

Measures of Variability

- Measures of central tendency, such as the mean and the median described, provide useful information. But it is important to recognize that these measures are limited and, by themselves, do not provide a great deal of information.
- There is an old saying that provides a caution about the mean: “If your head is in the freezer and your feet are in the oven, on average you’re comfortable.”
- Suppose I gave a sample of 100 fifth-grade children a survey to assess their level of depression. Suppose further that this sample had a mean of 10.0 on my depression survey and a median of 10.0 as well. All we know from this information is that the mean and the median are in the same place in my distribution, and this place is 10.0.
- Now consider what we do not know. We do not know if this is a high score or a low score. We do not know if all of the students in my sample have about the same level of depression or if they differ from each other. We do not know the highest depression score in our distribution or the lowest score. Simply put, we do not yet know anything about the dispersion (i.e., the spread) of scores in the distribution. In other words, we do not yet know anything about the variety of the scores in the distribution.
- There are three measures of dispersion that researchers typically examine: the range, the variance, and the standard deviation. Of these, the standard deviation is the most widely used because it is both easy to understand and provides a summary statistic of the average amount of variation within a distribution.

Range

- The range is simply the difference between the largest score (the maximum value) and the smallest score (the minimum value) of a distribution. This statistic gives researchers a quick sense of how spread out the scores of a distribution are, but it is not a particularly useful statistic because it can be quite misleading.
- the interquartile range (IQR). Unlike the range, which is the difference between the largest and smallest score in the distribution, the IQR is the difference between the score that marks the 75th percentile (the third quartile) and the score that marks the 25th percentile (the first quartile).

Variance

- The variance provides a statistical average of the amount of dispersion in a distribution of scores.
- the average difference, or deviation, between each score in the distribution and the mean of the distribution.
- Because of the mathematical manipulation needed to produce a variance statistic, variance, by itself, is not often used by researchers to gain a sense of a distribution.
- In general, variance is used more as a step in the calculation of other statistics (e.g., standard deviation, analysis of variance) than as a stand-alone statistic. But with a simple manipulation, the variance can be transformed into the standard deviation, which is one of the statistician’s favorite tools. -

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

where σ^2 is population variance, \sum is to sum, X is each score in the distribution, μ is the population mean, N is the number of cases in the population.

- For sample:

$$\sigma^2 = \frac{\sum(X - \bar{X})^2}{n - 1}$$

Standard Deviation

- The best way to understand a standard deviation is to consider what the two words mean. Deviation, in this case, refers to the difference between an individual score in a distribution and the average score for the distribution.
- So if the average score for a distribution is 10, and an individual child has a score of 12, the deviation is 2.
- The other word in the term standard deviation is standard. In this case, standard means typical, or average.
- So a standard deviation is roughly the typical, or average, deviation between individual scores in a distribution and the mean for the distribution. This is a very useful statistic because it provides a handy measure of how much spread there is in the scores in the distribution.
- When combined, the mean and the standard deviation provide a pretty good picture of what the distribution of scores is like. -

$$\sigma = \sqrt{\frac{\sum(X - \mu)^2}{N}}$$

where σ^2 is population variance, \sum is to sum, X is each score in the distribution, μ is the population mean, N is the number of cases in the population. \sum is to sum, X is each score in the distribution, μ is the population mean, N is the number of cases in the population. For Sample: -

$$\sigma = \sqrt{\frac{\sum(X - \bar{X})^2}{n - 1}}$$

Note

- In a sense, the range provides a measure of the total spread in a distribution (i.e., from the lowest to the highest scores), whereas the variance and standard deviation are measures of the average amount of spread within the distribution.

Why Have Variance?

- If the variance is a difficult statistic to understand, and rarely examined by researchers, why not just eliminate this statistic and jump straight to the standard deviation? There are two reasons.
- First, we need to calculate the variance before we can find the standard deviation anyway, so it is no more work. Second, the fundamental piece of the variance formula, which is the sum of the squared deviations, is used in a number of other statistics, most notably analysis of variance (ANOVA), Regression.
- You will see that each of these statistics uses **the sum of squares, which is just another way of saying the sum of the squared deviations**. Because the sum of squares is such an important part of so many statistics, the variance statistic has maintained its place in the teaching of basic statistics.