

How to Easily Calculate Confidence Intervals in SAS

Authored by
stats writer

November 22, 2025

RECOMMENDED CITATION

stats writer (2025). *How to Easily Calculate Confidence Intervals in SAS*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=99629>

Introduction to Calculating Confidence Intervals Using SAS

The calculation of confidence intervals (CIs) is a fundamental practice in statistical analysis, providing a range of plausible values for an unknown population parameter. In the context of SAS, these calculations are streamlined using powerful procedures such as **PROC TTEST** or, less commonly for basic intervals, **PROC MEANS**. A confidence interval is a calculated estimate that gives us a certain level of certainty--typically 90%, 95%, or 99%--that the true population value lies within the specified bounds. Understanding how to correctly implement the SAS syntax is essential for accurate inferential statistics.

While **PROC MEANS** can calculate confidence limits using the `CI` option (requiring specific statistical inputs like mean, standard deviation, and sample size), the PROC TTEST procedure is generally the most straightforward and robust method for deriving confidence intervals for population means, especially when performing related hypothesis testing simultaneously. This tutorial will focus on utilizing PROC TTEST due to its simplicity and comprehensive output, which includes critical information needed for interpreting the results, such as degrees of freedom and variance tests.

This guide will demonstrate the precise steps for calculating two common types of confidence intervals in SAS: the CI for a single population mean and the CI for the difference between two population means. By defining the necessary data structure, specifying the desired alpha level (e.g., `ALPHA=0.05` for 95% confidence), and correctly using the `VAR` and `CLASS` statements, users can quickly generate reliable estimates essential for research and data-driven decision-making.

A confidence interval is a range of values that is likely to contain a population parameter with a certain level of confidence.

This tutorial explains how to calculate the following critical confidence intervals in SAS:

Confidence Interval for a Population Mean

Confidence Interval for a Difference in Population Means

Let's jump into the practical examples!

Example 1: Confidence Interval for a Single Population Mean in SAS

Our first example focuses on calculating a 95% confidence interval for the true population mean height derived from a single random sample. This scenario is common when attempting to estimate a population characteristic based on limited sample data. We begin by defining our dataset, which contains the heights (in inches) of 12 randomly selected plants belonging to a specific species. This step ensures that SAS has the necessary variable and observation structure

to perform the subsequent calculations.

The data setup involves using the standard SAS data step. The following code snippet demonstrates the creation and immediate viewing of the dataset named `my_data`, which includes a single variable: `Height`. Note that the `datalines` section provides the observed measurements for the sample, which will serve as the input for our statistical procedure.

```
/*create dataset*/
```

```
data my_data;
```

```
input Height;
```

```
datalines;
```

```
14
```

```
14
```

```
16
```

```
13
```

```
12
```

```
17
```

```
15
```

```
14
```

```
15
```

```
13
```

```
15
```

```
14
```

```
;
```

```
run;
```

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

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

The visual representation of this data table confirms the successful loading of the 12 observations into the SAS environment, ready for statistical processing.

Obs	Height
1	14
2	14
3	16
4	13
5	12
6	17
7	15
8	14
9	15
10	13
11	15
12	14

Applying PROC TTEST for Single Mean Estimation

Suppose we would like to calculate a 95% confidence interval for the true population mean height of this species. With the dataset successfully loaded, we now employ the `PROC TTEST` procedure to generate the 95% confidence interval for the population mean height. The syntax is straightforward: we specify the data set, set the desired alpha level, and use the `VAR` statement to indicate the variable of interest, which is `Height`. Setting `ALPHA=0.05` ensures that the output provides the boundaries for a 95% CI ($1 - 0.05 = 0.95$).

The code below instructs SAS to execute the t-test procedure, automatically producing the descriptive statistics and the confidence limits for the mean based on the sample data.

```
/*generate 95% confidence interval for population mean*/  
proc ttest data=my_data alpha=0.05;  
var Height;  
run;
```

Upon execution, `SAS` generates a comprehensive output table detailing the statistical summary.

The TTEST Procedure

Variable: Height

N	Mean	Std Dev	Std Err	Minimum	Maximum
12	14.3333	1.3707	0.3957	12.0000	17.0000

Mean	95% CL Mean	Std Dev	95% CL Std Dev
14.3333	13.4624 15.2042	1.3707	0.9710 2.3273

DF	t Value	Pr > t
11	36.22	<.0001

Interpreting the Output for the Single Mean Confidence Interval

The primary table generated by `PROC TTEST` under Example 1 contains all the necessary information for interpretation. The value for **Mean** shows the calculated sample mean from our dataset, which is the point estimate for the population mean. Crucially, the values under **95% CL Mean** (Confidence Limits for the Mean) provide the lower and upper bounds of the specified 95% confidence interval.

Based on the provided output image, we observe that the calculated 95% confidence interval for the mean height of plants in this species is precisely . This result allows us to make an inferential statement: we are 95% confident that the true average height of all plants belonging to this population falls somewhere between 13.4624 inches and 15.2042 inches. This interval quantifies the uncertainty inherent in estimating a population parameter from a finite sample.

It is important to understand that the confidence level reflects the long-run frequency of the interval capturing the true parameter. This certainty level is determined by the `ALPHA=` value we specified in the SAS code, linking the input command directly to the statistical conclusion.

Example 2: Confidence Interval for the Difference in Two Population Means in SAS

Suppose we would like to calculate a 95% confidence interval for the difference in population mean height between species A and species B. For this second example, we aim to calculate the 95% confidence interval for the difference in population mean height between two species of plants, designated as Species A and Species B. This requires a dataset that includes both the measurement variable (Height) and a grouping variable (Species).

We start by creating the dataset, `my_data2`, which contains observations for both Species A and Species B. The grouping variable `Species` is defined as a character variable using the dollar sign (\$) in the `input` statement, which is critical for SAS to recognize it as a classification variable.

```
/*create dataset*/  
data my_data2;  
input Species $ Height;  
datalines;  
A 14  
A 14  
A 16  
A 13  
A 12  
A 17  
A 15  
A 14  
A 15  
A 13  
B 15  
B 14  
B 19  
B 19  
B 17  
B 18  
B 20  
B 19  
B 17  
B 15  
;  
run;  
  
/*view dataset*/  
proc print data=my_data2;
```

The output image confirms the data structure, clearly showing the two distinct groups and their corresponding height measurements, setting the stage for the comparative analysis.

Obs	Species	Height
1	A	14
2	A	14
3	A	16
4	A	13
5	A	12
6	A	17
7	A	15
8	A	14
9	A	15
10	A	13
11	B	15
12	B	14
13	B	19
14	B	19
15	B	17
16	B	18
17	B	20
18	B	19
19	B	17
20	B	15

Executing PROC TTEST for Two Independent Samples

Before running the main procedure, we first use **PROC SORT** to order the data by the classification variable `Species`. This practice is highly recommended to ensure the **PROC TTEST** calculation handles the groups correctly. We then apply **PROC TTEST**, using the `CLASS` statement to identify the grouping variable and the `VAR` statement for the measurement variable, `Height`. We maintain `ALPHA=0.05` for the 95% confidence interval.

We can use the following code in **SAS** to do so:

```
/*sort data by Species to ensure confidence interval is calculated correctly*/
```

```
proc sort data=my_data2;
```

```
by Species;
```

```
run;
```

```
/*generate 95% confidence interval for difference in population means*/
```

```
proc ttest data=my_data2 alpha=0.05;
class Species;
var Height;
run;
```

The resulting output provides an extensive set of statistics, including individual group means, variance estimates, and the crucial test for the Equality of Variances.

The TTEST Procedure

Variable: Height

Species	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
A		10	14.3000	1.4944	0.4726	12.0000	17.0000
B		10	17.3000	2.0575	0.6506	14.0000	20.0000
Diff (1-2)	Pooled		-3.0000	1.7981	0.8042		
Diff (1-2)	Satterthwaite		-3.0000		0.8042		

Species	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
A		14.3000	13.2309	15.3691	1.4944	1.0279	2.7283
B		17.3000	15.8281	18.7719	2.0575	1.4152	3.7562
Diff (1-2)	Pooled	-3.0000	-4.6895	-1.3105	1.7981	1.3587	2.6591
Diff (1-2)	Satterthwaite	-3.0000	-4.7011	-1.2989			

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	18	-3.73	0.0015
Satterthwaite	Unequal	16.429	-3.73	0.0017

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	9	9	1.90	0.3547

Analyzing Variance Equality and Selecting the Correct CI

The first table we need to look at in the output is **Equality of Variances**, which tests whether or not the variance between each sample is equal. This determination is crucial because it dictates which calculation method should be used for the t-test and the corresponding confidence interval: the **Pooled** method (assuming equal variances) or the **Satterthwaite** method (assuming unequal

variances).

We assess the results by looking at the associated p-value. Since the p-value is not less than the conventional significance level of .05 in this table, we fail to reject the null hypothesis of equal variances. Therefore, we can assume that the variances between the two groups are equal.

Thus, we must look at the row that uses **Pooled** variance to find the 95% confidence interval for the difference in population means. Relying on the correct variance assumption ensures the validity of our final interval estimate.

Interpreting the Confidence Interval for the Difference in Means

Following the determination that the pooled method is appropriate, we extract the confidence interval values from the **Pooled** row of the t-test results table. This interval specifically estimates the true difference between the population mean height of Species A and Species B (Mean A - Mean B).

From the output we can see that the 95% confidence interval for the difference in population means is . This tells us we can be 95% confident that the true difference between the mean height of plants in species A compared to species B is between -4.6895 inches and -1.1305 inches.

Since the entire interval is negative--meaning both the lower and upper bounds are less than zero--the interval **does not contain zero**. This indicates that there is a statistically significant difference between the two population means at the 95% confidence level, with Species B having a significantly greater mean height than Species A.

Summary and Further Applications in SAS

The PROC TTEST procedure in SAS serves as an efficient and reliable tool for calculating confidence intervals for population means, whether for a single sample or for the difference between two independent samples. The methodology involves careful data setup, specifying the alpha level, and correctly using `VAR` and `CLASS` statements. For two-sample comparisons, the critical step of checking the equality of variances is essential for selecting the statistically correct interval estimate (Pooled vs. Satterthwaite).

Mastery of these procedures enables analysts to move beyond simple point estimates and provide a robust measure of uncertainty around their statistical conclusions. The interpretation of whether the resulting interval contains zero is the key to determining the practical significance of mean differences. This structured approach ensures that statistical reporting is both accurate and comprehensive.

For users looking to expand their capabilities in SAS, there are numerous other procedures for

calculating different types of confidence intervals, such as those for proportions (using **PROC FREQ**) or regression coefficients (using **PROC REG**). Continual exploration of the SAS documentation, particularly around inferential procedures, will deepen analytical skills and enhance the quality of statistical reporting.

The following resources explain how to perform other common tasks in SAS:

ARABPSYCHOLOGY.COM