

Variance and Standard Deviation

Imagine you have a dartboard and a good number of darts. Imagine you throw the darts at the dartboard one at a time and aim each time for the bull's eye at the center. Let's look at some possible kinds of arrangements of the darts as a loose introduction to the idea of variability (of which standard deviation and variance are measures). Let's look at possible distributions of the darts over the dartboard.

If you were a precise dart-thrower, there might be a clustering of darts around the center. In this case we'd say there was relatively little variability in the distribution of darts – this is a way of saying that most of them fell in the same vicinity.

If you weren't as precise, it may look as if the darts were haphazardly arranged over the dartboard. If they were spread out all over the board, and there was little clustering, we'd say that the distribution of darts had a higher variability than in the first case.

Variability is a term that describes how spread out a distribution of scores (or darts) is. **Variance** and **standard deviation** are closely related ways of measuring, or quantifying, variability. [Standard deviation is simply the square root of variance; these concepts will be explained shortly.]

Finishing with the dartboard example, it is not necessary for the darts to cluster around the center in order to have low variability. If the darts clustered around any part of the dartboard, those darts would still have low variability. And the more closely packed they were, the lower the variability would be. We could quantify the degree of variability by assigning numerical values to the positions of the darts in some way, but that's not our concern just now. For now, it's enough to note that the variance and standard deviation are measures of the average amount of variability in where the darts land.

The example above was a rather simple way of talking about variability, but the concept of variability isn't much more complicated than that – even when applied to concrete numbers. By remembering a few easy formulas and a few easy ideas, you'll be a master at computing variances and standard deviations.

General problem: You are given a set of n numbers (or scores) and you are asked to find the **sum of squared deviations**, **variance**, and **standard deviation** of that set of numbers. Here is good way to proceed; examples will follow.

- 1) **Find the mean** (or arithmetic average) of the scores. To find the mean, add up the scores and divide by n where n is the number of scores.



Visit us on the web at www.uhv.edu/ac

Email at tutor@uhv.edu University West, room 129 (361) 570-4288

- 2) **Find the sum of squared deviations** (abbreviated SSD). To get the SSD, find the sum of the squares of the differences (or deviations) between the mean and all the individual scores.
- 3) **Find the variance.** If you are told that the set of scores constitute a population, divide the SSD by n to find the variance. If instead you are told, or can infer, that the set of scores constitute a sample, divide the SSD by $(n - 1)$ to get the variance.
- 4) **Find the standard deviation.** To get the standard deviation, take the square root of the variance.

Notes: It may be convenient to remember that there are two different formulas that can be used to compute the SSD, both yielding identical answers. The first formula, the definitional formula, proceeds directly from the definition of SSD, is easy to remember, and was the one used in the list of steps above: $SSD = \sum_i (m - x_i)^2$ where m is the mean of the scores and the x_i are all the individual scores. The second formula for computing SSD, called the computational formula, is sometimes handy for use in a calculator but harder to remember: $SSD = \sum_i (x_i)^2 - (1/n)(\sum_i x_i)^2$. The first term in this formula directs you to square each score and then add up the resulting values. The second term in this formula directs you to sum the scores, and then square the sum, and then divide the result by n where n is the number of scores. We will illustrate the use of both formulas for finding the SSD.

Example 1: Find the SSD, variance, and standard deviation for the following population of scores: 1, 2, 3, 4, 5 using the list of steps given above.

- 1) Find the mean. The mean of these five numbers (the population mean) is $(1+2+3+4+5)/5 = 15/5 = 3$.
- 2) Let's use the definitional formula for SSD for its calculation: SSD is the sum of the squares of the differences (squared deviations) between the mean and the individual scores. The squared deviations are $(3-1)^2$, $(3-2)^2$, $(3-3)^2$, $(3-4)^2$, and $(3-5)^2$. That is, 4, 1, 0, 1, and 4. The SSD is then $4 + 1 + 0 + 1 + 4 = 10$.
- 3) Divide SSD by n , since this is a population of scores, to get the variance. So the variance is $10/5 = 2$.
- 4) The standard deviation is the square root of the variance. So the standard deviation is the square root of 2.

For practice, let's also compute the SSD using the computational formula, $\sum_i (x_i)^2 - (1/N)(\sum_i x_i)^2$. $\sum_i (x_i)^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 1 + 4 + 9 + 16 + 25 = 55$. $(1/N)(\sum_i x_i)^2 = (1/5)(1 + 2 + 3 + 4 + 5)^2 = (1/5)(15^2) = 45$. So $SSD = 55 - 45 = 10$, just like before.

Example 2: Find the SSD, variance, and standard deviation for the following *sample* of scores: 1, 3, 3, 5.

