Measures of Central Tendency: Mean, Median, and Mode

A measure of **central tendency** is a summary statistic that represents the center point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution. You can think of it as the tendency of data to cluster around a middle value. In statistics, the three most common measures of central tendency are the mean, median, and mode. Each of these measures calculates the location of the central point using a different method.

The mean, median and mode are all valid measures of central tendency, but under different conditions, some measures of central tendency become more appropriate to use than others. In the following sections, we will look at the mean, mode and median, and learn how to calculate them and under what conditions they are most appropriate to be used.

Mean (Arithmetic)

The arithmetic mean is the most common measure of central tendency. It is simply the sum of the numbers divided by the number of numbers. The symbol "μ" is used for the mean of a population. The symbol "M" is used for the mean of a sample. The formula for μ is shown below:

$\mu = \Sigma X/N$

where ΣX is the sum of all the numbers in the population and

N is the number of numbers in the population.

The formula for M is essentially identical:

$M = \Sigma X/N$

where ΣX is the sum of all the numbers in the sample and

N is the number of numbers in the sample.

As an example, the mean of the numbers 1, 2, 3, 6, 8 is 20/5 = 4 regardless of whether the numbers constitute the entire population or just a sample from the population.

Although the arithmetic mean is not the only "mean" (there is also a geometric mean), it is by far the most commonly used. Therefore, if the term "mean" is used without specifying whether it is the arithmetic mean, the geometric mean, or some other mean, it is assumed to refer to the arithmetic mean.

Characteristics of Mean

1. It is the value in a given aggregate which would obtain if all the values were equal.
2. The sum of deviations on either side of the mean is equal; hence, the algebraic sum of the deviation is equal zero.
3. It reflects the magnitude of every value.
4. An array has one and only one mean.
5. Means may be manipulated algebraically: means of subgroups may be combined when properly weighted.
6. It may be calculated even when individual values are unknown, provided the sum of the values and the sample size n are known.
7. Values need not be ordered or grouped for this calculation.
8. It cannot be calculated from a frequency table when ends are open.
9. It is stable in that grouping procedures do not seriously affected it.

Merits:

1. It can be easily calculated; and can be easily understood. It is the reason that it is the most used measure of central tendency.
2. As every item is taken in calculation, it is affected by every item.
3. As the mathematical formula is rigid one, therefore the result remains the same.
4. Fluctuations are minimum for this measure of central tendency when repeated samples are taken from one and the same population.
5. It can further be subjected to algebraic treatment unlike other measures i.e. mode and median.
6. A.M. has also a plus point being a calculated quantity and is not based on position of terms in a series.
7. As it is rigidly defined, it is mostly used for comparing the various issues.

Demerits

1. It cannot be located graphically.

2. A single item can bring big change in the result. For example if there are three terms 4, 7, 10 ; X is 7 in this case. If we add a new term 95, the new X is 4+7+10+95/4 = 116/4 = 29. This is a big change as compared to the size of first three terms' X- .

3. Its value will be effective only if the frequency is normally distributed. Otherwise in case skewness is more, the results become ineffective.

4. In case of open end class intervals we have to assume the limits of such intervals and a little variation in X can take place. Such is not the case with median and mode, and there is no use of the open end intervals in its calculations.

5. Qualitative forms such as Cleverness, Riches etc. cannot give X as data can't be expressed numerically.

6. X cannot be located by inspection as in the case of mode and median.

7. Sometimes it gives impossible or laughable conclusions, e.g. if there are 60, 50 and 12 students in three classes then average number of students is 60+50+42/3 = 50.67, which is impossible as students can't be in fractions.

Median

The median is also a frequently used measure of central tendency. The median is the midpoint of a distribution: the same number of scores is above the median as below it. For the data in Table 1, there are 31 scores. The 16th highest score (which equals 20) is the median because there are 15 scores below the 16th score and 15 scores above the 16th score. The median can also be thought of as the 50th percentile.

Computation of the Median

When there is an odd number of numbers, the median is simply the middle number. For example, the median of 2, 4, and 7 is 4. When there is an even number of numbers, the median is the mean of the two middle numbers. Thus, the median of the numbers 2, 4, 7, 12 is (4+7)/2 = 5.5. When there are numbers with the same values, then the formula for the third definition of the 50th percentile should be used.

Characteristics of Median

1. It is the value of the middle point of the array (not midpoint of range), such that half the item are above and half below it.
2. The value of the media is fixed by its position in the array and doesn't reflect the individual value.
3. The aggregate distance between the median point and all the value in the array is less than from any other point.
4. Each array has one and only one median.
5. It cannot be manipulated algebraically: medians of subgroups cannot be weighted and combined.
6. It is stable in that grouping procedures do not affect it appreciably.
7. Value must be ordered, and may be grouped, for computation.
8. It can be compute when ends are open
9. It is not applicable to qualitative data.

Advantages of Median:

1. It is very simple to understand and easy to calculate. In some cases it is obtained simply by inspection.
2. Median lies at the middle part of the series and hence it is not affected by the extreme values.
3. It is a special average used in qualitative phenomena like intelligence or beauty which are not quantified but ranks are given. Thus we can locate the person whose intelligence or beauty is the average.
4. In grouped frequency distribution it can be graphically located by drawing ogives.
5. It is especially useful in open-ended distributions since the position rather than the value of item that matters in median.

Disadvantages of Median:

1. In simple series, the item values have to be arranged. If the series contains large number of items, then the process becomes tedious.

2. It is a less representative average because it does not depend on all the items in the series.

3. It is not capable of further algebraic treatment. For example, we can not find a combined median of two or more groups if the median of different groups are given.

4. It is affected more by sampling fluctuations than the mean as it is concerned with on1y one item i.e. the middle item.

5. It is not rigidly defined. In simple series having even number of items, median cannot be exactly found. Moreover, the interpolation formula applied in the continuous series is based on the unrealistic assumption that the frequency of the median class is evenly spread over the magnitude of the class interval of the median group.

Mode

The mode is the most frequently occurring value. For the data in Table 1, the mode is 18 since more teams (4) had 18 touchdown passes than any other number of touchdown passes. With continuous data such as response time measured to many decimals, the frequency of each value is one since no two scores will be exactly the same (see discussion of continuous variables). Therefore the mode of continuous data is normally computed from a grouped frequency distribution. Table 2 shows a grouped frequency distribution for the target response time data. Since the interval with the highest frequency is 600-700, the mode is the middle of that interval (650).

Characteristics of mode

1. It is the most frequent value in the distribution; it is the point of greatest density.

2. The value of the mode is established by the predominant frequency, not by the value in the distribution.

3. It is the most probable value, hence the most typical.

4. A distribution may have 2 or more modes. On the other hand, there is no mode in a rectangular distribution.

5. The mode does not reflect the degree of modality.

6. It cannot be manipulated algebraically: modes of subgroups cannot be combined.

7. It is unstable that it is influenced by grouping procedures.

8. Values must be ordered and group for its computation.

9. It can be calculated when table ends are open.

Merits of Mode:

1. It possesses the merit of simplicity. In a discrete series it can be located even by inspection. Hence like median it has an advantage over arithmetic average.

2. It is commonly understood. It is an average which people use in their day to day expressions. The average size of ready-made garments, the typical size of land holding is all examples of the common use of mode.

3. Mode is a value which exists in the series whereas arithmetic average may be a figure which may not be found in the series. It is the most common item of a series and is not an isolated example like the median.

4. It is not affected by the value of extreme items if the distribution follows the natural law relating to extremes. Usually, there is little concentration of items around extreme values.

5. It can be correctly calculated in open-end classes.

6. In a continuous series, mode can, be calculated even if all the item values are not given. Only the modal class and the frequencies of its adjoining classes are required to compute mode.

Demerits:

1. Mode is ill-defined in case of bi-modal, multi-modal series.

2. It is not a representative average as it is not based on all the items of the distributions, If in a series of 1000 items 20 have a particular value and other values have frequency less than 20, mode becomes 20. But certainly 20 is not the typical or representative average.

3. It is not capable of further algebraic treatment.

4. Mode is affected to a great extent by the fluctuations of sampling.