

SAS: How to use NWAY in PROC SUMMARY

Authored by
stats writer

November 19, 2025

RECOMMENDED CITATION

stats writer (2025). *SAS: How to use NWAY in PROC SUMMARY*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=96869>

Introduction: The Importance of Focused Summary Statistics in SAS

The PROC SUMMARY procedure in SAS is a foundational tool for analysts seeking to condense large datasets into meaningful summary measures. While incredibly versatile, the default behavior of PROC SUMMARY often generates output that includes summary statistics for all defined classification levels, including the overall dataset totals. For sophisticated data analysis, analysts frequently require statistics calculated only at the lowest, most granular level of grouping, excluding these grand totals which can clutter the output dataset and complicate subsequent processing steps.

This is where the powerful but often misunderstood NWAY option becomes indispensable. The purpose of the NWAY option is precisely to filter the generated summary dataset, retaining only the records that correspond to the analysis performed at the deepest level of classification specified in the PROC SUMMARY statement's **CLASS** variables. By employing NWAY, we ensure that the output focuses solely on group-level calculations, leading to cleaner, more efficient, and more easily managed datasets for reporting or further statistical modeling.

Understanding the mechanism behind PROC SUMMARY's grouping--specifically the automatic generation of the internal variable _TYPE_--is critical to appreciating the value of **NWAY**. When classification variables are defined, PROC SUMMARY calculates statistics for every possible combination of grouping levels, including the overall total (which is represented by a specific value in the _TYPE_ variable). The **NWAY** option acts as a selector, instructing PROC SUMMARY to bypass the default inclusion of these higher-level summaries, thus optimizing performance and simplifying data preparation.

The Role and Syntax of the NWAY Option

When executing a summary procedure in SAS, the PROC SUMMARY statement determines the scope of the analysis. Without any special options, if you define variables in the **CLASS** statement, the procedure calculates descriptive statistics not only for the lowest level of classification (the combinations of all **CLASS** variables) but also for intermediate subtotals and the overall total. The **NWAY** option directly addresses this, ensuring that the statistics are calculated and outputted exclusively at the level defined by the complete interaction of all grouping variables.

The syntax for implementing **NWAY** is elegantly simple, requiring only its inclusion immediately following the procedure declaration. Unlike other options that might require complex parameter definitions, **NWAY** operates as a binary switch. For instance, if you classify by three variables (A, B, and C), the **NWAY** option ensures that only the statistics for the A*B*C combination are saved to the output dataset, suppressing the totals for A, B, C, A*B, A*C, B*C, and the overall grand total. This behavior is fundamental when the objective is to analyze specific group performance without

the noise of aggregated data.

The core benefit of using **NWAY** lies in controlling the output dataset size and structure. In applications involving automated processing or macro loops, having extraneous total rows can complicate filtering logic and increase processing time. By using **NWAY**, the analyst can guarantee a predictable output structure where every row represents a unique combination of the classification variables, maximizing the utility of the output dataset for downstream processes such as merging, reporting, or generating visualizations. This level of control is essential for maintaining data integrity and streamlining large-scale data analysis pipelines.

Setting Up the Example: The sashelp.Fish Dataset

To demonstrate the practical application of the **NWAY** option, we will utilize a well-known, built-in dataset available in all SAS installations: **sashelp.Fish**. This dataset is commonly used for examples involving classification and aggregation, containing 159 observations detailing various physical measurements for different fish species caught in a lake in Finland. Key variables include **Species**, which we will use as our primary classification variable, and **Weight**, the continuous variable for which we will calculate the descriptive statistics.

Before performing the summarization, it is always beneficial to familiarize ourselves with the data structure. The dataset includes categorical identifiers like **Species** and numerical measurements like **Weight**, **Length1**, **Length2**, etc. Our goal is to calculate summary measures--specifically the count, mean, minimum, maximum, and standard deviation--of the fish **Weight**, segmented by the various **Species** present in the data. Viewing the initial observations confirms the structure and content we will be summarizing.

We begin by using **PROC PRINT** to display the first 10 observations of the **sashelp.Fish** dataset. This step serves as a validation check and provides context for the subsequent aggregation, ensuring we understand the input structure before moving to the summary phase.

Displaying Initial Data Structure

The following SAS code snippet employs **PROC PRINT**, limiting the output using the **OBS** option to show just the first 10 rows of the data. This provides a quick snapshot of the raw information we will be processing with PROC SUMMARY.

```
/*view first 10 observations from Fish dataset*/  
proc print data=sashelp.Fish (obs=10);  
  
run;
```

The corresponding output, typically displayed in the SAS output window, confirms the structure of the dataset, highlighting variables such as **Species** and **Weight**. This visualization is crucial for verifying that the data is prepared correctly for the grouping operation we intend to perform.

Obs	Species	Weight	Length1	Length2	Length3	Height	Width
1	Bream	242	23.2	25.4	30.0	11.5200	4.0200
2	Bream	290	24.0	26.3	31.2	12.4800	4.3056
3	Bream	340	23.9	26.5	31.1	12.3778	4.6961
4	Bream	363	26.3	29.0	33.5	12.7300	4.4555
5	Bream	430	26.5	29.0	34.0	12.4440	5.1340
6	Bream	450	26.8	29.7	34.7	13.6024	4.9274
7	Bream	500	26.8	29.7	34.5	14.1795	5.2785
8	Bream	390	27.6	30.0	35.0	12.6700	4.6900
9	Bream	450	27.6	30.0	35.1	14.0049	4.8438
10	Bream	500	28.5	30.7	36.2	14.2266	4.9594

Executing Standard PROC SUMMARY (Without NWAY)

To establish a baseline understanding of how PROC SUMMARY operates by default, we first run the procedure without the **NWAY** option. We specify **Weight** in the **VAR** statement and **Species** in the **CLASS** statement. This tells SAS to calculate descriptive statistics for **Weight**, grouping the results by **Species**. The output is directed to a new dataset called **summaryWeight** using the **OUTPUT OUT=** statement.

The core feature of this default run is the generation of hierarchical summary statistics. Since only one classification variable (**Species**) is used, PROC SUMMARY will naturally calculate two types of summary sets: statistics for the overall dataset (ignoring the **Species** grouping) and statistics for each individual **Species** group. Understanding how these two sets are differentiated is central to mastering the use of the **NWAY** option.

The following code executes the standard summary procedure and prints the resulting output dataset. Note the inclusion of the **OUTPUT OUT=** statement, which is essential for saving the summary results into a format that can be further analyzed or viewed with **PROC PRINT**.

```
/*calculate descriptive statistics for Weight, grouped by Species*/
proc summary data=sashelp.Fish;
var Weight;
class Species;
```

```
output out=summaryWeight;
run;
```

```
/*print output dataset*/
proc print data=summaryWeight;
```

Interpreting the Default Output: Understanding _TYPE_

Upon reviewing the output from the standard PROC SUMMARY run, we observe several automatic variables generated by the procedure, two of which are crucial for understanding data aggregation: _TYPE_ and _FREQ_. The output generated is significantly larger than the number of unique species because it includes multiple levels of aggregation. In this specific case, the output contains 40 rows in total, reflecting the summaries for each species group plus the summaries for the entire dataset, for each requested statistic.

Obs	Species	_TYPE_	_FREQ_	_STAT_	Weight
1		0	159	N	158.00
2		0	159	MIN	0.00
3		0	159	MAX	1650.00
4		0	159	MEAN	398.70
5		0	159	STD	359.09
6	Bream	1	35	N	34.00
7	Bream	1	35	MIN	242.00
8	Bream	1	35	MAX	1000.00
9	Bream	1	35	MEAN	626.00
10	Bream	1	35	STD	206.60
11	Parkki	1	11	N	11.00
12	Parkki	1	11	MIN	55.00
13	Parkki	1	11	MAX	300.00
14	Parkki	1	11	MEAN	154.82
15	Parkki	1	11	STD	78.76
16	Perch	1	56	N	56.00
17	Perch	1	56	MIN	5.90
18	Perch	1	56	MAX	1100.00
19	Perch	1	56	MEAN	382.24
20	Perch	1	56	STD	347.62

The internal variable _TYPE_ is an indicator of the level of grouping. When only one variable

(Species) is in the **CLASS** statement, **_TYPE_** will have values of 0 or 1. A value of **0** indicates that the statistics were calculated across the entire dataset, ignoring the classification variable--these are the grand totals. A value of **1** indicates that the statistics were calculated at the deepest level of classification defined, which, in this example, is the individual **Species** level.

Let us examine the key columns in the default output dataset, paying close attention to the rows where **_TYPE_** equals 0, as these represent the overall summary that the **NWAY** option is designed to suppress:

TYPE: Indicates the level of aggregation. 0 means overall total; 1 means classification level total (by Species).

FREQ: The number of observations used in calculating the descriptive statistics for that specific row.

STAT: Identifies the specific descriptive statistic calculated (e.g., N, MIN, MAX, MEAN, STD).

Weight: The calculated numerical value for the corresponding statistic.

The initial rows of the output, corresponding to **_TYPE_** = 0, show descriptive statistics for all 158 observations in the dataset. For example, the mean weight is 398.70, and the standard deviation is 359.09. Following these overall summaries, the output transitions to rows where **_TYPE_** = 1, providing summaries for specific groups, such as the Bream species. While comprehensive, this inclusion of the overall totals is often redundant if the primary goal is only group-level comparison.

Applying the NWAY Statement for Focused Analysis

When the analytical requirement is strictly limited to the summaries generated by the complete interaction of the **CLASS** variables, the **NWAY** option provides the necessary control. By including **NWAY** in the **PROC SUMMARY** statement, we instruct **SAS** to discard all summary records where the **_TYPE_** value does not correspond to the highest classification level.

In the context of our single classification variable (**Species**), adding **NWAY** ensures that only rows where **_TYPE_** equals 1 are written to the output dataset. Essentially, we are filtering out the overall total rows (**_TYPE_** = 0). If we had two classification variables (e.g., Species and Location), the **NWAY** option would ensure that only the Species*Location interaction summaries (where **_TYPE_**=3, depending on binary encoding) are retained, excluding all grand totals and intermediate subtotals. This mechanism is key to producing clean, ready-to-use output datasets for specific group comparisons.

The code below integrates the **NWAY** option directly into the **PROC SUMMARY** statement. This small modification dramatically changes the resulting output dataset, making it significantly shorter and easier to handle, as it exclusively contains the group-level descriptive statistics for each unique species.

Executing PROC SUMMARY with NWAY

We re-run the aggregation procedure, this time incorporating the essential **NWAY** option:

```
/*calculate descriptive statistics for Weight, grouped by Species*/
```

```
proc summary data=sashelp.Fish nway;
```

```
var Weight;
```

```
class Species;
```

```
output out=summaryWeight;
```

```
run;
```

```
/*print output dataset*/
```

```
proc print data=summaryWeight;
```

The output dataset **summaryWeight** is now much more focused, containing only the summary statistics for the individual fish **Species**.

Obs	Species	_TYPE_	_FREQ_	_STAT_	Weight
1	Bream	1	35	N	34.00
2	Bream	1	35	MIN	242.00
3	Bream	1	35	MAX	1000.00
4	Bream	1	35	MEAN	626.00
5	Bream	1	35	STD	206.60
6	Parkki	1	11	N	11.00
7	Parkki	1	11	MIN	55.00
8	Parkki	1	11	MAX	300.00
9	Parkki	1	11	MEAN	154.82
10	Parkki	1	11	STD	78.76
11	Perch	1	56	N	56.00
12	Perch	1	56	MIN	5.90
13	Perch	1	56	MAX	1100.00
14	Perch	1	56	MEAN	382.24
15	Perch	1	56	STD	347.62

Reviewing the Output with NWAY Applied

A direct comparison between the two outputs reveals the efficacy of the **NWAY** option. The second output image, generated using the **NWAY** statement, is noticeably shorter. Crucially, the records

corresponding to the overall dataset total, where `_TYPE_` was equal to 0, have been completely omitted. The remaining rows all have a `_TYPE_` value of 1 (or the highest possible value based on the number of **CLASS** variables), confirming that the statistics relate only to the specified classification groups.

This targeted output is highly beneficial in environments where consistency and precision are paramount. For analysts working on comparative studies, having the output dataset contain only the distinct group metrics simplifies filtering and prevents errors that might arise from accidentally including grand totals in group comparisons. The **NWAY** option transforms PROC SUMMARY from a generalized aggregation tool into a precise instrument for group-level analysis.

In conclusion, mastering the **NWAY** option is a fundamental skill for advanced SAS programming. It provides an immediate and efficient way to manage the output complexity inherent in multi-level classification analyses, ensuring that only the desired statistics--those derived from the lowest, most specific group interaction--are retained for further examination or reporting. This contributes significantly to clean code practices and reliable data analysis workflows.

Further Learning and Related Topics in SAS

The principles demonstrated here using PROC SUMMARY and the NWAY option extend to other powerful SAS procedures, most notably **PROC FREQ** and **PROC TABULATE**, which also handle aggregation and classification. Understanding how grouping variables interact and how to control the output hierarchy is transferable across the entire SAS statistical ecosystem.

For those seeking to delve deeper into controlling aggregation output, exploring the **WAYS** and **TYPES** options within PROC SUMMARY is highly recommended. While **NWAY** selects only the highest level of interaction, **WAYS** allows the user to specify the number of classification variables to interact (e.g., `WAYS=2` would include all two-way interactions), and **TYPES** allows for the explicit selection of `_TYPE_` values to include. These options offer greater granularity than the simple **NWAY** switch.

The following tutorials explain how to perform other common tasks in SAS, building upon the foundational knowledge of data aggregation and classification presented here:

How to calculate percentiles using PROC SUMMARY.

Using **PROC FREQ** for generating frequency tables and cross-tabulations.

Advanced reporting with **PROC TABULATE** for complex hierarchical summaries.

You can use the **NWAY** statement in **PROC SUMMARY** in SAS to only calculate summary statistics at a group level rather than calculating them for an entire dataset.

The following example shows how to use the NWAY statement in practice.

Example: How to Use NWAY in PROC SUMMARY

For this example, we'll use the [SAS](#) built-in dataset called **sashelp.Fish**, which contains various measurements for 159 different fish caught in a lake in Finland.

We can use **PROC PRINT** to view the first 10 observations from this dataset:

```
/*view first 10 observations from Fish dataset*/
```

```
proc print data=sashelp.Fish (obs=10);
```

```
run;
```

Obs	Species	Weight	Length1	Length2	Length3	Height	Width
1	Bream	242	23.2	25.4	30.0	11.5200	4.0200
2	Bream	290	24.0	26.3	31.2	12.4800	4.3056
3	Bream	340	23.9	26.5	31.1	12.3778	4.6961
4	Bream	363	26.3	29.0	33.5	12.7300	4.4555
5	Bream	430	26.5	29.0	34.0	12.4440	5.1340
6	Bream	450	26.8	29.7	34.7	13.6024	4.9274
7	Bream	500	26.8	29.7	34.5	14.1795	5.2785
8	Bream	390	27.6	30.0	35.0	12.6700	4.6900
9	Bream	450	27.6	30.0	35.1	14.0049	4.8438
10	Bream	500	28.5	30.7	36.2	14.2266	4.9594

We can use the following code with **PROC SUMMARY** to calculate [descriptive statistics](#) for the variable **Weight**, grouped by the variable **Species**:

```
/*calculate descriptive statistics for Weight, grouped by Species*/
```

```
proc summary data=sashelp.Fish;
```

```
var Weight;
```

```
class Species;
```

```
output out=summaryWeight;
```

```
run;
```

```
/*print output dataset*/
```

```
proc print data=summaryWeight;
```

Obs	Species	_TYPE_	_FREQ_	_STAT_	Weight
1		0	159	N	158.00
2		0	159	MIN	0.00
3		0	159	MAX	1650.00
4		0	159	MEAN	398.70
5		0	159	STD	359.09
6	Bream	1	35	N	34.00
7	Bream	1	35	MIN	242.00
8	Bream	1	35	MAX	1000.00
9	Bream	1	35	MEAN	626.00
10	Bream	1	35	STD	206.60
11	Parkki	1	11	N	11.00
12	Parkki	1	11	MIN	55.00
13	Parkki	1	11	MAX	300.00
14	Parkki	1	11	MEAN	154.82
15	Parkki	1	11	STD	78.76
16	Perch	1	56	N	56.00
17	Perch	1	56	MIN	5.90
18	Perch	1	56	MAX	1100.00
19	Perch	1	56	MEAN	382.24
20	Perch	1	56	STD	347.62

Note: There are a total of 40 rows in the output but we've only taken a screenshot of the first 20 rows.

Here's how to interpret the output table:

TYPE: This column shows whether or not every row in the dataset was used to calculate the descriptive statistics. 0 = Every row was used (Grand Total).

FREQ: The number of rows used to calculate each descriptive statistic.

STAT: The name of the descriptive statistic (e.g., N, Mean, Std).

Weight: The numerical value for the corresponding descriptive statistic.

The first five rows show summary statistics for the entire dataset (**_TYPE_** = 0).

For example:

The total number of observations was **158**.

The minimum weight value was **0**.

The maximum weight value was **1,650**.

The mean weight value was **398.70**.

The standard deviation of weight values was **359.09**.

The next five rows show these summary statistics only for the rows in the dataset where the Species is equal to **Bream** (**_TYPE_** = 1).

And so on.

If we use the **NWAY** statement in PROC SUMMARY, we specify that we only want to display the rows with the highest value in the **_TYPE_** column of the output.

This means that only rows with a value of **1** in the **_TYPE_** column will be shown (in this single-classification example). In other words, the first five rows that show summary statistics for the entire dataset will no longer be shown.

The following code shows how to use the NWAY statement in practice:

```
/*calculate descriptive statistics for Weight, grouped by Species*/
```

```
proc summary data=sashelp.Fish nway;
```

```
var Weight;
```

```
class Species;
```

```
output out=summaryWeight;
```

```
run;
```

```
/*print output dataset*/
```

```
proc print data=summaryWeight;
```

Obs	Species	_TYPE_	_FREQ_	_STAT_	Weight
1	Bream	1	35	N	34.00
2	Bream	1	35	MIN	242.00
3	Bream	1	35	MAX	1000.00
4	Bream	1	35	MEAN	626.00
5	Bream	1	35	STD	206.60
6	Parkki	1	11	N	11.00
7	Parkki	1	11	MIN	55.00
8	Parkki	1	11	MAX	300.00
9	Parkki	1	11	MEAN	154.82
10	Parkki	1	11	STD	78.76
11	Perch	1	56	N	56.00
12	Perch	1	56	MIN	5.90
13	Perch	1	56	MAX	1100.00
14	Perch	1	56	MEAN	382.24
15	Perch	1	56	STD	347.62

Notice that the summary statistics for the entire dataset are no longer shown.

Only the summary statistics for the individual Species are shown.

The following tutorials explain how to perform other common tasks in [SAS](#):