- 1) The average of these four numbers (the *sample mean*) is $(1+3+3+5)/4 = 12/4 = 3$.
- 2) So, $SSD = (3-1)^2 + (3-3)^2 + (3-3)^2 + (3-5)^2 = 4 + 0 + 0 + 4 = 8$.
- 3) Now, because we were told that these scores constitute a sample, we'll divide SSD by $n-1$ to get the sample variance. In our case we have four scores, so $n = 4$ so $n-1 = 3$. Therefore, our sample variance is $8/3$.



Visit us on the web at www.uhv.edu/ac

Email at tutor@uhv.edu University West, room 129 (361) 570-4288

4) And the sample standard deviation is square root of $8/3$.

As you can see, calculating these statistical measures is quite easy, and it's only a matter of remembering a few formulas and practicing their application.

Further Facts: If a constant number is added to each score in a sample or population, the standard deviation will not change. [This makes sense. Adding a constant will not change how disparate the scores are. The mean will increase by the constant, but the SSD will stay the same, and, consequently, so will the variance and standard deviation.]

Example: Let A be the population of scores $\{1, 3, 3, 5\}$ and let B be the population of scores $\{3, 5, 5, 7\}$. Notice that B is produced by adding 2 to each score in A. The mean of A is 3; the mean of B is 5. The SSD of A is $(3-1)^2 + (3-3)^2 + (3-3)^2 + (3-5)^2 = 4 + 0 + 0 + 4 = 8$; the SSD of B is $(5-3)^2 + (5-5)^2 + (5-5)^2 + (5-7)^2 = 4 + 0 + 0 + 4 = 8$ also. So the SSD of A equals the SSD of B. So the variance and standard deviation of A and B are both the same; they are 2 and square root of 2, respectively.

Also, multiplying each score in a sample or population by a constant factor will multiply the standard deviation by that same factor.

An algebraic explanation: in the original set of scores, to get the SSD we are summing terms of the form $(a-b)^2$, where a is the mean and b is a score. If you multiply each score by a constant, x, the new mean would be xa, and the score that had a value of b would be translated to a score with the value xb. Now, for SSD, we would be summing terms of the form $(xa-xb)^2$, which is $x^2(a-b)^2$. So the SSD is multiplied by x^2 , so the variance is also multiplied by x^2 (since variance is just SSD divided by the number of scores), and so the standard deviation would be multiplied by x, because standard deviation is the square root of the variance.

Example: Let A be the population of scores $\{1, 3, 3, 5\}$ and let B be the population of scores $\{2, 6, 6, 10\}$. Notice that B is produced by multiplying 2 by each score in A. From the preceding paragraph, we know that the mean, SSD, variance, and standard deviation of A are 3, 8, 2, and square root of two, respectively. Let's now compute these for B. The mean of B is 6 (which is twice the mean of A). The SSD of B is $(6-2)^2 + (6-6)^2 + (6-6)^2 + (6-10)^2 = 16 + 0 + 0 + 16 = 32$, which is four (or two squared) times the SSD of A. The variance of B is 8 (or two squared times the variance of A) and the standard deviation of B is square root of eight (which simplifies to two times the square root of two, which is two times the standard deviation of A).

This handout covered the calculation of SSD, variance, and standard deviation. For a comprehensive explanation of the use these concepts, see for example Chapter 4 of *Statistics for the Social Sciences* by Gravetter & Wallnau, a copy of which is owned by the UHV Academic Center located in the UHV Center Building.

Standard deviation is a measure of how far from the mean you would expect an average score to be. The higher the standard deviation, the farther you would expect a randomly chosen score to be from the mean, numerically. Weight is usually more variable in human populations than height, so the standard deviation of the weight variable should be bigger than the standard deviation of the height variable for a population. Also, the



Visit us on the web at www.uhv.edu/ac

Email at tutor@uhv.edu University West, room 129 (361) 570-4288

standard deviation is a way of relating a particular score to the rest of the sample. If a score is within one standard deviation from the mean, it can be considered a pretty ordinary score. On the other hand, if a score is, say, two standard deviations away from the mean (either below or above), then that score can be considered an extreme value. Standard deviation, as the name suggests, helps to give a standardized picture of distributions, in the sense that no matter what the mean or what is being studied, a score that's two standard deviations away from the mean can be considered extreme (in the context of that set of scores). Standard deviation is important to the use of z-scores.

Resources

Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the social sciences* (6th Ed.).

Belmont, CA: Wadsworth.

Copyright 2005 by the Academic Center and the University of Houston-Victoria.
Created 2005 by Hari Damodaran.



Visit us on the web at www.uhv.edu/ac

Email at tutor@uhv.edu University West, room 129 (361) 570-4288