How do I calculate descriptive statistics using SAS?

Authored by

stats writer

June 25, 2024

Descriptive statistics refer to a set of quantitative measures that summarize and describe the characteristics of a dataset. These statistics provide valuable information about the central tendency, variability, and distribution of the data. SAS (Statistical Analysis System) is a powerful statistical software that offers various tools and procedures to calculate descriptive statistics.

To calculate descriptive statistics using SAS, one can use the PROC MEANS procedure. This procedure allows users to obtain measures such as mean, median, standard deviation, and quartiles for numerical variables. The PROC UNIVARIATE procedure can also be used to generate a comprehensive report with descriptive statistics for multiple variables in a dataset.

To begin, the user must first import the dataset into SAS and ensure that the variables are correctly coded as numerical or categorical. Then, the desired procedure can be called and the necessary variables can be specified. The output will provide a detailed summary of the data, including measures of central tendency, variability, and distribution.

In conclusion, SAS offers a reliable and efficient way to calculate descriptive statistics for a dataset. Its procedures provide a comprehensive overview of the data, allowing users to gain valuable insights and make informed decisions. With its user-friendly interface and powerful analytical capabilities, SAS is a valuable tool for data analysis and research.

Calculate Descriptive Statistics in SAS

Descriptive statistics are values that describe a dataset.

They help us gain an understanding of where of the dataset is located along with how the values are in the dataset.

There are two common ways to calculate descriptive statistics for variables in SAS:

1. Use PROC MEANS to Calculate Summary Statistics

2. Use PROC UNIVARIATE to Calculate Detailed Descriptive Statistics

The following examples show how to use each method in practice with the following dataset in SAS:

```
/*create dataset*/
data my_data;
input team $ points assists;
datalines;
A 10 2
A 17 5
A 17 6
A 18 3
A 15 0
B 10 2
B 14 5
B 13 4
B 29 0
B 25 2
C 12 1
C 30 1
C 34 3
C 12 4
```

C 11 7

;

run;

/*view dataset*/

proc printdata=my_data;

Obs	team	points	assists
1	A	10	2
2	A	17	5
3	A	17	6
4	A	18	3
5	A	15	0
6	B	10	2
7	B	14	5
8	B	13	4
9	B	29	0
10	B	25	2
11	C	12	1
12	C	30	1
13	C	34	3
14	C	12	4
15	C	11	7

Example 1. Use PROC MEANS to Calculate Summary Statistics

The following code shows how to use PROC MEANS to calculate summary statistics for the points variable in the dataset:

```
/*calculate summary statistics for points variable*/
proc meansdata=my_data;
var points;
run;
```

The MEANS Procedure

Analysis Variable : points				
N	Mean	Std Dev	Minimum	Maximum
15	17.8000000	7.8848861	10.0000000	34.0000000

The PROC MEANS procedure calculates the following descriptive statistics:

N: The total number of observationsMean: The mean value of pointsStd Dev: The standard deviation of pointsMinimum: The minimum value of pointsMaximum: The maximum value of points

Also note that you can use the class statement to calculate descriptive statistics for one variable, grouped by another variable.

For example, we can use the following code to calculate summary statistics for the points variable, grouped by the team variable:

```
/*calculate summary statistics for points, grouped by team*/
proc meansdata=my_data;
class team;
var points;
run;
```

The MEANS Procedure

Analysis Variable : points						
team	N Obs	N	Mean	Std Dev	Minimum	Maximum
A	5	5	15.4000000	3.2093613	10.0000000	18.0000000
B	5	5	18.2000000	8.2885463	10.0000000	29.0000000
C	5	5	19.8000000	11.2338773	11.0000000	34.0000000

The output displays the summary statistics for the points variable, grouped by each of the unique team values.

Example 2. Use PROC UNIVARIATE to Calculate Detailed Descriptive Statistics

The following code shows how to use PROC UNIVARIATE to calculate detailed descriptive statistics for the points variable:

```
/*calculate detailed descriptive statistics for points
```

```
variable*/
proc univariate data=my_data;
var points;
run;
```

The UNIVARIATE Procedure
Variable: points

Moments			
N	15	Sum Weights	15
Mean	17.8	Sum Observations	267
Std Deviation	7.88488608	Variance	62.1714286
Skewness	1.00931793	Kurtosis	-0.2991564
Uncorrected SS	5623	Corrected SS	870.4
Coeff Variation	44.2971128	Std Error Mean	2.03586883

Basic Statistical Measures			
Location		Variability	
Mean	17.80000	Std Deviation	7.88489
Median	15.00000	Variance	62.17143
Mode	10.00000	Range	24.00000
		Interquartile Range	13.00000

Note: The mode displayed is the smallest of 3 modes with a count of 2.

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	8.743196	Pr > |t|	<.0001
Sign	M	7.5	Pr >= |M|	<.0001
Signed Rank	S	60	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	34
99%	34
95%	34
90%	30
75% Q3	25
50% Median	15
25% Q1	12
10%	10
5%	10
1%	10
0% Min	10

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
10	6	18	4
10	1	25	10
11	15	29	9
12	14	30	12
12	11	34	13

The PROC UNIVARIATE procedure calculates detailed descriptive statistics for the points variable including the mean, median, mode, standard deviation, variance, range, interquartile range, and more.

We can also use the class statement to calculate these detailed descriptive statistics for the points variable, grouped by the team variable:

```
/*calculate detailed descriptive statistics for points, grouped by team*/
proc univariate data=my_data;
class team;
var points;
run;
```

This will produce three groups of output tables that display detailed descriptive statistics for the points variable, grouped by each of the unique team values.

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