

# How to Compare Standard Deviations: A Step-by-Step Guide with Example

Authored by  
**stats writer**

January 24, 2026

## RECOMMENDED CITATION

stats writer (2026). *How to Compare Standard Deviations: A Step-by-Step Guide with Example*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=127391>

## 1. Introduction: Understanding Standard Deviation and Variability

The standard deviation is a foundational metric in statistics, quantifying the amount of variability or dispersion within a set of data. Essentially, it tells us how spread out the values are relative to the data set's arithmetic mean. A low standard deviation indicates that data points tend to be close to the mean, while a high standard deviation indicates that the data points are spread out over a wider range of values. Understanding this measure is crucial when analyzing the consistency and reliability of data across different samples or populations.

When dealing with multiple data sets--perhaps comparing the performance of two different marketing campaigns, the consistency of two manufacturing lines, or, as in our example, the effectiveness of two study methods--it becomes essential to formally compare their levels of variability. Simply calculating the standard deviation for each set is the first step, but a deeper analysis is often required to determine if the differences observed are statistically significant or merely due to random chance. This comparison helps inform critical decision-making processes where the level of spread or risk needs to be accurately assessed.

Consider a practical scenario: assessing test scores. If Study Method A results in scores with a standard deviation of 5, and Study Method B results in scores with a standard deviation of 10, we immediately see that Method B leads to greater score spread. This suggests that while Method B might produce some very high scores, it is inherently less consistent than Method A. The larger the standard deviation, the greater the scatter, implying less reliability or greater risk depending on the context of the data. Determining whether this difference (10 vs. 5) is mathematically significant requires specialized comparison methods, which we will detail below.

## 2. The Necessity of Comparing Standard Deviations

The need to compare standard deviations often arises in inferential statistics, especially when preparing to use tests that assume equal variances, such as the independent samples T-test. Before pooling data or running comparative tests, researchers must first verify the homogeneity of variances (or homoscedasticity). Since the standard deviation is simply the square root of the variance, comparing standard deviations is mathematically equivalent to comparing variances, though statistical tests typically operate directly on the variance ( $\sigma^2$ ).

Comparing variances allows us to statistically confirm whether two populations exhibit the same degree of dispersion. If the variances are found to be significantly different, it violates the assumptions of many common parametric statistical tests, necessitating the use of alternative test statistics (like Welch's T-test, which does not assume equal variances) or data transformations. Therefore, determining if the population variances ( $\sigma_1^2$  and  $\sigma_2^2$ ) are equal is a prerequisite step in robust data analysis.

We will explore two primary methods used for this comparison: a quick, practical assessment known as the Rule of Thumb, and a highly rigorous, formal method utilizing the F-test. While the Rule of Thumb provides a fast, initial assessment suitable for exploratory data analysis, the F-test offers a precise, hypothesis-driven statistical conclusion, essential for published research or high-stakes decision-making.

## Compare Standard Deviations (With Example)

### 3. Method 1: The Variance Rule of Thumb (A Quick Assessment)

The Rule of Thumb provides a straightforward, although non-statistical, method for assessing whether two data sets likely share approximately equal variances. This method is particularly useful in preliminary data screening when formal statistical testing might be overkill or when a quick decision needs to be made about proceeding with certain assumptions. It relies solely on calculating the ratio between the larger and the smaller sample variance.

The core premise of this heuristic approach is simplicity and practicality. If the ratio calculated is small, it suggests the difference in spread is negligible for practical purposes. Specifically, the generally accepted guideline states that if the ratio of the larger sample variance to the smaller sample variance is less than 4, we can reasonably assume that the population variances are approximately equal (homoscedasticity

holds). This threshold of four is arbitrary but widely adopted in many introductory statistical contexts and serves as a pragmatic guideline.

Conversely, should the calculated ratio be equal to or greater than 4, the observed difference in variance is considered substantial enough to assume that the underlying population variances are not equal (heteroscedasticity). In such cases, one must proceed with caution, acknowledging that tests assuming equal variances may yield unreliable results, thereby requiring the application of alternative methods like the **F-test** for formal confirmation or immediate use of variance-unequal procedures. This simple ratio provides a fast initial filter for variability comparison.

#### 4. Method 2: Employing the Formal F-Test for Equality of Variances

A far more rigorous and statistically sound approach to comparing variances is the use of the **F-test**. This test is a type of hypothesis test that utilizes the F-distribution to determine if the variation observed in two samples is likely to have come from populations with the same variance. Since standard deviation is the square root of variance, this test directly addresses the equality of

standard deviations in a statistically verifiable manner.

The F-test operates under a predefined set of competing hypotheses. The primary goal is to assess the plausibility of the null hypothesis ( $H_0$ ), which posits that the population variances are identical. The alternative hypothesis ( $H_1$ ) suggests they are not equal, meaning one population exhibits significantly more dispersion than the other. These hypotheses are formalized as follows:

$H_0: \sigma_1^2 = \sigma_2^2$  (The population variances are equal, implying equal standard deviations.)

$H_1: \sigma_1^2 \neq \sigma_2^2$  (The population variances are *not* equal, implying unequal standard deviations.)

To execute the F-test, we calculate the F-statistic, which is derived from the ratio of the two sample variances. This calculated F-statistic is then compared against the F-distribution to obtain a p-value. Modern statistical practice almost universally relies on dedicated statistical software (such as R, Python, SAS, or Excel) to compute this test accurately, as manual calculation

involving F-distribution tables can be cumbersome and prone to error, especially given the various degrees of freedom involved in defining the critical regions.

#### 5. Detailed Example Setup and Data Presentation

To illustrate both comparison methods effectively, let us use a concrete example involving educational data. We compare the exam scores of students who utilized two distinct study methodologies, Method 1 and Method 2, to prepare for a standardized test. Our objective is to determine whether these two methods result in scores with statistically similar levels of dispersion (i.e., whether the standard deviations are equal).

The raw data collected for this hypothetical study is presented below. Notice that both groups contain  $n=15$  students. The scores are quantitative continuous data, making them ideal for variance analysis. Before any sophisticated testing, it is often insightful to visually inspect the raw data to get a preliminary sense of the spread, though this intuition must always be validated by formal statistical measures to ensure objectivity.

This visualization represents the collected scores for both Study Method 1 and Study Method 2, providing the essential input for our subsequent calculations and tests.

Method 1	Method 2
68	77
70	80
71	81
72	81
74	82
74	83
78	83
82	84
83	84
88	85
90	88
92	89
93	90
96	92
97	95

## 6. Application of Method 1: Step-by-Step Analysis

The first step in applying the Rule of Thumb is to calculate the sample variance ( $s^2$ ) for each dataset. Recall that the sample variance is the average squared deviation of each data point from the sample mean. These calculations are necessary because we are

comparing the variances, not the standard deviations directly, although the conclusion pertains to both measures of spread.

After calculation (typically done via spreadsheet software or statistical packages), we obtain the following results for the sample means, standard deviations, and variances:

	Method 1	Method 2
	68	77
	70	80
	71	81
	72	81
	74	82
	74	83
	78	83
	82	84
	83	84
	88	85
	90	88
	92	89
	93	90
	96	92
	97	95
$s^2$	103.41	24.21

Once the variances are established ( $s^2_{\text{Method 1}}$  approx 90.95\$ and  $s^2_{\text{Method 2}}$  approx 21.24\$), we

apply the Rule of Thumb by determining the ratio of the larger variance to the smaller variance. In this case, the calculation is  $\$90.95 / 21.24$  approx 4.28\$. Since this calculated ratio (4.28) is greater than the established threshold of 4, the Rule of Thumb strongly suggests that the variances are not equal. Consequently, we would preliminarily conclude that the standard deviations between the two study method score sets are unequal, necessitating the use of statistical procedures designed for unequal variances. This initial finding, though not formal proof, provides a strong indication of heteroscedasticity.

#### 7. Application of Method 2: Interpreting the R F-Test Results

For a definitive and statistically robust conclusion regarding the equality of standard deviations, we must perform the formal F-test. We use R, a powerful environment for statistical computing, to execute the test on our two sets of scores. The resulting output below provides all the necessary components for hypothesis testing.

The output provided by the statistical software is critical for making a statistically informed decision

regarding the null hypothesis and confirms the findings suggested by the Rule of Thumb.

**#enter exam scores for both groups of students**

```
method1 <- c(68, 70, 71, 72, 74, 74, 78, 82, 83, 88, 90, 92, 93, 96, 97)
```

```
method2 <- c(77, 80, 81, 81, 82, 83, 83, 84, 84, 85, 88, 89, 90, 92, 95)
```

**#perform an F-test to determine if the variances are equal**

```
var.test(method1, method2)
```

**F test to compare two variances**

**data: method1 and method2**

**F = 4.2714, num df = 14, denom df = 14, p-value = 0.01031**

**alternative hypothesis: true ratio of variances is not equal to 1**

**95 percent confidence interval:**

**1.434049 12.722857**

**sample estimates:**

**ratio of variances**

**4.27144**

The key results extracted from the R output are the F-test statistic and the corresponding p-value. The F-statistic (4.2714) is derived from the ratio of the sample variances. The p-value (0.01031) represents the probability of observing a variance ratio this extreme or more extreme, assuming that the null hypothesis is true.

F-test statistic: 4.2714

p-value: .01031

#### 8. Conclusion: Deciding on Variance Equality

The decision to accept or reject the null hypothesis ( $H_0: \sigma_1^2 = \sigma_2^2$ ) is made by comparing the p-value to a predetermined significance level ( $\alpha$ ), typically set at 0.05. The significance level represents the maximum risk we are willing to accept of incorrectly rejecting a true null hypothesis (Type I Error).

In our analysis, the obtained p-value is 0.01031. Since 0.01031 is less than 0.05, we have statistically sufficient evidence to reject the null hypothesis. Rejecting the null hypothesis leads us to accept the alternative

**hypothesis ( $H_1$ ), which states that the population variances are not equal. This conclusion confirms the initial finding suggested by the Rule of Thumb, providing a statistically robust confirmation that the variability differs significantly between the two methods.**

**Therefore, the definitive conclusion is that the standard deviations between the exam scores resulting from the two study methods are statistically unequal. Study Method 1, with the higher variance, exhibits significantly greater score dispersion compared to Study Method 2. This difference in variability is an important finding, suggesting that while the average score might be similar, the outcomes of Study Method 1 are much less predictable, a crucial piece of information for researchers and educators.**

#### **Further Resources on Variability Analysis**

**This detailed example demonstrates that while a simple comparison of calculated standard deviations provides an intuitive understanding of variability, a formal statistical test, such as the F-test, is essential for determining if the observed difference is statistically**

significant across the populations from which the samples were drawn.

**Bonus:** Specialized online calculators, such as the Statology F-Test Calculator, are also available for performing this variance comparison quickly and reliably without needing direct coding access to statistical software.

The following tutorials provide additional information about using standard deviations and related metrics in statistical analysis, helping solidify the understanding of dispersion and risk assessment: