

A Comprehensive Guide to Generating Summary Statistics in SAS with PROC SUMMARY and the NWAY Statement

Authored by
Mohammed loot

November 14, 2025

RECOMMENDED CITATION

Mohammed loot (2025). *A Comprehensive Guide to Generating Summary Statistics in SAS with PROC SUMMARY and the NWAY Statement*. PSYCHOLOGICAL STATISTICS.

Retrieved from <https://statistics.arabpsychology.com/?p=1462>

In the realm of statistical computing, particularly when leveraging the robust capabilities of [SAS](#), data analysts frequently require the generation of concise and highly targeted [summary statistics](#). The primary tool for this within the SAS environment is the [PROC SUMMARY](#) procedure, which efficiently aggregates large volumes of data and calculates essential descriptive measures. A frequent challenge arises because `PROC SUMMARY`'s default operation, known as "rollup" processing, yields a comprehensive, multi-level output. This output invariably includes statistics for the overall population in addition to the desired specific [group level](#) summaries. When the analytical goal is strictly limited to obtaining metrics at the most granular level of grouping, this excess information can significantly clutter the report, making interpretation challenging. The powerful, yet remarkably simple, statement, **NWAY**, offers an indispensable solution to this complexity.

When integrated into the [PROC SUMMARY](#) syntax, the **NWAY** statement acts as a precise filter for the analytical output. By invoking **NWAY**, the [SAS](#) system is specifically directed to compute and display results only for the combinations defined by the specified [class variables](#). This action systematically filters out all higher-level aggregations and overall totals, resulting in a dramatically cleaner, more focused report that addresses the exact [group levels](#) of interest, thereby significantly enhancing both the readability and interpretability of the summary results. This comprehensive guide will detail the practical application of the **NWAY** statement within standard statistical workflows. We will use a concrete, step-by-step example, contrasting the verbose default output generated by [PROC SUMMARY](#) with the streamlined results achieved when **NWAY** is properly employed. Mastering this fundamental utility is essential for efficient data analysis and professional reporting within the [SAS](#) environment.

Exploring the `SASHELP.FISH` Dataset

To effectively demonstrate the precise filtering capabilities afforded by the **NWAY** option, we will utilize a foundational, built-in [SAS dataset](#) called `SASHELP.FISH`. This dataset is widely recognized and frequently used as a standard teaching resource in SAS documentation and training modules, making it an ideal candidate for illustrating complex statistical and data manipulation procedures. The `SASHELP.FISH` [dataset](#) contains detailed biological measurements for 159 individual fish specimens collected from a lake in Finland, representing a diverse range of species. Each [observation](#) within the dataset includes several critical [variables](#), such as the fish's species, various length measurements, and its weight. This rich categorical and quantitative structure makes it perfectly suited for generating [group-level](#) summary statistics based on variables like species.

Before proceeding with any form of [aggregation](#) or summary calculation using [PROC SUMMARY](#), best practice in data analysis dictates a preliminary inspection to confirm the data's structure and content. We employ the [PROC PRINT](#) procedure to display the initial [observations](#), providing a rapid overview of the raw data we intend to analyze.

The following [SAS](#) code snippet executes this crucial preliminary step, displaying the first 10 observations extracted from the `SASHELP.FISH` 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

This initial examination confirms the presence and naming conventions of key [variables](#), such as `Species` and `Weight`. Gaining this deep understanding of the source data is a mandatory first step before performing any meaningful [aggregation](#) or generating summary statistics.

Default Behavior of PROC SUMMARY Without NWAY

When a user executes [PROC SUMMARY](#) without explicitly including the **NWAY** statement, the procedure is designed to produce [descriptive statistics](#) across all possible hierarchical combinations of the [CLASS statement variables](#). Crucially, it also computes statistics for the entire [dataset](#), which represents the highest level of [aggregation](#). This default multi-level output, often referred to as a "rollup," provides a comprehensive, albeit frequently overly detailed, perspective on the underlying data structure. For our illustration, our primary objective is singular: to calculate [descriptive statistics](#) specifically for the `Weight` variable, categorized and grouped distinctly by the `Species` variable. In the absence of the **NWAY** option, the resulting output will contain not only the statistics for each unique species (our desired output) but also statistics representing the overall aggregate data, combining all species together in a single summary.

Examine the following SAS code. It invokes [PROC SUMMARY](#) to calculate standard summary

statistics for `Weight`, categorized by `Species`. The `OUTPUT OUT=summaryWeight;` statement is essential, as it directs the calculated statistics into a new [output dataset](#) named `summaryWeight`. Subsequently, the [PROC PRINT](#) procedure is used to display the contents of this newly created [output dataset](#).

```
/*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

Although the screenshot above only displays the initial 20 rows, the full [output dataset](#),

`summaryWeight`, actually contains 40 distinct rows of data due to the hierarchical rollup. This visual snippet is sufficient, however, to clearly demonstrate the inherent structure and comprehensive content of the default, multi-level output produced by [PROC SUMMARY](#) when the **NWAY** statement is intentionally omitted.

Deciphering the Comprehensive Output Structure

The standard output produced by [PROC SUMMARY](#), when unrestricted by **NWAY**, delivers a rich array of information, meticulously organized into several key, automatically generated columns. These columns are essential metadata required for decoding the various [aggregation](#) levels present in the rollup. Understanding the precise role of each column is vital for accurate interpretation of the results:

TYPE: This unique, automatically generated [variable](#) is included by `PROC SUMMARY` to signify the specific level of [aggregation](#) represented by that row. A value of **0** invariably indicates summary statistics calculated for the entire [dataset](#)--the overall total. Progressively higher values correspond to summaries based on different combinations of the [CLASS statement](#) variables. The maximum value (**1** in this case, as we only used one grouping variable, `Species`) represents the most detailed [group level](#), where all defined [CLASS statement](#) variables are utilized for grouping.

FREQ: This column reports the number of [observations](#) or records that contributed to the calculation of the [descriptive statistics](#) for that particular [group level](#) or overall summary. Essentially, it functions as the frequency count of the records used in the calculation.

STAT: This [variable](#) clearly labels the specific [descriptive statistic](#) being reported in the corresponding row. Standard values reported by default include 'N' (the count of non-missing observations), 'MIN' (the minimum value), 'MAX' (the maximum value), 'MEAN' (the [arithmetic mean](#)), and 'STD' (the [standard deviation](#)).

Weight: Finally, this column presents the actual numerical result for the [descriptive statistic](#) indicated in the **_STAT_** column, specifically for the `Weight` [variable](#) under analysis.

Upon closer inspection of the default output, the initial rows, which all show **_TYPE_ = 0**, provide the summary statistics for the entire `SASHELP.FISH` [dataset](#), completely disregarding the grouping by `Species`. For instance, these overall rows immediately tell us that the total count of non-missing observations (N) for `Weight` was **158**, the minimum recorded weight was **0**, and the [mean](#) weight across all fish was approximately **398.70**. Following these comprehensive, overall statistics, the subsequent rows then display the identical set of [descriptive statistics](#), but meticulously calculated for each unique species, starting with 'Bream' and continuing for all other defined categories. This tiered, hierarchical presentation is the fundamental default behavior of [PROC SUMMARY](#).

Introducing the `NWAY` Statement for Focused Summaries

While the default hierarchical output of [PROC SUMMARY](#) is undoubtedly thorough, it frequently includes aggregate summary levels--like the overall total--that are superfluous or entirely irrelevant to a specific analytical question. If the sole purpose of an analysis is to conduct a direct comparison of summary statistics among the distinct [groups](#) defined by the [CLASS statement](#) variables, the inclusion of overall total statistics (the rollup) simply introduces unnecessary complexity and visual noise. The **NWAY** statement is deployed precisely to mitigate this issue.

This powerful option, utilized within `PROC SUMMARY`, explicitly instructs the [SAS](#) procedure to restrict the [aggregation](#) process. It limits the output to only the "N-way" combinations of the specified [CLASS statement](#) variables. Operationally, this means that only the rows corresponding to the highest possible value in the `_TYPE_` column of the output will be generated and displayed, thus eliminating the less granular summaries.

Considering our current example, which involves grouping by a single [CLASS statement](#) variable (`Species`), the highest `_TYPE_` value achievable is 1. Consequently, by activating **NWAY**, we ensure that SAS produces summary statistics solely for each individual species. This crucial filtering step completely eliminates the initial rows that represented the overall [dataset](#) summary (where `_TYPE_` = 0), resulting in a direct, concise, and focused report centered exclusively on inter-species comparisons.

Applying `NWAY` for Targeted Group Analysis

Integrating the **NWAY** statement into the [PROC SUMMARY](#) procedure is remarkably straightforward, requiring only the addition of the keyword `NWAY` directly onto the `PROC SUMMARY` statement itself. Despite this minimal change in syntax, the modification profoundly transforms the output structure, making it perfectly suited for analyses that require only detailed [group-level](#) summary statistics. Let us re-examine our running example using the `SASHELP.FISH` [dataset](#). We proceed to calculate [descriptive statistics](#) for the `Weight` [variable](#), grouped by `Species`. This time, however, the critical **NWAY** option is explicitly included in the procedure call. The subsequent [PROC PRINT](#) statement will then display the contents of the refined [output dataset](#), `summaryWeight`.

Note carefully the placement of `NWAY` within the `PROC SUMMARY` statement in the code provided below:

```
/*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

Reviewing the output generated with the **NWAY** statement reveals a stark and immediate difference: the summary statistics corresponding to the entire [dataset](#) (all rows where `_TYPE_ = 0`) are completely absent. Instead, the [output dataset](#) is now streamlined, containing only the summary statistics calculated individually for each unique `Species`. This results in an output that is significantly more manageable, concise, and directly relevant when the focus of the analysis is solely on comparing specific [group levels](#) against one another. The effective use of **NWAY** drastically reduces the verbosity inherent in the default `PROC SUMMARY` output, facilitating easier identification and interpretation of the statistics pertinent to your [CLASS statement](#) variables. This targeted approach is particularly advantageous in scenarios involving complex, multi-variable groupings, automated reporting requirements, or when dealing with extremely large [datasets](#) where comprehensive rollups would be overwhelmingly cumbersome.

Conclusion: Mastering **NWAY** for Streamlined Reporting

The **NWAY** statement stands as an indispensable feature within [PROC SUMMARY](#), providing

every [SAS](#) user with the ability to precisely refine and focus their statistical output. By ensuring that only the summary statistics corresponding to the most granular [group level](#), as defined by the [CLASS statement](#) variables, are displayed, **NWAY** successfully eliminates extraneous rollup statistics. The outcome is cleaner, more efficient, and far more readily interpretable results.

This powerful, targeted [aggregation](#) capability is especially beneficial in analytical contexts where precise, direct comparisons between categories are paramount, or when the outputs from `PROC SUMMARY` must be seamlessly integrated into subsequent automated reports or downstream analytical steps. Proficiency in utilizing **NWAY** empowers the analyst to produce highly focused and relevant [descriptive statistics](#), significantly enhancing the overall clarity and measurable impact of their data analysis within the SAS ecosystem. By mastering the strategic deployment of **NWAY**, you gain the assurance that your [PROC SUMMARY](#) results are not only accurate but also optimally structured for your specific analytical requirements, thereby saving considerable time and substantially improving the overall efficiency of your statistical workflows.

Additional Resources

To further enhance your technical proficiency in SAS programming and statistical analysis, we recommend exploring the following related tutorials that explain how to perform other common data manipulation and descriptive tasks:

How to Calculate [Mean](#) in SAS

How to Calculate [Standard Deviation](#) in SAS

Introduction to [PROC PRINT](#) in SAS