

## 4.3: Standard Deviation

Though the ranges help show how far scores could vary from each other, they do not indicate how far scores tended to be from a central score. This brings us to a measure of variability called the standard deviation. A **standard deviation** is a descriptive statistic that summarizes how far raw scores tended to fall from the mean in standard units. It can also be thought of as an estimate of the error that occurs when the mean is used to estimate the value of raw scores.

### Interpreting the Standard Deviation

Standard deviations indicate how far scores tended to be from their sample mean. The smaller the standard deviation is, the closer the individual scores tended to be from the mean. The lowest a standard deviation can be is 0; when a standard deviation is 0 it means that scores did not deviate from the mean because they were all equal to the mean (and therefore are also all equal to each other). The larger the standard deviation, the farther the individual scores tended to be from the mean.

Standard deviation can also be interpreted as the expected error when using the mean to estimate a raw score from a sample. Means are used to summarize what tended to be true. Standard deviations summarize how far individual scores tended to be from the mean. Another way to say this is that the standard deviation estimates how wrong the mean is when used to represent the raw scores. Thus, this can be reworded to state that standard deviations summarize how far the mean is estimated or tended to be from raw scores. When the mean is more accurate in estimating the raw scores, the standard deviation will be lower. When the mean is perfect at estimating the raw scores because they are equal to the mean and each other, the standard deviation will be 0. When the mean is less accurate in estimating the raw scores, the standard deviation will be higher.

Notice that each of these two ways to interpret the standard deviation are simply different ways of saying the same thing: standard deviation estimates how similar or dissimilar scores tended to be from the mean and, thus, each other.

### Calculating the Standard Deviation

There are two formulas for standard deviation: one for populations and one for samples. The formulas look complicated but actually use simple operations and can be quite easy to use once you have practiced them a few times. The core of each version of the formula is the same; standard deviation focuses on how far scores tend to deviate (i.e. be different) from the mean.

### Decoding the Symbols

The symbol  $\sigma$  is used to refer to standard deviations of populations. The symbols  $SD$  or  $s$  are used to refer to the standard deviations of samples.  $x$  refers to an individual raw score.  $\mu$  refers to the population mean, whereas  $\bar{X}$  refers to the sample mean. Remember, the mean is calculated the same way for populations and samples, even though the symbols are different.  $N$  refers to the size of a population, whereas  $n$  refers to the size of a sample.

### Sum of Squares Within Formula

As is often the case in statistics, there are smaller formulas inside the standard deviation formulas. One which is subsumed into the standard deviation formula is known as the sum of squares within (often referred to simply as the “sum of squares” or  $SS$ ). The **sum of squares within** is the sum of squared deviations from the mean within a sample. Its name is just a simplified version of its definition and operations. The sum of squares within can be represented using either the symbol  $SS$  or  $SS_w$ .

The  $SS$  formula and steps, in order, are summarized below.

Sum of Squares Within

Formula	Calculation Steps
$SS = \sum (x - \bar{x})^2$	<ol style="list-style-type: none"> <li>1. Find the mean.</li> <li>2. Subtract the mean from each raw score to find each deviation.</li> <li>3. Square each deviation.</li> <li>4. Sum the squared deviations.</li> </ol>

The utility of this formula is not readily apparent on its own as it is rarely used on its own. Instead, it is best for summarizing some core steps that are built upon by other formulas such as the standard deviation and some inferential statistics such as ANOVA (which we will cover in Chapter 10). For now, we will focus on what the steps of the  $SS$  formula are doing.

The SS is a calculation of total deviations from the mean. However, the mean is a balance point of deviations. Therefore, if we were to just find each deviation and then sum, we would always get 0 for every data set. Therefore, a step is added where each deviation is squared to get rid of

all the negatives before summing. Thus, the SS reports the total squared deviations from the mean. In order to bring it from squared units back to standard units, we could add a step at the end where we square root the SS but you will notice this step is not included in the formula and you may wonder why. Recall that the SS is rarely used on its own and, instead, appears within other, larger formulas. Those formulas have more steps that will either include a step to square-root or some other procedures that require the sum of deviations be left in square units. Thus, the SS is left in squared units so that it is ready to go into other formulas that will deal with square rooting later on.

**Population Standard Deviation Formula.** You can see the population formula for standard deviation below. As we begin learning more complex formulas, it can be useful to notice their similarities and differences in structure relative to each other. There are a few core steps, components, and structural aspects of formulas that repeat in several different formulas in statistics. For example, you may notice a structural similarity between the standard deviation formula and the mean formula. Each includes a sum as a numerator and sample size as the denominator. The numerator of each formula sums whatever is the focus of the formula; thus, the numerator sums raw scores in the mean formula and sums deviations in the standard deviation formula. This is because the standard deviation uses the same logic as the mean to find how far scores tended to deviate from the mean. You can think of the standard deviation as the mean of deviations to help you remember what the formula is doing and understand why the formulas have similar structures. The standard deviation uses SS for its numerator because it allows the deviations to be summed without the negative deviations causing the sum to be 0. Thus, the first four steps of the standard deviation formula are the same as the steps for calculating SS.

The SS formula and steps, in order, are summarized below.

Standard Deviation for a Population

Formula	Calculation Steps
$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$	<ol style="list-style-type: none"> <li>1. Find the mean.</li> <li>2. Subtract the mean from each raw score to find each deviation.</li> <li>3. Square each deviation.</li> <li>4. Sum the squared deviations to find SS.</li> <li>5. Divide SS by <math>N</math>.</li> <li>6. Square root the result of step 5.</li> </ol>

Notice that the last step is square rooting. This brings the calculation back into standard units which are easier to interpret than squared units and which are typically reported.

### Sample Standard Deviation Formula

You can see the sample formula below. This formula is very similar to the population formula with two exceptions: 1. It uses sample symbols such as  $\bar{x}$  in place of  $\mu$  and 2. It uses an adjusted sample size in its denominator. This adjustment to the sample size is necessary because data for samples are incomplete and, thus, have more risk of error than data from populations. The standard deviation is an estimate of error so it is presumed to be higher when using samples. Decreasing the denominator, causes the quotient (i.e. the result from division) to increase. In keeping, the adjusted sample size is used to inflate the estimate of error when data are from samples rather than populations. Here are the steps of the sample standard deviation formula, in order:

Standard Deviation for a Sample

Formula	Calculation Steps
$s = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}}$	<ol style="list-style-type: none"> <li>1. Find the mean.</li> <li>2. Subtract the mean from each raw score to find each deviation.</li> <li>3. Square each deviation.</li> <li>4. Sum the squared deviations to find SS.</li> <li>5. Find the adjusted sample size by subtracting 1 from <math>n</math>.</li> <li>6. Divide SS by the adjusted sample size.</li> <li>7. Square root the result of step 6.</li> </ol>

Notice that the last step is square rooting to bring results back to standard units, just as it was for the population version of the formula.

### Connecting Concepts

The formula used to find  $s$  includes a few parts. One important part of the formula looks like this “ $\sum(x - \bar{x})^2$ ” and is referred to as the *sum of squares* or with the symbol  $SS$ . Sum of squares is actually short for “the sum of squared deviations from the mean.” Therefore, the  $s$  formula can also be written as:

$$s = \sqrt{\frac{SS}{n-1}}$$

### Sample Standard Deviation Formula Walkthrough

Let’s work through an example of each step of the sample standard deviation formula using Data Set 4.1. One way to help organize the calculations is to use a table rather than writing them into a formula to make it easier to find and double-check each piece. Therefore, a table format will be used here to show the steps. Take a look at Table 1. The age data from Data Set 4.1 appear in the first column of Table 1 under the heading “Raw Score.” Now that we have the data ready, we can start working through the steps of the formula.

#### 1. Find the Mean.

We found that the mean was 29.00 for these data using the mean formula in Chapter 3 (see Chapter 3 to review the steps to calculate the mean, if needed).

#### 2. Find Each Deviation.

The second step is to find the deviation for each raw score. This means that we must subtract the mean from each raw score. To make it easier to see and to organize this step, a second column appears in Table 1 which shows the sample mean to the right of each raw score. A third column appears to the right of that titled *Deviation*; this column shows what you would get when you subtracted the mean from each raw score. Notice that there are several positive deviations and several negative deviations and that, if you sum the deviations, the result will be 0. This will always be the case because the mean balances deviations and, thus, the sum of deviations from the mean must always be 0.

#### 3. Square Each Deviation.

The third step is to square each deviation. This gets rid of the negatives seen in some of the deviations. The fourth column of Table 1 shows the squared deviation for each raw score.

#### 4. Sum the Squared Deviations.

The fourth step is to add up, or sum, the squared deviations. This gives the Sum of Squares (or  $SS$ ). You can see the sum of squared deviations is 1,850 in the bottom right corner of Table 1.

#### 5. Divide by the Adjusted Sample Size.

The fifth step is to divide the sum of squares ( $SS$ ) by the adjusted sample size. In step 4, we found that  $SS = 1,850$ . The adjusted sample size is calculated as  $n - 1$ . The sample size ( $n$ ) for Data Set 4.1 is 22 making the adjusted sample size for this data set 21. To complete this step, we divide 1,850 by 21 as shown in the bottom of Table 1. The result, which appears under the square root sign in the formula, is 88.0952381...(remember that the ellipsis indicates that the number continues but is being abbreviated).

#### 6. Square Root.

The sixth step is to square root to bring the value back into standard units. This step is necessary because everything was rounded in step 3 and we are now ready to put things back into standard units. When we square root 88.0952381... we get 9.3859063..., or 9.39 when rounded to the hundredths place. You can see this final step and the rounded answer in the bottom of Table 1.

### Interpreting Standard Deviation

The standard deviation provides a summary of how different the raw scores tended to be from the mean. You can think of this as indicating how accurate the mean tends to be in describing each individual in the sample (or population). Remember, a sample mean provides a summary of the whole sample. Thus, the standard deviation is also a summary of the whole sample in relation to its mean. The standard deviation can be described as the average error we would have if we assumed each member of the sample

had a score for the variable that was the same as the sample mean. Of course, the raw scores are often similar, but not identical, to the mean.

Remember, the smaller the standard deviation, the more similar sample raw scores tended to be to the mean. If the standard deviation was 0, it would indicate that there was no deviation. That is the same as saying every raw score was the exact same value as the mean and, thus, that the data were constant rather than varied. The larger the standard deviation, the less similar the sample raw scores tended to be to the mean. There are a few other ways we could convey this same concept. We could say that when standard deviations are smaller, there is less dispersion and that when they are larger, there is more dispersion. Another way we could interpret this is to say that the mean has less error in estimating raw scores when the standard deviation is small compared to when the standard deviation is large. As should be clear here, the standard deviation is a descriptive statistic that builds from and accompanies the mean in describing data.

Table 1 Standard Deviation Calculations for Data Set 4.1 (n = 22)

Raw Score	Mean	Deviation	Squared Deviation
47	29	18	324
46	29	17	289
42	29	13	169
39	29	10	100
36	29	7	49
34	29	5	25
33	29	4	16
33	29	4	16
32	29	3	9
29	29	0	0
29	29	0	0
29	29	0	0
28	29	-1	1
27	29	-2	4
25	29	-4	16
23	29	-6	36
20	29	-9	81
19	29	-10	100
19	29	-10	100
18	29	-11	121
16	29	-13	169
14	29	-15	225

Table 1 Standard Deviation Calculations for Data Set 4.1 (n = 22)

Sum of Deviations = 0	Sum of Squared Deviations (SS) = 1,850
$s = \sqrt{\frac{1,850}{22 - 1}} = \sqrt{\frac{1,850}{21}} = \sqrt{88.0952381...} = 9.3859063... \approx 9.39$	

## The Mean and Standard Deviation Go Together

In statistics and research, we typically report the mean and the standard deviation together because they complement each other in summarizing data. This is because they tell us a more complete story together than either does on its own. The mean tells where the group of scores tended to be. The standard deviation tells us how far raw scores tended to be from the mean. When the mean and standard deviation appear together, they provide summaries of data that aid in making useful comparisons in a way that neither can do alone. For example, you can have the same mean in two samples but different standard deviations. The sample with the smaller standard deviation tended to have raw scores that were closer to the mean than the sample with the larger standard deviation.

Let's review this concept with some examples. Assume five groups of individuals (i.e. five samples) reported how many hours they spent working per week. Assume also that the sample size of each of the five groups is the same. We can look at the means and standard deviations (*SDs*) for each group instead of all of the raw scores. This makes it easier to compare the groups because the mean and *SD* summarize each group in a comparable way.

Group 1:  $M = 40.00$ ,  $SD = 0.00$

Group 2:  $M = 40.00$ ,  $SD = 5.00$

Group 3:  $M = 40.00$ ,  $SD = 20.00$

Group 4:  $M = 20.00$ ,  $SD = 1.00$

Group 5:  $M = 20.00$ ,  $SD = 5.00$

Groups 1, 2, and 3 have the same mean but different *SDs*. The *SD* for Group 1 indicates that everyone worked for 40 hours because there is no deviation (i.e. no differences from the mean). The *SD* for Group 2 indicates that folks worked about 40 hours but that individuals varied. Some worked more, others less. The same is true of Group 3, however, the *SD* for group 3 is very large. This means that scores in Group 3 varied much more greatly around the mean than scores in Group 2. Compare the means and *SDs* in groups 4 and 5. You should notice that the raw scores in Group 4, on average, were more similar to the mean than the raw scores in Group 5 were, on average.

Remember, variables are things that vary. The *SD* is a way to summarize how greatly raw scores of a variable tended to vary from their group's mean. The larger the *SD*, the greater the dispersion. The smaller the *SD*, the lesser the dispersion.

---

This page titled [4.3: Standard Deviation](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .