How can confidence intervals be calculated in SAS?

Authored by

stats writer

June 25, 2024

Confidence intervals can be calculated in SAS using the PROC MEANS or PROC UNIVARIATE procedures. These procedures allow for the calculation of confidence intervals for a given variable, based on a specified confidence level. The confidence interval is a range of values within which the true population parameter is likely to fall with a certain level of confidence. SAS calculates the confidence interval by using the sample mean, standard deviation, and sample size. The resulting confidence interval can then be used to make inferences about the population parameter. Additionally, SAS offers the option to calculate confidence intervals for different types of data, such as continuous or categorical, making it a versatile tool for statistical analysis. Overall, SAS provides a reliable and efficient method for calculating confidence intervals and is commonly used in research and data analysis.

Calculate Confidence Intervals in SAS

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 confidence intervals in R:

1. Confidence Interval for a Population Mean

2. Confidence Interval for a Difference in Population Means

Let's jump in!

Example 1: Confidence Interval for Population Mean in SAS

Suppose we have the following dataset that contains

the height (in inches) of a random sample of 12 plants that all belong to the same species:

```
/*create dataset*/
data my_data;
input Height;
datalines;
14
14
16
13
12
17
15
14
15
13
15
14
;
run;

/*view dataset*/
proc printdata=my_data;
```

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

Suppose we would like to calculate a 95% confidence for the true population mean height of this species.

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

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

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

The value for Mean shows the sample mean and the values under 95% CL Mean show the 95% confidence interval for the population mean.

From the output we can see that the 95% confidence interval for the mean weight of plants in this population is .

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

Suppose we have the following dataset that contains the height (in inches) of a random sample of plants that belong to two different species:

```
/*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 printdata=my_data2;
```

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

Suppose we would like to calculate a 95% confidence for difference in population mean height between species A and species B.

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

```
/*sort data by Species to ensure confidence interval is calculated correctly*/
proc sortdata=my_data2;
by Species;
run;

/*generate 95% confidence interval for difference in population means*/
proc ttestdata=my_data2 alpha=0.05;
class Species;
var Height;
run;
```

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

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.

Since the p-value is not less than .05 in this table, we can assume that the variances between the two groups is equal.

Thus, we can look at the row that uses Pooled variance

to find the 95% confidence interval for difference in population means.

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 , this indicates that there is a statistically significant difference between the two population means.

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