How can the SAS Proc Univariate procedure be used to analyze data by group?

Authored by

stats writer

June 26, 2024

The SAS Proc Univariate procedure is a statistical tool that allows for the analysis of data by group. This procedure allows for the comparison of multiple groups within a dataset, making it a valuable tool for understanding the differences and similarities between various subsets of data. By specifying a group variable, the procedure can provide descriptive statistics, histograms, and other graphical representations for each group, allowing for a comprehensive and efficient analysis. This procedure is particularly useful for identifying patterns and trends within data, as well as for detecting any potential outliers or anomalies within a particular group. Overall, the SAS Proc Univariate procedure is an essential tool for any researcher or analyst looking to gain insights into their data through group analysis.

SAS: Use Proc Univariate by Group

You can use proc univariate in SAS with the by statement to calculate descriptive statistics for each numeric variable in a dataset, grouped by a particular variable.

This procedure uses the following basic syntax:

```
proc univariatedata=my_data normal;
by group_variable;
run;
```

The following example shows how to use this procedure in practice.

Example: Proc Univariate by Group in SAS

Suppose we have the following dataset in SAS that

contains information about various basketball players:

```
/*create dataset*/
data my_data;
input team $ points rebounds;
datalines;
A 12 8
A 12 8
A 12 8
A 23 9
A 20 12
A 14 7
A 14 7
B 20 2
B 20 5
B 29 4
B 14 7
B 20 2
B 20 2
B 20 5
;
run;

/*view dataset*/
```

```
proc printdata=my_data;
```

Obs	team	points	rebounds
1	A	12	8
2	A	12	8
3	A	12	8
4	A	23	9
5	A	20	12
6	A	14	7
7	A	14	7
8	B	20	2
9	B	20	5
10	B	29	4
11	B	14	7
12	B	20	2
13	B	20	2
14	B	20	5

We can use proc univariate with the by statement to calculate descriptive statistics for the points and rebounds variables, grouped by the team variable:

```
proc univariatedata=my_data;
by team;
run;
```

This procedure will produce the following results:

Descriptive statistics for points for team ADescriptive

statistics for rebounds for team BDescriptive statistics for points for team ADescriptive statistics for rebounds for team B

Here is what the descriptive statistics looks like for the points variable for team A:

The UNIVARIATE Procedure
Variable: points

team=A

Moments			
N	7	Sum Weights	7
Mean	15.2857143	Sum Observations	107
Std Deviation	4.42396073	Variance	19.5714286
Skewness	1.22128564	Kurtosis	-0.0509457
Uncorrected SS	1753	Corrected SS	117.428571
Coeff Variation	28.9417992	Std Error Mean	1.67209999

Basic Statistical Measures			
Location		Variability	
Mean	15.28571	Std Deviation	4.42396
Median	14.00000	Variance	19.57143
Mode	12.00000	Range	11.00000
		Interquartile Range	8.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	9.141627	Pr > |t|	<.0001
Sign	M	3.5	Pr >= |M|	0.0156
Signed Rank	S	14	Pr >= |S|	0.0156

Quantiles (Definition 5)	
Level	Quantile
100% Max	23
99%	23
95%	23
90%	23
75% Q3	20
50% Median	14
25% Q1	12
10%	12
5%	12
1%	12
0% Min	12

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
12	3	12	3
12	2	14	6
12	1	14	7
14	7	20	5
14	6	23	4

If you only want to calculate descriptive statistics for one specific variable grouped by another variable, then you can use the var statement.

For example, you can use the following syntax to calculate descriptive statistics only for the points variable, grouped by the team variable:

```
proc univariatedata=my_data;
var points;
```

by team;
run;

Feel free to specify as many variables as you'd like in both the var and by statements to calculate descriptive statistics for whichever variables you'd like.

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