

Perform Welch's t-Test in SAS

Authored by
Mohammed looti

November 1, 2025

RECOMMENDED CITATION

Mohammed looti (2025). *Perform Welch's t-Test in SAS*. PSYCHOLOGICAL STATISTICS.
Retrieved from <https://statistics.arabpsychology.com/?p=7623>

The Necessity of Welch's t-Test in Statistical Analysis

The [Welch's t-test](#) stands as a cornerstone statistical procedure, primarily utilized for comparing the means derived from two **independent groups**. This test is a critical modification of the classical Student's t-test, specifically engineered to handle complex scenarios often encountered in real-world data analysis where underlying population characteristics may differ significantly between the groups being studied.

Unlike its predecessor, the Welch test eliminates the restrictive assumption that the two populations must possess **equal variances**. This condition, often referred to as homoscedasticity, is frequently violated in practice, leading to a state known as heteroscedasticity. By relaxing this stringent requirement, the Welch test offers a far more robust and reliable method for accurately assessing mean differences, especially when sample sizes are dramatically unequal or when the variability within each group varies substantially.

Utilizing the appropriate statistical tool is essential, as ignoring unequal variances can result in distorted statistical inferences, specifically leading to an inflation of [Type I error rates](#) (false positives). Consequently, the Welch test is indispensable across fields such as social sciences, medical research, and applied statistics. This comprehensive tutorial will detail the step-by-step process of executing and interpreting the [Welch's t-test](#) using the widely respected statistical software package, [SAS](#).

Assumptions of the Standard t-Test and the Problem of Heteroscedasticity

Standard parametric hypothesis tests, including the basic two-sample t-test, rely on several foundational mathematical assumptions for their validity. Key among these are the assumption of normality--meaning the data within each group should be normally distributed--and the crucial assumption of homogeneity of variances. When researchers conduct preliminary data checks and discover that the variances between the comparison groups are not equal, the results generated by the standard t-test become statistically unreliable, potentially leading to incorrect conclusions.

The fundamental innovation of the [Welch's t-test](#) lies in its sophisticated method for calculating the degrees of freedom. Instead of using a simple pooled variance estimate, the Welch test employs the [Satterthwaite approximation](#). This adjustment allows the test statistic to remain accurate even when the assumption of equal variance is severely violated, thereby making the test exceptionally robust against this common data challenge.

Therefore, selecting the Welch test is not merely a preference but a statistical necessity in certain scenarios. If prior diagnostic testing, such as the Levene's test or the preliminary F-test for equality of variances, yields a statistically significant result (indicating heterogeneity), the standard procedure should be abandoned in favor of the Welch modification. By choosing the [Welch's t-test](#),

analysts ensure that their final inference about the mean difference is correctly weighted based on the observed variability within the two [independent groups](#), maintaining the integrity and accuracy of the statistical conclusion.

Case Study: Preparing Data for Score Comparison in SAS

To demonstrate the practical application of the [Welch's t-test](#) within the [SAS](#) environment, we will examine a clear, hypothetical scenario rooted in educational research. Imagine a study designed to evaluate the efficacy of a new exam preparation booklet. The objective is to statistically compare the final exam performance of one group of 12 students who actively used the prep booklet against a control group of 12 students who did not utilize it.

The core research question centers on whether a statistically significant difference exists in the average (mean) exam scores between these two distinct groups. Given that study habits, prior knowledge, and motivation levels are likely to differ between participants--potentially leading to disparate score distributions and unequal variances--the Welch's t-test is the most appropriate analytical method to employ for a fair and accurate comparison.

The initial step requires structuring the collected raw data into a format that [SAS](#) can process effectively. We must transform the independent score lists into a long-format dataset, meaning we need two key variables: a categorical variable (`group`) identifying the treatment level, and a continuous variable (`score`) representing the outcome measure. The scores collected for the two groups are listed below, followed by the requisite SAS code to prepare the data:

Booklet Group Scores: 90, 85, 88, 89, 94, 91, 79, 83, 87, 88, 91, 90

No Booklet Group Scores: 67, 90, 71, 95, 88, 83, 72, 66, 75, 86, 93, 84

The following code segment utilizes the **DATA** and **DATALINES** statements to create the necessary dataset named `exam_scores`, making the information ready for statistical processing.

```
/*create dataset*/  
data exam_scores;  
input group $ score;  
datalines;  
booklet 90  
booklet 85  
booklet 88  
booklet 89  
booklet 94  
booklet 91
```

```
booklet 79
booklet 83
booklet 87
booklet 88
booklet 91
booklet 90
no_booklet 67
no_booklet 90
no_booklet 71
no_booklet 95
no_booklet 88
no_booklet 83
no_booklet 72
no_booklet 66
no_booklet 75
no_booklet 86
no_booklet 93
no_booklet 84
;
run;
```

This careful preparation ensures that the subsequent statistical procedure, the T-Test, can correctly identify the groups and calculate the required test statistics based on the provided input structure.

Executing the Analysis using PROC TTEST

With the dataset successfully defined and populated, the next logical step is to invoke the statistical procedure designed for mean comparison: [PROC TTEST](#). This powerful procedure in SAS is highly versatile; it is designed not only to perform the standard two-sample t-test but also, crucially, to automatically generate the results for the [Welch's t-test](#) simultaneously, thereby accommodating both the equal and unequal variance scenarios.

