

How do you perform an F-Test in SAS?

Authored by
stats writer

November 19, 2025

RECOMMENDED CITATION

stats writer (2025). *How do you perform an F-Test in SAS?*. PSYCHOLOGICAL SCALES.
Retrieved from <https://scales.arabpsychology.com/?p=96983>

The **F-Test** is a fundamental statistical tool utilized across many fields, particularly within the **SAS** statistical software environment. While it can be executed using advanced procedures like **PROC GLM** (General Linear Model) for testing complex model comparisons or specific contrasts, the most straightforward application for comparing two group variances often involves **PROC TTEST**, which conveniently provides F-test results by default. When employing **PROC GLM**, successful execution requires careful specification of the statistical model within the MODEL statement, followed by explicitly defining the **null and alternative hypotheses** using the TEST statement to derive the necessary F-statistic. The resulting output furnishes crucial metrics, including the calculated F-value, the associated degrees of freedom, and the critical **p-value**, which dictates the decision regarding the equality of the tested variances.

Understanding the F-Test: A Test of Variances

At its core, the **F-Test** is formally defined as a hypothesis test used to determine whether the **population variances** (σ^2) of two independent groups or samples are statistically equal. This test is essential for validating underlying assumptions required by many subsequent statistical analyses, such as the standard Student's t-test or Analysis of Variance (ANOVA). Understanding the core premise--that variability across groups is comparable--ensures that the subsequent analysis, which often focuses on differences in means, is interpreted correctly. The ratio of the variances forms the basis of the F-statistic, which follows an F-distribution characterized by two separate degrees of freedom parameters.

The primary purpose of the two-sample F-test for variances is encapsulated by its formal definition of the **null and alternative hypotheses**. These hypotheses precisely define the statistical claim being tested against the available sample data. Given the fundamental nature of this test in preempting further analysis, it is critical to correctly establish these competing statements before executing the procedure within **SAS** or any other statistical platform. The conclusion drawn from the F-test determines which version of downstream tests, such as pooled or unpooled t-tests, is appropriate for comparing means, thereby maintaining the integrity of the overall statistical investigation.

The hypotheses for the two-tailed F-Test comparing two **population variances** are mathematically expressed as follows:

H0: $\sigma_1^2 = \sigma_2^2$ (The population variances are considered statistically equal.)

HA: $\sigma_1^2 \neq \sigma_2^2$ (The population variances are statistically *unequal*, indicating significant difference in variability.)

Key Applications and Use Cases of the F-Test

The **F-Test** serves several critical roles in applied statistics, primarily addressing questions about data dispersion and variability rather than central tendency. One of the most common applications involves checking the homogeneity of variance assumption--a mandatory prerequisite for numerous parametric tests, including the independent samples t-test. If this assumption is violated, meaning we reject the **null hypothesis** that variances are equal, statistical procedures must be adjusted, typically by using correction factors like the Satterthwaite method when conducting t-tests, or by employing non-parametric alternatives.

Specifically, the **F-Test** is frequently employed to answer practical questions concerning dataset properties and experimental outcomes. These questions often revolve around quality control, process optimization, or initial data validation before complex modeling. Researchers often apply this test early in the data exploration phase to gain insight into the inherent spread of measurements within different treatment groups or categories. This early insight can significantly influence the selection of appropriate methodologies and the final interpretation of experimental results, ensuring that conclusions are robust and statistically sound.

Common scenarios where the F-test provides indispensable statistical evidence include:

Do two independent samples originate from populations possessing equivalent levels of variability, ensuring comparability for mean testing?

Has a newly introduced treatment, intervention, or manufacturing process successfully reduced the inherent variability (risk or spread) associated with an existing, baseline process?

Executing the F-Test in SAS using PROC TTEST

While the **SAS** environment offers multiple procedures for performing the F-Test, the most accessible and commonly used method for comparing the variances of two samples is through the **PROC TTEST** statement. Although primarily designed for calculating t-statistics to compare population means, this robust procedure calculates and reports the F-test for the equality of variances automatically as a standard diagnostic output. This integrated approach simplifies the workflow significantly, allowing researchers to simultaneously assess both the equality of means and the required assumption of variance homogeneity using a single block of code.

The structure of the **PROC TTEST** syntax is highly efficient. It requires specifying the input dataset, defining the grouping variable (the independent variable) using the CLASS statement, and identifying the continuous measurement variable (the dependent variable) using the VAR statement. This clear structure ensures that the procedure correctly partitions the data into the two comparison groups, Team A and Team B in our example, and calculates the variances required for

the F-statistic ratio. Utilizing this procedure streamlines the initial data analysis steps, making it the preferred method for quick and reliable two-sample comparisons in **SAS** programming.

The following detailed example illustrates the step-by-step process of preparing data and using **PROC TTEST** to execute the variance F-test, providing a practical demonstration of how to interpret the resulting statistical measures, including the F-statistic and its associated **p-value**.

Detailed Example: Preparing the Basketball Dataset in SAS

To demonstrate the practical application of the F-test in **SAS**, let us consider a hypothetical dataset concerning sports statistics. Suppose we have collected data representing the points scored by various basketball players, categorized across two distinct teams. The objective is to determine whether the variability in individual player performance (points scored) differs significantly between these two teams. A large difference in variance would suggest that one team exhibits greater inconsistency in scoring than the other, which is a critical piece of information for coaching staff or performance analysts.

The first step in **SAS** involves creating and populating the dataset, usually using a DATA step followed by the DATALINES statement to input the raw observations directly into the system. This structured data preparation ensures that the subsequent statistical procedure (PROC TTEST) can correctly identify the grouping variable ('team') and the continuous variable for analysis ('points'). Proper data input is foundational to obtaining reliable statistical results, and viewing the printed dataset confirms data integrity.

