

# How to calculate Skewness & Kurtosis in SAS?

Authored by  
**stats writer**

November 19, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to calculate Skewness & Kurtosis in SAS?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=96807>

Skewness and kurtosis are fundamental statistical metrics used to quantify the shape of a data distribution, moving beyond simple measures of central tendency and dispersion. In the realm of advanced data analysis, particularly when utilizing statistical software like SAS, understanding how to calculate and interpret these values is crucial for validating modeling assumptions and gaining deeper insights into data characteristics. These metrics provide essential context regarding the symmetry and the peakedness, or tail heaviness, of the distribution relative to a standard benchmark, such as the Normal Distribution.

Proper assessment of distributional shape is vital in many statistical applications, as many inferential tests, such as t-tests and ANOVA, rely on the assumption of normality. Deviations from normality, particularly pronounced skew or high kurtosis, can necessitate data transformations or the use of non-parametric methods. Fortunately, SAS provides straightforward procedures, notably PROC UNIVARIATE and PROC MEANS, that efficiently calculate these shape parameters, allowing analysts to rapidly diagnose the nature of their data set and proceed with appropriate modeling strategies.

## Understanding Measures of Distribution Shape

In the field of statistics, analyzing the distribution of a variable is often the first and most critical step in any rigorous data analysis workflow. While descriptive statistics like the mean, median, and standard deviation tell us about the center and spread of the data, they do not fully describe the distribution's unique form. This is where higher-order moments, specifically skewness and kurtosis, become indispensable tools. They function as comprehensive measures of distribution shape, offering quantitative insight into how data points are clustered around the central tendency and how frequently extreme values occur in the tails.

The shape of a distribution dictates the validity and reliability of subsequent statistical inferences. For instance, a distribution that is severely asymmetric might suggest that the mean is not the best measure of central tendency, as it is heavily influenced by outliers in the long tail. Similarly, a distribution with very heavy tails (high kurtosis) indicates an elevated risk of observing extreme values, which is a crucial consideration in fields like risk management or quality control. Understanding these shape parameters is essential for building robust statistical models that accurately reflect the underlying data generating process.

To efficiently obtain these summary statistics in the SAS environment, analysts typically rely on established procedures. The **UNIVARIATE** procedure is designed specifically for detailed single-variable analysis, providing comprehensive output including tests for normality, graphical representations, and precise calculations of moments. Alternatively, the **MEANS** procedure offers a streamlined way to calculate many common statistics, including skewness and kurtosis, often utilized when summary statistics for multiple variables or groups are required rapidly.

## Defining and Interpreting Skewness

Skewness is a statistical measure that quantifies the degree of asymmetry present in a probability distribution. A perfectly symmetrical distribution, such as the Normal Distribution, has a skewness value of zero, meaning that data points are distributed equally on both sides of the mean. Deviations from zero indicate asymmetry and suggest that the distribution is stretched or "skewed" in one direction due to the presence of outliers pulling the tail.

The interpretation of the skewness coefficient is straightforward and highly informative about the distribution's shape:

A **negative skew** indicates that the tail is on the left side of the distribution. This implies that the majority of data points have high values, while a few extreme low values pull the mean toward the left.

A **positive skew** indicates that the tail is on the right side of the distribution. The majority of observations are clustered on the left (low values), and the mean is pulled towards the right by extreme high values.

A value of **zero** indicates that there is no skewness in the distribution at all, meaning the distribution is perfectly symmetrical, where the mean, median, and mode are ideally equal.

Understanding the direction and magnitude of the skew is critical for statistical modeling. For instance, high positive skewness in residuals may violate the assumptions of linear regression, requiring transformation techniques like the log transform to normalize the variable before modeling can proceed effectively. Skewness analysis provides a foundational understanding of data bias and concentration.

## Defining and Interpreting Kurtosis

Kurtosis measures whether a distribution is heavy-tailed or light-tailed relative to a benchmark distribution, typically the Normal Distribution. This statistic quantifies the "tailedness" and peakedness, giving insight into the probability of extreme observations. When calculating excess kurtosis (the standard output in SAS), the normal distribution serves as the zero reference point.