The code requires defining three essential components. First, we specify the source of the data using the `data=` option, pointing to our newly created `exam_scores` dataset. Second, the `class` statement is used to designate the categorical predictor variable, `group`, which defines the two independent populations we are comparing. Third, the `var` statement identifies the continuous outcome variable, `score`, upon which the mean comparison will be performed. Finally, for robustness and clarity, we explicitly set the significance level (alpha) at the conventional 0.05

threshold.

By default, when `PROC TTEST` runs, it calculates the necessary statistics for two versions of the t-test: one assuming equal variances (Pooled) and one adjusting for unequal variances (Satterthwaite/Welch). This dual output allows the researcher to make an informed decision based on the preliminary test for variance equality, which is also provided in the output. The following simple procedure initiates the analysis:

```
/*perform Welch's t-test*/  
proc ttest data=exam_scores alpha=0.05;  
class group;  
var score;  
run;
```

Executing this code generates a comprehensive report that moves beyond simple descriptive statistics, providing the critical inferential statistics necessary for drawing a valid conclusion about the effect of the exam preparation booklet.

Detailed Interpretation of SAS Output (F-Test and Welch's Results)

The output generated by `PROC TTEST` is structured logically, providing descriptive summaries followed by the critical inferential tables needed for hypothesis testing. The analysis requires focusing on two main tables: the test for variance equality and the final T-Tests table.

The first crucial step in interpretation involves examining the table titled **Equality of Variances**. This section presents the results of an [F-test](#), designed to rigorously determine if the variances of the two samples are statistically equal. This test addresses the core assumption underlying the choice between the Student's and Welch's t-tests. The associated hypotheses for this initial variance check are clearly defined:

H₀: The population variances are equal (Homogeneity holds).

H_A: The population variances are not equal (Heterogeneity exists).

For our educational example, we must scrutinize the [p-value](#) reported for the F-test. If this value is below our predetermined alpha level of 0.05, we must reject the **null hypothesis** (H_0). In this specific instance, the calculated p-value is **0.0046**. Since 0.0046 is significantly less than 0.05, we confidently reject the null hypothesis, providing strong statistical evidence that the variability in scores differs significantly between the booklet group and the no-booklet group.

The TTEST Procedure

Variable: score

group	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
booklet		12	87.9167	4.0330	1.1642	79.0000	94.0000
no_bookl		12	80.8333	10.2055	2.9461	66.0000	95.0000
Diff (1-2)	Pooled		7.0833	7.7594	3.1678		
Diff (1-2)	Satterthwaite		7.0833		3.1678		

group	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
booklet		87.9167	85.3542	90.4791	4.0330	2.8570	6.8476
no_bookl		80.8333	74.3491	87.3176	10.2055	7.2295	17.3276
Diff (1-2)	Pooled	7.0833	0.5138	13.6529	7.7594	6.0011	10.9823
Diff (1-2)	Satterthwaite	7.0833	0.3048	13.8618			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	22	2.24	0.0358
Satterthwaite	Unequal	14.354	2.24	0.0417

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	11	11	6.40	0.0046

The rejection of the equal variances assumption confirms that the standard pooled t-test results would be inaccurate. We must, therefore, direct our attention exclusively to the row labeled **Unequal** in the subsequent T-Tests table, as this row contains the results derived from the Satterthwaite approximation--the [Welch's t-test](#) result. Focusing on this row, we extract the calculated test statistic and the associated probability value:

t Value: **2.24**

[p-value](#): **0.0417**

These values pertain to the primary hypotheses concerning the population means:

H₀: $\mu_1 = \mu_2$ (The mean scores of both groups are equal.)

H_A: $\mu_1 \neq \mu_2$ (The mean scores are not equal; a significant difference exists.)

Finally, we compare the Welch's [p-value](#) (**0.0417**) against the 0.05 significance level. Because 0.0417 is less than 0.05, we reject the **null hypothesis** (H₀). This final statistical decision leads to the conclusion that there is sufficient evidence to assert that the mean exam score of students who

used the prep booklet is statistically and significantly different from those who did not, having appropriately controlled for the observed differences in score variability.

Conclusion and Recommended Next Steps

The successful execution of the Welch's t-test in [SAS](#) provides a clear and statistically sound answer to our research question. The analysis confirms that the exam preparation booklet had a statistically measurable and significant effect on student performance. Crucially, by employing the Welch correction, we ensured that this finding remained reliable despite the violation of the equal variance assumption, which was clearly demonstrated by the preliminary [F-test](#) results.

Researchers should always prioritize the Welch test when analyzing data that exhibits heteroscedasticity, especially when sample sizes are uneven. This methodological rigor strengthens the credibility of the findings and minimizes the risk of drawing false conclusions based on inappropriate statistical models.

For those looking to practice or verify their results outside of the SAS environment, several resources are available. **Bonus Tool:** Feel free to use this [online calculator](#) to automatically perform Welch's t-test for any two samples, allowing for quick checks and validation of manual calculations.

To further enhance your statistical programming skills in SAS, consider exploring tutorials on related hypothesis tests:

The following tutorials explain how to perform other common statistical tests in SAS: