

# SAS: Display Median in PROC MEANS

Authored by  
**Mohammed looti**

November 15, 2025

## RECOMMENDED CITATION

Mohammed looti (2025). *SAS: Display Median in PROC MEANS*. PSYCHOLOGICAL STATISTICS. Retrieved from <https://statistics.arabpsychology.com/?p=1951>

## Introduction to Descriptive Statistics with SAS

In the advanced world of statistical analysis, [SAS](#) remains a foundational and powerful software suite, highly valued for its robust capabilities in data management, advanced modeling, and comprehensive reporting. The initial phase of any thorough data investigation must necessarily begin with [descriptive statistics](#), which serve to provide simple, yet crucial, summaries of the sample data and the observed characteristics. These calculated [summary statistics](#) are vital for understanding the fundamental features of the data structure. By simplifying large volumes of information into an intelligible format, descriptive statistics offer a clear visualization of the data's distribution, central tendency, and inherent variability.

When analyzing measures of central tendency--metrics designed to pinpoint the typical or central value of a given [variable](#)--the [median](#) carries unique importance. Unlike the [mean](#), which is highly susceptible to distortion by extreme values or [outliers](#), the [median](#) provides a far more stable and robust measure. It represents the exact middle value of a dataset once all observations are ordered sequentially. For distributions that are heavily skewed or contain significant extreme values, the median frequently delivers a more accurate and representative measure of the "average" experience within the data landscape.

This comprehensive guide focuses on maximizing the utility of one of SAS's most frequently deployed procedures, [PROC MEANS](#), specifically demonstrating how to calculate and display the median alongside other essential [summary statistics](#). While [PROC MEANS](#) is powerful, the [median](#) is notably absent from its default statistical output. Therefore, we will meticulously examine the precise [syntax](#) required to explicitly request the median, thereby ensuring a complete and nuanced understanding of your data's central characteristics.

## Understanding the `PROC MEANS` Procedure

The [PROC MEANS](#) procedure within the [SAS](#) environment is an essential utility designed for efficiently generating [summary statistics](#) for all numeric [variables](#). By default, it swiftly computes key measures such as the total count of valid [observations](#) (N), the [mean](#), the [standard deviation](#) (Std Dev), and the [minimum](#) and [maximum](#) values. Collectively, these statistics provide an immediate and valuable snapshot of the data's underlying distribution, enabling users to quickly identify potential anomalies or critical patterns requiring further investigation.

To activate [PROC MEANS](#) and specify the desired analytical outputs, statistical keywords must be appended directly to the procedure statement. For example, if the goal is to calculate the number of observations (N), [mean](#), median, standard deviation (Std), minimum (Min), and maximum (Max) for a variable named **points**, sourced from a [dataset](#) labeled **my\_data**, the correct [syntax](#) structure must be implemented precisely:

```
proc means data=my_data N Mean Median Std Min Max;  
var points;  
run;
```

Within this foundational [syntax](#), the `data=my_data` option explicitly designates the input [dataset](#) upon which the analysis will be performed. Crucially, the string of keywords `N Mean Median Std Min Max` directs [PROC MEANS](#) to calculate and display these specific [summary statistics](#). Finally, the `var points;` statement isolates **points** as the target [variable](#) for calculation. This highly structured approach grants the user precise control over the analytical output, enabling tailored statistical summaries that directly address specific research objectives.

## Constructing Your SAS Dataset for Analysis

To effectively demonstrate the practical necessity of instructing [PROC MEANS](#) to display the [median](#), we will first establish a working example. This involves creating a simplified [dataset](#) within [SAS](#), containing hypothetical performance metrics for several basketball players. This dataset, which we will name **my\_data**, will include essential details such as the player's assigned team, their individual points scored, and the number of assists they recorded during gameplay.

The foundational method for building and managing [SAS datasets](#) is through the use of a [DATA step](#). Within this crucial step, we define the data structure: the [INPUT statement](#) is used to name the [variables](#) (`team $, points, assists`) and specify their types, while the [DATALINES statement](#) allows us to embed the raw data records directly into the program for immediate execution. This streamlined approach ensures the data is correctly structured before any statistical procedures are applied.

Immediately following the data creation, we utilize [PROC PRINT](#) procedure to display the complete contents of the newly generated **my\_data**. [PROC PRINT](#) is indispensable for data validation, providing a clear, tabular view of the dataset. This verification step is absolutely essential to confirm the integrity and structure of our [observations](#) before we proceed to any formal [descriptive statistics](#) 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 resulting output confirms that our **my\_data dataset** is correctly populated. It comprises 15 complete [observations](#) and three distinct [variables](#): **team** (a character variable), **points** (a numeric variable), and **assists** (another numeric variable). With the data validated and structured, we are

prepared to move forward with the descriptive statistical analysis.

## Analyzing Default `PROC MEANS` Output

Having successfully prepared and validated our sample data, the next logical step is to calculate the preliminary [summary statistics](#) for the **points** [variable](#). We will initially execute [PROC MEANS](#) using its most basic form--without explicitly requesting the [median](#). This exercise clearly illustrates the procedure's default behavior and underscores the necessity of using specific keywords when a more comprehensive set of analytical metrics is required for the study.

```
/*calculate summary statistics for points variable without specifying keywords*/
proc means data=my_data;
var points;
run;
```

Executing the simple code above instructs [PROC MEANS](#) to generate a standard output table containing several fundamental [descriptive statistics](#). The resulting table, displayed below, provides immediate insights into the central tendency and dispersion of the **points** variable, though it is intentionally incomplete regarding robust measures.

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

By default, the [PROC MEANS](#) procedure typically calculates the following core [summary statistics](#):

**N:** Represents the total number of non-missing [observations](#) available for the analyzed [variable](#). In our basketball data, this is the count of valid player scores.

**Mean:** The arithmetic [mean](#), calculated by dividing the sum of all point values by N. This figure provides the calculated average points scored.

**Std Dev:** The [standard deviation](#), a metric that quantifies the amount of variation or data dispersion around the [mean](#). A higher standard deviation suggests greater inconsistency in scoring.

**Minimum:** The [minimum](#) value observed for the **points** variable, denoting the lowest score recorded in the dataset.

**Maximum:** The [maximum](#) value observed, which represents the single highest score achieved by any player in the sample.

As clearly visible in the output, while these measures offer a solid initial overview of the data's spread and center, the [median](#) value--an essential metric for assessing central tendency without sensitivity to extreme scores--is entirely absent. This highlights the crucial necessity of explicitly requesting the median through proper [syntax](#) when it is a required component of the statistical summary.

## Incorporating the Median into `PROC MEANS`

To achieve a truly complete and robust picture of our data's central tendency, it is mandatory to instruct [PROC MEANS](#) to calculate and include the [median](#) in its generated output. This vital inclusion is accomplished through a simple yet powerful modification: adding the specific keyword `Median` to the list of requested statistics within the [PROC MEANS](#) statement. This straightforward adjustment allows the analyst to harness the full potential of the procedure, ensuring that the statistical summary is comprehensive and reliable.

The enhanced [syntax](#), which now ensures the calculation of the middle value, is structured as follows:

```
/*calculate summary statistics for points and include median value*/  
proc means data=my_data N Mean Median Std Min Max;  
var points;  
run;
```

By explicitly listing `Median` among the desired metrics, we override the procedure's default settings and compel [SAS](#) to compute and present this crucial measure. This is particularly important when analyzing numeric [variables](#) derived from real-world data, where distributions are often skewed or contain influential outliers. In such cases, the [median](#) offers a much more representative central value than the standard [mean](#), leading to better-informed analytical conclusions.

## Interpreting the Enhanced Statistical Output

After running the modified [PROC MEANS](#) code, the resulting output table is significantly improved, now featuring the calculated [median](#) value. This provides a substantially richer set of [summary statistics](#) for the `points` variable, which is essential for conducting a nuanced assessment of both central tendency and overall data distribution characteristics.

### The MEANS Procedure

Analysis Variable : points					
N	Mean	Median	Std Dev	Minimum	Maximum
15	17.8000000	15.0000000	7.8848861	10.0000000	34.0000000

As clearly indicated in the output table above, the [median](#) value for the **points variable** is now prominently displayed as **15**. This figure signifies that when the scores of all players in our sample [dataset](#) are arranged in ascending order, 15 is the value located precisely in the middle. Consequently, we can conclude that 50% of the players scored 15 points or less, and conversely, 50% of the players scored 15 points or more. This measure is highly valuable because it is less sensitive to the highest or lowest single scores recorded.

A crucial step in statistical interpretation is comparing the [mean](#) against the [median](#). If these two values are close, it generally indicates that the data distribution is relatively symmetrical. However, a noticeable divergence, such as the one observed here--where the mean (16.53) is higher than the median (15)--often suggests a degree of positive skewness. This positive skew implies the presence of a few high scores that are pulling the [mean](#) upward, away from the typical center of the data. This comprehensive comparison allows analysts to draw far more informed conclusions regarding player performance dynamics and data normality.

## Conclusion: Enhancing Descriptive Analysis in SAS

Achieving mastery of the [PROC MEANS](#) procedure in [SAS](#) is absolutely essential for anyone conducting serious [descriptive statistics](#). As demonstrated throughout this guide, while the procedure offers a standardized set of default [summary statistics](#), the strategic ability to explicitly request the [median](#) significantly deepens the analytical output. This is particularly critical when the goal is to obtain a measure of central tendency that is inherently robust and less influenced by extreme or anomalous values.

The straightforward act of including the `Median` keyword in your [PROC MEANS](#) statement immediately results in a clearer, more representative understanding of your data's typical value. This technique is universally valuable across diverse disciplines--including finance, sociology, engineering, and public health--where data distributions frequently deviate from the normal curve, making reliable, robust statistics paramount for accurate decision-making.

We strongly encourage practitioners to continue experimenting with various [variables](#) and statistical keywords available within [PROC MEANS](#) to solidify their expertise. The inherent flexibility and customizability of the [SAS](#) system empower users to extract precisely the insights

required from complex [datasets](#), moving beyond basic averages to a truly comprehensive statistical view.

## Additional Resources

To expand your proficiency in [SAS](#) and explore other common data analysis tasks, consider the following tutorials:

How to calculate percentiles in SAS.

Performing frequency analysis with SAS.

Running ANOVA in SAS.