

Merits of the Median

Although the mean is usually the preferred measure of central tendency, there are several situations in which the median is preferred. As I mentioned earlier, the median is not sensitive to the values of the scores above and below it—only to the number of such scores. Unlike the mean, it is not affected by extreme scores, and thus it is a more representative measure of central tendency for very skewed distributions. Also, it can be computed when the values of the extreme scores are unknown. Suppose, for example, that I recorded the number of trials required to learn a list of paired adjectives and Japanese kana (writing) symbols. The data are as follows: 12, 17, 17, 18, 21, 24, > 41 . After the 41st trial, the poorest learner was still unable to learn the list and gave up; his score is some number greater than 41. The distribution is open-ended because the value of the extreme score is unknown. Although the exact value of one of the scores is unknown, the median can be computed for these data. Notice that three scores are above 18 (21, 24, > 41) and three are below (12, 17, 17); hence, the median is 18. The mean cannot be computed because the value of the extreme score is unknown.

The median has the added advantage of being easy to compute; when the number of scores is small, it can be determined by inspection.

The principal disadvantages of the median relative to the mean are (1) its poorer sampling stability and (2) its poorer mathematical tractability. For these and other reasons, the median is not used as frequently as the mean in advanced descriptive and inferential statistical procedures.

Merits of the Mode

The mode is the only measure of central tendency that can be used with un-ordered qualitative variables such as eye color, blood type, race, and political party affiliation. For quantitative variables that are inherently discrete, such as family size, it is sometimes a more meaningful measure of central tendency than the mean or the median. Who ever heard of an average family with 3.7 members? It makes more sense to say that the most typical family size is 3, the mode. Other than these two applications, the mode has little to recommend it except its ease of estimation.

Let us consider why the mode is called the most typical score. Because the mode is the score that occurs most frequently, the number of scores not equal to the mode is as small as it possibly can be. In Figure 3.5-1, for example, three scores differ from the mode; they are 2, 4, and 8. However, four scores differ from the mean (2, 3, 3, and 8), and five scores differ from the median (2, 3, 3, 4, and 8). Hence, the mode is the most typical score.

The mode has a number of limitations. Its sampling stability is much poorer than that of the mean and the median, and it also is less mathematically tractable. Therefore, it is rarely used in advanced descriptive and inferential statistics. However, the mode, like the median, can be computed for an open-ended distribution if the distribution is known to be unimodal and if the unknown scores do not have the greatest frequency. However, because of the median's superior mathematical properties, it is preferred for this application.

Consider another limitation of the mode. A mode may not exist for a set of data, as when the distribution is bi- or multimodal. In such cases, it is customary to report the two or more scores with the same maximum frequency. Because many variables in the behavioral sciences are approximately normally distributed, the existence of two scores with the same maximum frequency suggests the presence of two underlying distributions. This would occur if I administered a test of masculinity to a sample containing an equal number of men and women. To report a mean or a median for such data would be misleading without also reporting that the distribution is bimodal and the values of the maximum scores.

Summary of the Properties of the Mean, Median, and Mode

The mean is

1. the balance point of a distribution, the point for which $\sum_{i=1}^n (X_i - \bar{X}) = 0$;
2. the preferred measure for relatively symmetrical distributions and quantitative variables;
3. the measure with the best sampling stability;
4. widely used in advanced statistical procedures;
5. mathematically tractable;
6. the only measure whose value is dependent on the value of every score in the distribution;
7. more sensitive to extreme scores than the median and the mode and, hence, is not recommended for markedly skewed distributions;
8. not appropriate for qualitative data; and
9. not appropriate for open-ended distributions.

The median is

1. the point that divides the ordered scores into two samples of equal size;
2. second to the mean in usefulness;
3. widely used for markedly skewed distributions because it is sensitive only to the number rather than to the values of scores above and below it;
4. the most stable measure that can be used with open-ended distributions;
5. more subject to sampling fluctuation than the mean;
6. less mathematically tractable than the mean; and
7. less often used in advanced statistical procedures.

The mode is

1. the score that occurs most often and, therefore, the most typical value;
2. the only measure appropriate for unordered qualitative variables;
3. more appropriate than the mean or the median for quantitative variables that are inherently discrete;
4. the easiest measure to compute;
5. much more subject to sampling fluctuation than the mean and the median;
6. less mathematically tractable than the mean and the median;
7. not necessarily existent, as when a distribution has two or more scores with the same maximum frequency; and
8. rarely used in advanced statistical procedures.