is 5-10. Hence, the median class for the given data is equal to 5-10.

The next step, we will interpolate the median value from the class interval 5-10 using the formula Median is equal to l one plus N by two minus cf, whole divided by f into c, where in the values are
l1 is the lower limit of the class interval = 5. N/2 is the size of the item 96th item. cf is the cumulative frequency of the preceding class which is equal to 44, f is the frequency of the median class which is equal to 60 and c width of the class interval is equal to 5 when these values are substituted and calculated we get the median as 9.333.
Important points to be noted while solving the median problems are:

If the given data is in the inclusive series form, it should be converted into exclusive series.

If class intervals are unequal, then it need not be converted into equal form.
If the given data is open-ended class interval, then it need not be converted into close ended class interval expect for the median class.
If cumulative frequencies are given, then convert it into simple frequencies.

The partition values are those values which divides the data into various parts a few of them are mentioned below:
The median as a positional average divides the data set into two equal parts similarly there are other positional averages that divide the data four called as quartiles, five called as quintiles, six called as hextiles, eight called as octiles, ten called as deciles and hundred called as percentiles.
These measurements depend on the median and the procedure for calculating these positional averages is also same as that of the median.

Mode is also a positional average, which represents the most frequently occurring items of the series. It means it represent the item that is repeated maximum number of items in the series.

The word is derived from the French word 'la made' which means fashion or the most popular phenomenon.

The extreme items do not affect mode. It is easy to calculate and it helps in determining the popularity of a commodity. It is denoted by the alphabet 'Z'.

In the words of Croxton and Cowden, "The mode of a distribution is the value point around which items and tends to be most heavily concentrated. It may be regarded as the most typical of a series of values."
The various types of modes are unimodal, bimodal, multimodal and no mode depending on the values in the data set.

Unimodal- when a single value occurs more frequently than any other value the distribution is called unimodal.

Bimodal- when there are two different values having equal and maximum frequencies associated with them then the distribution is called as bimodal.

Multimodal – when there are more than two values having equal and maximum frequencies associated with them then the distribution is called multimodal.

No mode – if all the values of the series are unique, then there is no mode to determine we call this as no mode or indeterminate mode.

The different methods to Calculate mode are:

- Observation method
- Grouping & Analysis method
- Interpolation
- Graphical method
- Empirical Relation method

Observation Method is a method in which we will inspect the data set to identify the values that are repeating the maximum number of times and then, that value will be the mode for the data set.

This method is popular while calculating the mode for individual series.

Grouping and Analysis method is a method in which we use a grouping and analysis table to calculate the mode. This method is generally applied when the difference between the frequency preceding or succeeding the highest value is very small and the items are heavily concentrated on either side. This method is used both in discrete frequency distribution and continuous frequency series.

The grouping table consists of seven columns. They are:
Column I: We take the variables of the given data.
Column II: We take the original frequencies and encircle the maximum value.
Column III: Frequencies are added in two and the highest value is encircled.
Column IV: Leaving the first frequency the remaining is added in two and the highest value is encircled.

Column V: Frequencies are added in three and the highest value is encircled.
Column VI: Leaving the first frequency the remaining is added in three and the highest value is encircled.
Column VII: Leaving the first two values add the remaining in threes and encircle the highest value.
This finishes the preparing of the grouping table.

4. Calculation of Mode using Interpolation Method

Now we prepare the analysis table. In an analysis table, we take the variables horizontal and the column values vertically and each column highest value is taken and crossed against the variables. Then, the total of each variable is taken the variable with the highest total is called the mode.

In Interpolation method, the mode is interpreted using a formula. Z is equal to the lower limit (l1) of the modal class plus the frequency of the modal class (f1) minus the frequency of the preceding class (f0) divided by twice the frequency of the modal class (2(f1) minus the frequency of the preceding class (f0) minus the frequency of the succeeding class (f2) the whole thing multiplied by the width or magnitude of the class interval (c).
Depending on the size of the class interval the modal class will change. It is affected by the frequencies of the neighboring classes. This method is used to calculate the mode in the continuous frequency distribution.

Graphical method is also called as curve fitting method. In this method, we use the histogram of the given data and draw two lines diagonally inside the modal class starting from each upper corner of the adjacent rectangles and then draw a perpendicular line from the intersection of the diagonal lines to the X-axis, which indicates the value of the mode. This method is used to calculate the mode when the class intervals are unequal.

A famous statistician Karl Pearson gives empirical Relation method. It establishes a relationship between the three measures of central tendency that is Mean (X bar), Median (M) and Mode (Z).

The relationship is stated as follows
Mode is equal to thrice the median minus twice the mean.
This formula is used to compute the third value when we know any two values of a given set of data. This formula is also used when the data has a bimodal.

Let us discuss the calculation of Mode for Individual series.
An individual series is a set of raw data. In this series, we use the inspection method to calculate the Mode.

Let us take an example to understand the calculation.
Calculate the mode for the following data:
10, 27, 24, 12, 27, 27, 20, 18, 15, 30

As this is an individual series, we use the inspection method.
When the data is observed we find that the value 27 repeats the maximum number of times that is it is repeated for 3 times in the data and so 27 is the mode for this data.

Let us discuss the calculation of Mode for discrete series.

A discrete frequency distribution is a set of raw data grouped according to the variable and its frequency.
Here, we can use both the inspection method and the Grouping and Analysis table for calculation of the table.

The whole procedure can be made clear with the help of an example.
Calculate the mode for the following data.

In this example, we use the inspection method as there is a gradual rise or fall in the frequencies and the highest frequency and the adjacent frequencies are not too close. Thus, the variable 12 with the highest frequency 15 is the mode for the given data.

From the following data of the ages of different persons, determine the modal.
In this data, we can see that the adjacent frequencies are close to each other and hence, we will use the grouping and analysis table for calculating the mode.

First, let us prepare the grouping table.

In this table, we have taken the variable age as 15, 20…in the first column, in the second column is the original frequencies 2, 3, 4, 10….. and the highest value 12 is encircled. The third column we add the frequencies in two that is 2 plus 3=5, 10 plus 4 =15, and so on and encircle the highest value 23. In column four, we leave the first item that is 2 and then add the frequencies in two that is 3 plus 4 =7, 10 plus 11 = 21 and so on and the highest value 21 is encircled. In column five, we add the frequencies in threes 2 plus 3 plus 4 =9, 10 plus 11 plus 12 is equal to 3 and so on and the highest value 33 is encircled. In column six, we leave the first item 2, add the frequencies in three that is 3 plus 4 plus 10 = 17, and so on, and encircle the value 26. In the last column, we leave away the first two items, add the remaining frequencies in three that is 4 plus 10 plus 11 =25, and encircle the highest value 25. Thus, the grouping table is done.
Next, we will prepare the analysis table.
Here, we have taken the age variable horizontal and the column values vertically. In the first row, we will take the highest value of the first column and mark a cross against the corresponding variable 40. In the second row, let us take the value of the second column 23, which is corresponding to the variable 35 and 40 and put a cross against these variables. Similarly, we do for all the values of column 3, 4, 5, 6 and 7 and then take the total. We see that the total of variable 35 is high and hence it is the modal age.

When we compare mode in inspection and grouping analysis methods, we will see that the mode through inspection is 40 and through grouping is 35. The position of the mode shifts with the weightage of the preceding and succeeding items.

5. Calculation of Mode for Continuous Frequency Distribution

Now, let us discuss the calculation of mode for continuous frequency distribution.
In a continuous series, we cannot locate the mode directly as the mode lies in a class interval. Hence, with the help of a formula we interpolate the mode with the assumption that the values in each class interval are equally distributed.

Let us take an example to understand the interpolation calculation of mode in continuous frequency distribution.

Calculate the mode for the following data:

The solution is that by inspection we can identify the modal class as 10-20 as that is the class with the highest frequency.

Now, let us apply the formulae for interpolation of the mode.
Let us substitute the value in the equation:
L1 = 10, f1 = 520, f0=278, f2=464 and c=10
Z is equal to 10 plus (520 minus 278) divided by (2 multiplied by 520) minus 278 minus 464) multiplied by 10
Z is equal to 10 plus 242 divided by 298 multiplied by 10
Z is equal to 10 plus 2420 divided by 298 is equal to 10 plus 8.12 is equal to 18.12
Thus, the modal value is 18.12.

Let us take an example to understand how we use the graphical method in calculating the mode.
Determine the mode by constructing a histogram for the following data:

By inspection, the modal class is 9-17. We take the three rectangles that is the rectangle corresponding to the modal class, pre-modal class and post modal class, draw two lines diagonally inside the modal class starting from each upper corner of the adjacent rectangles and then draw a perpendicular line from the intersection of the diagonal lines to the X-axis which indicates the value of the mode. In this case it is 14.7

To discuss the Empirical relation method, let us take an example and understand the calculation of mode through this method.

For a given set of data, the mean value is 150 and the median is 163. Find the mode of the data.

Mean (X) is equal to 150
Median (M) is equal to 163
To find the Mode (Z), we can use the following formula:
Mode is equal to thrice the median minus twice the mean

Which is equal to 3 multiplied by 163 minus 2 multiplied by 150
Which is equal to 489 minus 300
Which is equal to 189
This relationship explains that the data is asymmetrical and is positively skewed because the mean is less than the median and the median is less than the mode.

The following are the important points to be considered while solving the problems of mode.

- If the data is given in inclusive form, we should convert it to exclusive form.
- If the data is given in unequal class interval, converting it into equal class interval is compulsory.
- If cumulative frequency is given, then convert it into simple frequencies.
- If the given data is in open-ended class interval, then it is not necessary to convert the data into close ended except for the modal class.

Following are the important points to be noted while solving the Mode problems.

- If the given data is having frequencies very near to the maximum frequency, then it is compulsory to prepare the grouping analysis table for getting accurate results.
- If first class of the given data is the modal class, then take the frequency of the preceding class as zero.

Let us list the merits and demerits of mode.

The main merits of mode are:

- Mode is simple to calculate.
- It is commonly understood and implemented by people in day-to-day life.
- It is the most common item of the series.
- It is not affected by the values of extreme items.
- It can be determined in open-ended classes without knowing the limits.
- It is used to describe the qualitative phenomenon.

The main demerits of mode are:

- Mode cannot be determined always, it could be bimodal, multimodal, etc
- It is not capable of further algebraic treatment
- It is not based on each and every item of the data
- It is not rigidly defined, because there are different methods to calculate and each of the method does not render the same results
- It does not possess any essential of a good average
- It is affected by the fluctuations of the sampling
- Many times, it is ill defined, indefinite and indeterminate

Here's a summary of our learning in this session, where we have understood:

- The concept of Median and Mode
- Calculation of Median and Mode for statistical series
- Application of Median and Mode
- Merits and Limitations of Median and Mode