The following code block demonstrates the creation of the dataset named my_data, followed by the use of proc print to display the data structure and verify correct entry:

```
/*create dataset*/  
data my_data;  
input team $ points;  
datalines;  
A 18  
A 19  
A 22  
A 25  
A 27  
A 28  
A 41  
A 45  
A 51  
A 55
```

B 14

B 15

B 15

B 17

B 18

B 22

B 25

B 25

B 27

B 34

;

run;

```
/*view dataset*/
```

```
proc print data=my_data;
```

After running the data creation and printing steps, the generated output confirms the structure of the dataset, illustrating the assignment of points to their respective teams. This visual confirmation is crucial for debugging and ensuring that the statistical procedures are applied to the correct variables, preparing the environment for the subsequent F-test analysis.

Obs	team	points
1	A	18
2	A	19
3	A	22
4	A	25
5	A	27
6	A	28
7	A	41
8	A	45
9	A	51
10	A	55
11	B	14
12	B	15
13	B	15
14	B	17
15	B	18
16	B	22
17	B	25
18	B	25
19	B	27
20	B	34

Executing the F-Test Code and Obtaining Results

Our specific goal is to execute the **F-Test** to statistically determine if the variability in points scored is equivalent across Team A and Team B, thereby testing the equality of the two **population variances**. As previously established, the most efficient method in **SAS** for this comparison is the **PROC TTEST** procedure, which handles the necessary calculations automatically.

The required syntax is straightforward and highly standardized. It begins with the invocation of **PROC TTEST**, specifying the dataset name, `my_data`. The crucial part involves the **CLASS** statement, which identifies the grouping variable (`team`), and the **VAR** statement, which specifies the measurement variable (`points`) upon which the variances will be compared. SAS then performs the internal computations, resulting in an output table dedicated specifically to testing the assumption of variance homogeneity, which is essential for two-sample comparisons.

Use the following minimal syntax to instruct **PROC TTEST** to perform the required statistical comparison:

```
/*perform F-test for equal variances*/  
proc ttest data=my_data;  
class team;  
var points;  
run;
```

Upon execution, **PROC TTEST** generates several output tables, including descriptive statistics for each group and the results of the t-tests. However, the section most relevant to the variance test is typically labeled "Equality of Variances," which summarizes the outcomes of the F-test and provides the statistical basis for our decision.

Interpreting the F-Test Results in SAS Output

The comprehensive output generated by **PROC TTEST** includes a dedicated section titled **Equality of Variances**. This table presents the essential statistics derived from the F-test, allowing us to formally test the **null hypothesis** that the variability across the two basketball teams is identical. Interpreting this table is critical for concluding the variability comparison and determining the appropriate path for any subsequent mean comparison.

The TTEST Procedure

Variable: points

team	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
A		10	33.1000	13.6744	4.3242	18.0000	55.0000
B		10	21.2000	6.5286	2.0645	14.0000	34.0000
Diff (1-2)	Pooled		11.9000	10.7147	4.7918		
Diff (1-2)	Satterthwaite		11.9000		4.7918		

team	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
A		33.1000	23.3179 42.8821	13.6744	9.4057 24.9641
B		21.2000	16.5297 25.8703	6.5286	4.4906 11.9186
Diff (1-2)	Pooled	11.9000	1.8329 21.9671	10.7147	8.0962 15.8452
Diff (1-2)	Satterthwaite	11.9000	1.5399 22.2601		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	18	2.48	0.0231
Satterthwaite	Unequal	12.9	2.48	0.0276

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	4.39	0.0383

The table provides two key figures necessary for drawing a statistical conclusion:

The calculated F-Test statistic, which is the ratio of the sample variances. In this specific output, the statistic is reported as **4.39**.

The corresponding **p-value** (Pr > F), which measures the probability of observing the current data, or data more extreme, assuming the null hypothesis of equal variances is true. Here, the p-value is **0.0383**.

To make a decision, we compare the calculated **p-value** to a pre-determined significance level (α), typically set at 0.05. Since 0.0383 is less than 0.05, we possess statistically significant evidence to reject the **null hypothesis**. This powerful rejection implies that we must conclude that the **population variances** in points scored between the two teams are, in fact, unequal. Practically, this means Team A and Team B do not exhibit the same level of consistency in player scoring performance, with one team showing significantly more dispersion in individual scores.

Implications for Further Statistical Analysis

The finding that the variances are unequal has significant methodological implications, especially if

the research objective extends beyond variance comparison to include a test of means (a standard two-sample t-test). When the assumption of homogeneity of variance is violated, the traditional pooled variance t-test becomes inappropriate because it assumes a single, common estimate of variance exists for both groups, leading to potentially inaccurate standard error calculations and misleading p-values.

In such scenarios, researchers must pivot to a corrected test statistic that does not require the assumption of equal variances. In the context of the **PROC TTEST** output, this correction is implemented through the **Satterthwaite approximation**, which adjusts the degrees of freedom downward to account for the disparity in variances. Therefore, if one were to subsequently perform a two-sample t-test to assess whether the mean points values are equal between the two teams, the correct p-value to reference would be found in the row labeled **Satterthwaite**, rather than the pooled variance row, ensuring the validity of the hypothesis testing regarding the means.

Additional SAS Tutorials for Data Analysis

Mastering the **F-Test** is a crucial step in advanced statistical analysis using **SAS**. For users seeking to broaden their proficiency, the following structured tutorials and resources explain how to perform other common and essential data manipulation and inferential tasks within the SAS programming environment, building upon the foundational knowledge demonstrated here for hypothesis testing: