

# How can I calculate skewness and kurtosis in SAS?

Authored by  
**stats writer**

June 23, 2024

## RECOMMENDED CITATION

stats writer (2024). *How can I calculate skewness and kurtosis in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=148129>

Skewness and kurtosis are two important statistical measures that help to determine the shape and distribution of a dataset. In SAS, there are built-in procedures and functions that can be used to calculate skewness and kurtosis. To calculate skewness, the PROC MEANS or PROC UNIVARIATE procedure can be used, where the SKEWNESS option can be specified. Similarly, to calculate kurtosis, the KURTOSIS option can be specified in these procedures. Additionally, the SKEWNESS and KURTOSIS functions can also be used in the DATA step to calculate these measures for individual variables. By understanding and utilizing these methods in SAS, one can accurately assess the degree of asymmetry and peakedness in their data, aiding in the analysis and interpretation of statistical results.

## Calculate Skewness & Kurtosis in SAS

**In statistics, skewness and kurtosis are two ways to measure the shape of a distribution.**

**Skewness measures the asymmetry of a distribution.**

**A negative skew indicates that the tail is on the left side of the distribution. A positive skew indicates that the tail is on the right side of the distribution. A value of zero indicates that there is no skewness in the distribution at all, meaning the distribution is perfectly symmetrical.**

**Kurtosis measures whether or not a distribution is heavy-tailed or light-tailed relative to a normal distribution.**

**The kurtosis of a normal distribution is 0. If a given distribution has a kurtosis less than 0, it is said to be**

***platykurtic***, which means it tends to produce fewer and less extreme outliers than the normal distribution. If a given distribution has a kurtosis greater than 0, it is said to be ***leptokurtic***, which means it tends to produce more outliers than the normal distribution.

To calculate skewness and kurtosis for variables in SAS, you can use the **SKEWNESS** and **KURTOSIS** statements in **PROC MEANS**.

The following example shows how to use these statements in practice.

**Example: Calculating Skewness & Kurtosis 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 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

We can use **PROC MEANS** with the **SKEWNESS** and **KURTOSIS** statements to calculate the skewness and kurtosis for each numeric variable in the dataset:

```
/*calculate skewness and kurtosis for each numeric variable*/
```

```
proc means data=my_data SKEWNESSKURTOSIS;  
run;
```

**The MEANS Procedure**

Variable	Skewness	Kurtosis
points	1.0093179	-0.2991564
assists	0.3035599	-0.7824519

The table in the output displays the skewness and kurtosis values for each numeric variable in the dataset:

**(1) points**

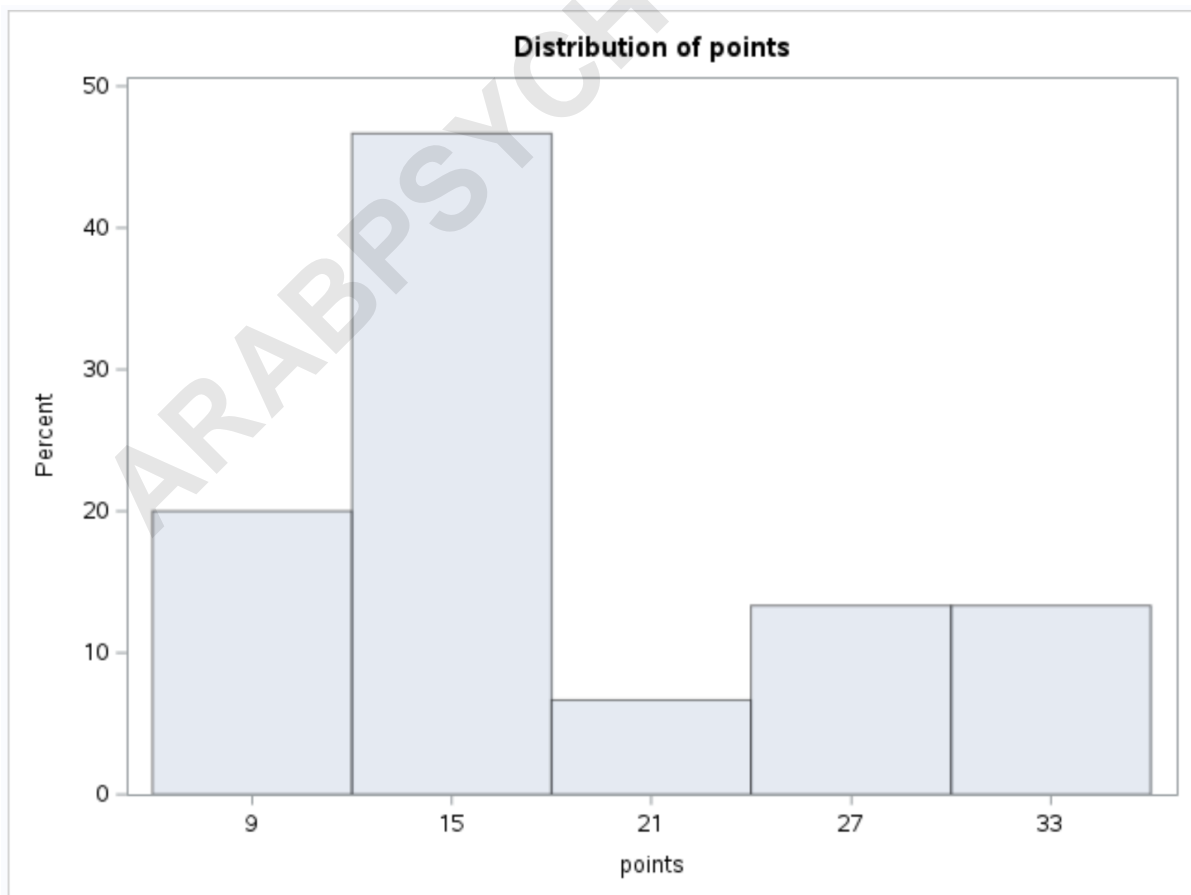
The points variable has a skewness of 1.009. Since this value is greater than 0, it means the tail is on the right side of the distribution. The points variable has a kurtosis of -0.299. Since this value is less than 0, it means the distribution has slightly fewer and less extreme outliers than the normal distribution. The assists variable has a skewness of 0.304. Since this value is greater than 0, it means the tail is on the right side of the distribution. The assists variable has a kurtosis of -0.782. Since this value is less than 0, it means the distribution has fewer and less extreme outliers than the normal distribution.

To visualize the distribution of values for each numeric variable in the dataset, you can use PROC UNIVARIATE

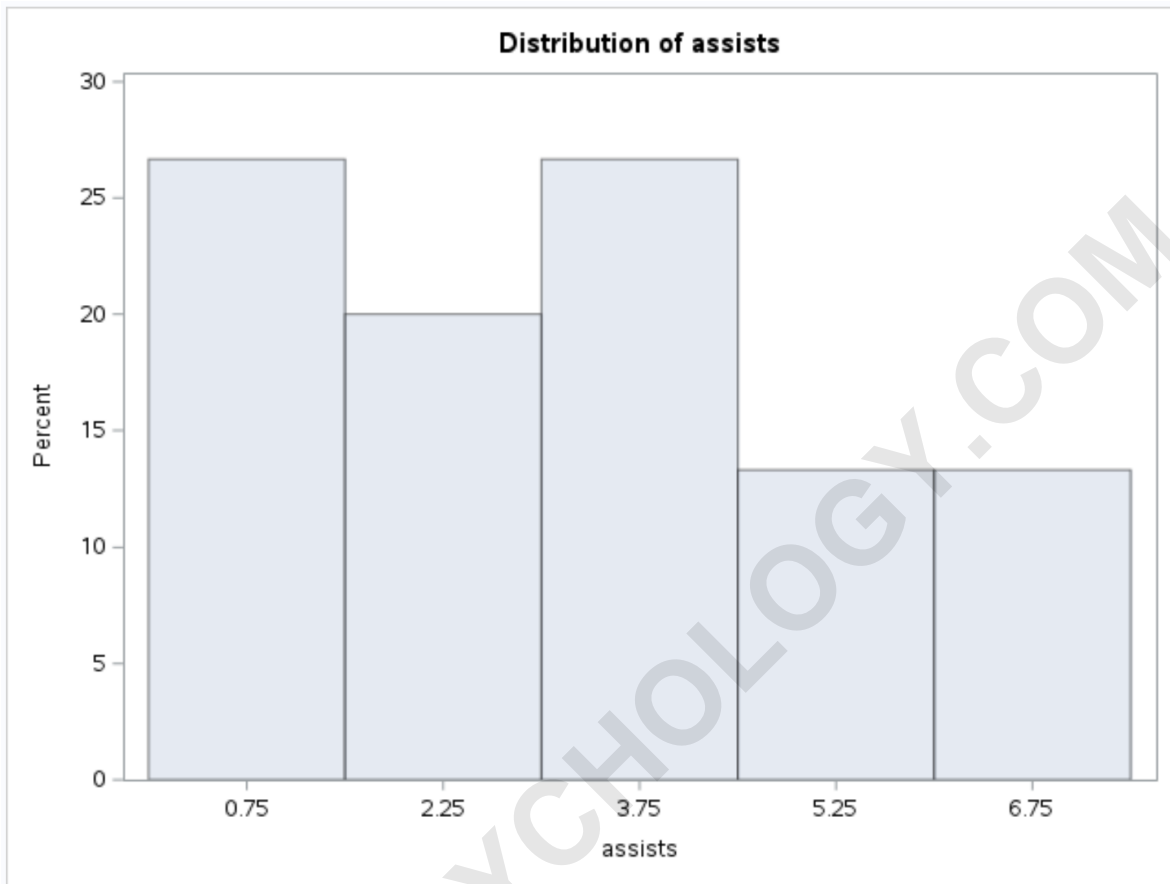
to create histograms for both the points and assists variables:

```
/*create histograms for points and assists variables*/  
proc univariate data=my_data;  
var points assists;  
histogram points assists;  
run;
```

This produces the following histogram for the points variable:



**And the following histogram for the assists variable:**



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