How to use Proc Summary in SAS with examples?

Authored by

stats writer

July 1, 2024

Proc Summary is a powerful tool in SAS that allows users to efficiently summarize and analyze data. It is used to calculate descriptive statistics such as mean, median, minimum, maximum, and standard deviation for numeric variables. Additionally, it can also be used to create frequency tables, as well as perform grouping and sorting of data. To use Proc Summary, users need to specify the data set, the variables to be summarized, and the desired statistics to be calculated. For example, to calculate the mean and standard deviation of a variable called "age" in a data set called "patients", the syntax would be "proc summary data = patients mean std; var age; run;". This would generate a table with the mean and standard deviation values for the "age" variable. Overall, Proc Summary is a useful tool for quickly summarizing and analyzing data in SAS.

Use Proc Summary in SAS (With Examples)

You can use proc summary in SAS to quickly calculate the following for one or more variables in a dataset:

N: The total number of observationsMIN: The minimum valueMAX: The maximum valueMEAN: The meanSTD: The standard deviation

The following examples show how to use this procedure with the SAS built-in dataset called , which contains various measurements for 159 different fish caught in a lake in Finland.

We can use proc print to view the first 10 from this dataset:

/*view first 10 observations from *Fish* dataset*/
proc printdata=sashelp.Fish (obs=10);

run;

Obs	Species	Weight	Length1	Length2	Length3	Height	Width
1	Bream	242	23.2	25.4	30.0	11.5200	4.0200
2	Bream	290	24.0	26.3	31.2	12.4800	4.3056
3	Bream	340	23.9	26.5	31.1	12.3778	4.6961
4	Bream	363	26.3	29.0	33.5	12.7300	4.4555
5	Bream	430	26.5	29.0	34.0	12.4440	5.1340
6	Bream	450	26.8	29.7	34.7	13.6024	4.9274
7	Bream	500	26.8	29.7	34.5	14.1795	5.2785
8	Bream	390	27.6	30.0	35.0	12.6700	4.6900
9	Bream	450	27.6	30.0	35.1	14.0049	4.8438
10	Bream	500	28.5	30.7	36.2	14.2266	4.9594

Related:

Example 1: Proc Summary with One Variable

We can use the following code to calculate descriptive statistics for the Weight variable:

```
/*calculate descriptive statistics for Weight variable*/
proc summarydata=sashelp.Fish;
var Weight;
output out=summaryWeight;
run;

/*print output dataset*/
```

```
proc printdata=summaryWeight;
```

Obs	_TYPE_	_FREQ_	_STAT_	Weight
1	0	159	N	158.00
2	0	159	MIN	0.00
3	0	159	MAX	1650.00
4	0	159	MEAN	398.70
5	0	159	STD	359.09

Here's how to interpret the output table:

TYPE: This column shows whether or not every row in the dataset was used to calculate the descriptive statistics. 0 = Every row was used._FREQ_: The number of rows used to calculate each descriptive statistic._STAT_: The name of the descriptive statistic.Weight: The numerical value for the corresponding descriptive statistic.

From the output we can see:

The total number of observations was 158.The minimum weight value was 0.The maximum weight value was 1,650.The mean weight value was 398.70.The standard deviation of weight values was 359.09.

From these five values we can gain a pretty good understanding of the distribution of values for the Weight variable.

Example 2: Proc Summary with Multiple Variables

For example, we can use the following code to calculate descriptive statistics for the Weight and Height variables:

```
/*calculate descriptive statistics for Weight and Height variables*/
proc summarydata=sashelp.Fish;
var WeightHeight;
output out=summaryWeightHeight;
run;

/*print output dataset*/
proc printdata=summaryWeightHeight;
```


Obs	_TYPE_	_FREQ_	_STAT_	Weight	Height
1	0	159	N	158.00	159.000
2	0	159	MIN	0.00	1.728
3	0	159	MAX	1650.00	18.957
4	0	159	MEAN	398.70	8.971
5	0	159	STD	359.09	4.286

From the output we can see the five descriptive statistics for both Weight and Height.

Example 3: Proc Summary with One Variable Grouped by Another Variable

To calculate descriptive statistics for one variable grouped by another variable, we can use the class statement.

For example, we can use the following code to calculate descriptive statistics for Weight grouped by Species:

```
/*calculate descriptive statistics for Weight grouped by Species*/
proc summarydata=sashelp.Fish;
var Weight;
class Species;
output out=summaryWeightSpecies;
run;

/*print output dataset*/
proc printdata=summaryWeightSpecies;
```

Obs	Species	_TYPE_	_FREQ_	_STAT_	Weight
1		0	159	N	158.00
2		0	159	MIN	0.00
3		0	159	MAX	1650.00
4		0	159	MEAN	398.70
5		0	159	STD	359.09
6	Bream	1	35	N	34.00
7	Bream	1	35	MIN	242.00
8	Bream	1	35	MAX	1000.00
9	Bream	1	35	MEAN	626.00
10	Bream	1	35	STD	206.60
11	Parkki	1	11	N	11.00
12	Parkki	1	11	MIN	55.00
13	Parkki	1	11	MAX	300.00
14	Parkki	1	11	MEAN	154.82
15	Parkki	1	11	STD	78.76
16	Perch	1	56	N	56.00
17	Perch	1	56	MIN	5.90
18	Perch	1	56	MAX	1100.00
19	Perch	1	56	MEAN	382.24
20	Perch	1	56	STD	347.62
21	Pike	1	17	N	17.00
22	Pike	1	17	MIN	200.00
23	Pike	1	17	MAX	1650.00
24	Pike	1	17	MEAN	718.71
25	Pike	1	17	STD	494.14
26	Roach	1	20	N	20.00
27	Roach	1	20	MIN	0.00
28	Roach	1	20	MAX	390.00
29	Roach	1	20	MEAN	152.05
30	Roach	1	20	STD	88.83

The output table displays the descriptive statistics for each Species of fish.

For example, we can observe the following descriptive statistics for just the Bream fish:

The total number of observations was 34.The minimum weight value was 242.The maximum weight value was 1,000.The mean weight value was 626.The standard deviation of weight values was 206.60.

We can observe these descriptive statistics for every other species as well.

Additional Resources

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