One of the important objectives of statistical analysis is to determine various numerical measures which describe the inherent characteristics of a frequency distribution. The first of such measures is average. Averages are the typical values around which other items of the distribution congregate. They are the values which lie between the two extreme observations of the distribution and give us an idea about the concentration of the values in the central part of the distribution. Accordingly, they are also sometimes referred to as the measures of central tendency or central location. The three most commonly used measures are the mean, median and mode.

The mean of a set of observations is their arithmetic mean. If we think of observations as a series of different weights strung out along a thin rod, the mean is the point at which the rod would balance. So mean is the balance point for data.

Formula for the Mean



But, the strongest drawback of mean is that it is very much affected by extreme observations. Two or three very large values of the variable may unduly affect the value of the mean. For example, let us consider an industrial complex which houses the workers and some officials in the top hierarchy like general manager, chief engineer, architect etc. The average salary of the workers (skilled and unskilled) is, say, Rs.8000 per month. If the salaries of the few officials who draw very high salaries are also included, the average wage per worker comes out to be Rs. 12000 say. Thus, if we say that the average salary of the workers in the factory is Rs. 12000 p.m. it gives a very good impression and one is tempted to think that the workers are well paid and their standard of living is good. But the real picture is entirely different. Thus, in the case of extreme observations, the mean gives a distorted picture and is no longer representative of the distribution and quite often leads to very misleading conclusions. Thus, while dealing with extreme observations, mean should be used with caution.

In these situations, a second measure called the median may be more appropriate.

The median is the middle observation when the data are arranged from smallest to largest. If there is an odd number of observations, the median is the middle observation. For example, if there are nine observations, the median is the fifth smallest (or fifth largest) observation. If there is an even number of observations, the median is usually defined to be the mean of the two middle observations. For example, if there are 10 observations, the median is usually defined to be the mean of the fifth and sixth smallest values. So, the median is the point with half the observations on either side.

The mode is another measure that is occasionally used as a measure of central tendency. The mode is the most frequently occurring value. If no value occurs more frequently than any other, the data set is said to have no mode. The mode might be particularly useful in describing the central location value for clothes sizes. For example, shoes come in full sizes. Consider the following sample data that have been sorted from low to high:



The mean for these sample data is



Although 9.417 is the numerical mean, the mode is 10, because more people wore that size shoe than any other. IN making purchasing decisions, a shoe store manager would order more shoes at the modal size than at any other size. The mean isn’t of any particular value in the purchasing decision.

Sometimes the highest frequency occurs at more than one measurement. When this happens, two or more modes exist. When exactly two modes exist, we say the data are bimodal. When more than two modes exist, we say the data are multimodal.