The interpretation of the excess kurtosis coefficient relies on comparison to the mesokurtic standard:

The kurtosis of a normal distribution (excess kurtosis) is **0** (mesokurtic).

If a given distribution has a kurtosis **less than 0** (negative), it is said to be **platykurtic**, which means it tends to produce fewer and less extreme outliers than the normal distribution. It has a flatter peak and lighter tails.

If a given distribution has a kurtosis **greater than 0** (positive), it is said to be **leptokurtic**. This

signifies that the distribution has heavier tails and a sharper peak than the normal distribution, implying that it tends to produce more outliers or extreme values.

High positive kurtosis is a critical finding in risk assessment, as it implies that the occurrence of rare, high-impact events is more likely than a normal model would predict. SAS facilitates the calculation of this fourth moment, ensuring analysts can properly account for the presence of heavy tails in their modeling process.

## Setting Up the Analysis: Data Preparation

To calculate these shape parameters for variables in SAS, we will utilize a practical example involving a dataset of basketball player statistics. This dataset contains numerical variables, 'points' and 'assists', for which we want to determine the degree of skewness and kurtosis. The first step involves creating and verifying the dataset within the SAS environment using the standard DATA step syntax.

The following SAS code snippet creates the dataset `my_data` and populates it using the DATALINES statement. We define the team as a character variable and points and assists as numeric variables that will be the focus of our distributional analysis.

```
/*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 print data=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

The `PROC PRINT` output confirms that the data has been loaded correctly. With the dataset prepared, we can now proceed to the computational step, utilizing the power of SAS procedures to derive the skewness and kurtosis coefficients efficiently for both 'points' and 'assists'.

## Executing the Calculation using PROC MEANS

To calculate skewness and kurtosis for the variables in our SAS dataset, we can use the **SKEWNESS** and **KURTOSIS** statements within PROC MEANS. This procedure is optimized for generating quick summary statistics. By omitting a `VAR` statement, `PROC MEANS` automatically calculates the requested metrics for all numerical variables in the input dataset, which is ideal for a broad initial assessment.

The following example shows how to use these statements in practice, focusing solely on obtaining the shape parameters from the `my_data` table:

```
/*calculate skewness and kurtosis for each numeric variable*/  
proc means data=my_data SKEWNESS KURTOSIS;  
run;
```

Executing this code produces a concise output table displaying the computed values. This streamlined approach minimizes unnecessary statistical output and provides a direct, quantitative measure of asymmetry and tailedness for each variable, crucial for the next step of interpretation.

Variable	Skewness	Kurtosis
points	1.0093179	-0.2991564
assists	0.3035599	-0.7824519

## Interpreting the Calculated Results

The table in the output displays the numerical skewness and kurtosis values for each numerical variable in the dataset, providing the raw data necessary for drawing conclusions about distribution shape relative to the Normal Distribution.

### Interpretation for 'points' variable:

The points variable has a **skewness** of **1.009**. Since this value is greater than 0, it indicates a strong **positive skew**, meaning the distribution has a long tail on the right side.

The points variable has an excess **kurtosis** of **-0.299**. Since this value is less than 0, the distribution is **platykurtic**, indicating it has slightly fewer and less extreme outliers than the normal distribution.

### Interpretation for 'assists' variable:

The assists variable has a **skewness** of **0.304**. Since this value is greater than 0, it indicates a mild **positive skew**, with the tail extending to the right.

The assists variable has an excess **kurtosis** of **-0.782**. Since this value is significantly less than 0, the distribution is highly **platykurtic**, suggesting the distribution is quite flat with very light tails compared to the normal distribution.

These numerical findings confirm that both variables exhibit asymmetry and possess a flatter, lighter-tailed nature, strongly suggesting that statistical models relying on the assumption of strict

normality should be applied with caution or preceded by data transformation.

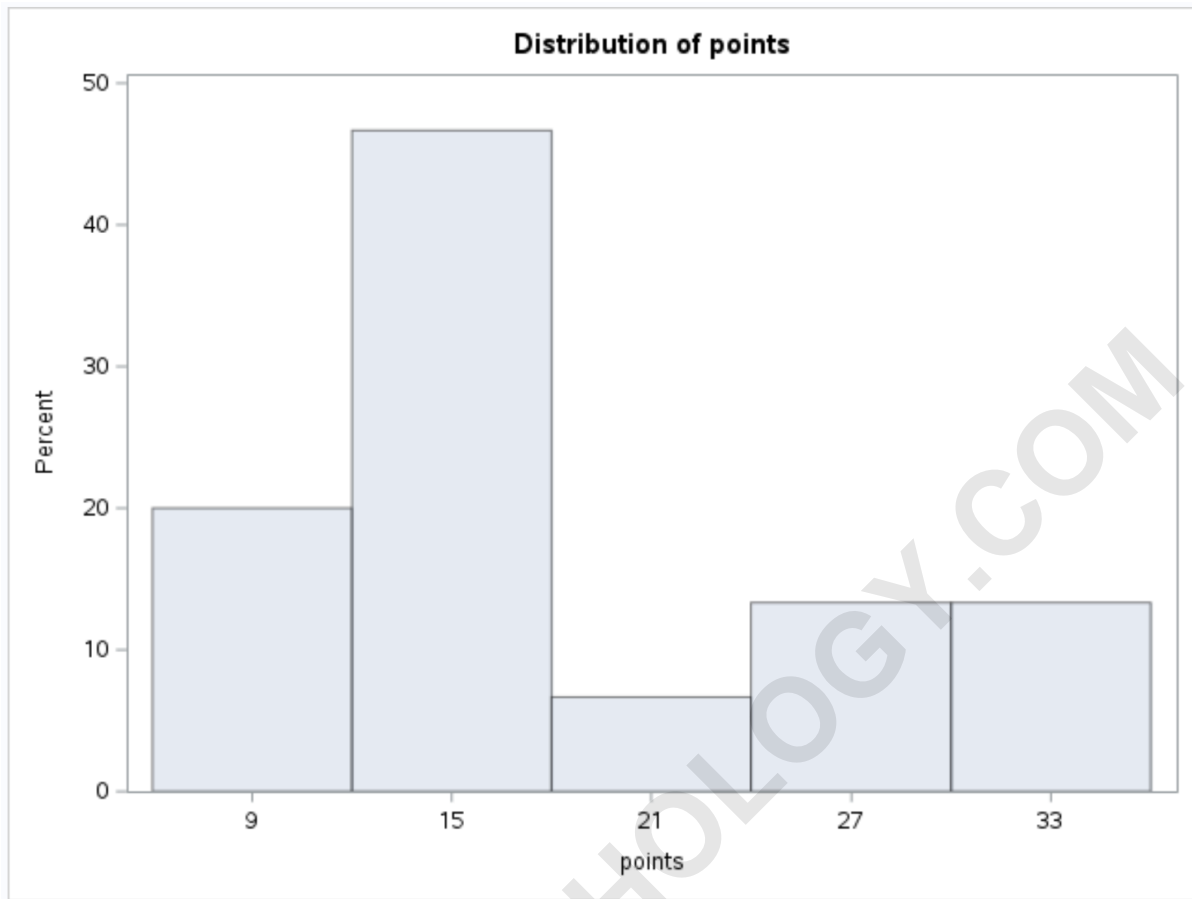
## Visualizing and Confirming Results with PROC UNIVARIATE

To visually validate the numerical results obtained from PROC MEANS, we use PROC UNIVARIATE to generate histograms. Histograms provide an immediate, intuitive representation of the distribution shape, allowing us to confirm the presence of skewness and the characteristics of the tails.

The following code uses PROC UNIVARIATE and the HISTOGRAM statement to create graphical outputs for both the points and assists variables:

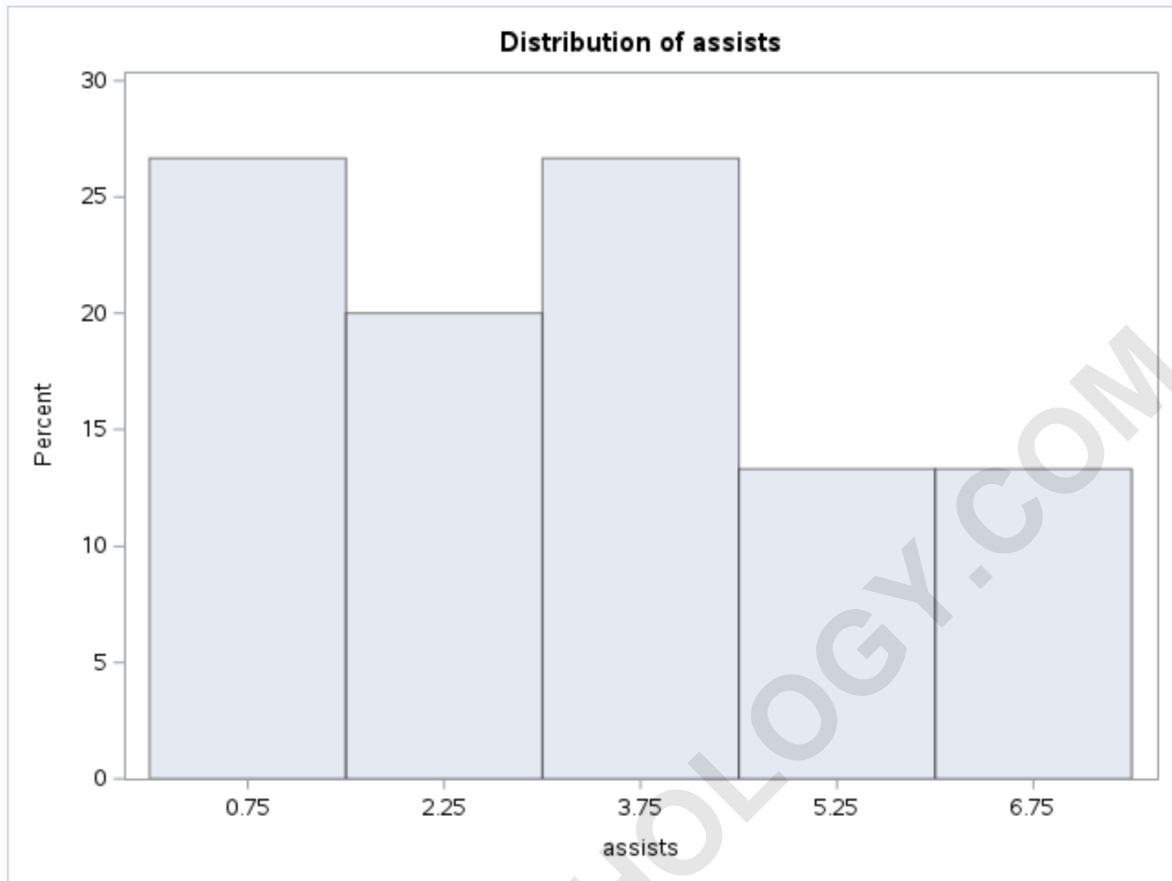
```
/*create histograms for points and assists variables*/  
proc univariate data=my_data;  
var points assists;  
histogram points assists;  
run;
```

This produces the following histogram for the **points** variable:



The histogram for points clearly shows the bulk of the data clustered on the left, with a distinct tail extending right, confirming the positive skewness value of 1.009. The distribution's overall appearance is also consistent with the platykurtic finding.

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



The assists histogram also displays a slight rightward pull and a flat top, graphically supporting the mild positive skew (0.304) and the strongly platykurtic nature (-0.782). The combined use of numerical coefficients and visual inspection ensures a thorough and validated assessment of the data's distributional properties.

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