How to Easily Calculate Correlation Matrices in SAS with PROC CORR

Authored by

stats writer

December 1, 2025

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2025). *How to Easily Calculate Correlation Matrices in SAS with PROC CORR*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=103462

In data analysis and statistics, quantifying the relationship between variables is often a crucial first step. The PROC CORR procedure in SAS provides a powerful and straightforward method for calculating the correlation between two or more variables. This procedure is fundamental for generating insights into linear associations within your dataset.

When executed, PROC CORR generates a detailed output, including the correlation coefficients and the corresponding p-values for every pair of variables analyzed. This comprehensive report, often presented as a correlation matrix, helps analysts determine the strength and statistical significance of these relationships. Throughout this guide, we will provide practical, step-by-step examples demonstrating how to effectively utilize the PROC CORR procedure in SAS.

Understanding the Pearson Correlation Coefficient

One of the most common methods used to quantify the linear relationship between two continuous variables is the Pearson correlation coefficient. This statistic, denoted by r, measures the strength and direction of the linear association, providing an easily interpretable summary of how changes in one variable relate to changes in another. Understanding the range and interpretation of the coefficient is essential for accurate statistical reporting.

The Pearson correlation coefficient always assumes a value between **-1** and **1**. These boundaries provide critical context for interpreting the relationship:

-1 indicates a perfectly negative linear correlation between two variables. As one variable increases, the other decreases consistently.
0 indicates no linear correlation between two variables. They are linearly independent.
1 indicates a perfectly positive linear correlation between two variables. As one variable increases, the other increases consistently.

Crucially, the further the absolute value of the correlation coefficient is from zero (i.e., closer to -1 or 1), the stronger the linear relationship between the two variables is considered to be. A coefficient near 0.8 suggests a strong relationship, while one near 0.1 suggests a very weak relationship.

Preparing the Data: The SASHELP.FISH Dataset

To illustrate the application of PROC CORR, we will utilize a classic, built-in SAS dataset called `sashelp.Fish`. This dataset contains 159 observations, representing various physical measurements (such as Weight, Length, Height, and Width) collected from different species of fish caught in a lake in Finland. Analyzing the correlation among these measurements is useful for understanding the biological structure of the sampled fish population.

ARABPSYCHOLOGY.COM

Before performing the correlation analysis, it is good practice to inspect the structure and initial observations of the dataset. We can use the **proc print** procedure in SAS to view the first 10 observations from the `Fish` dataset, ensuring we understand the data structure and variable names we intend to analyze:

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

run;
```

The output provides a snapshot of the raw data, allowing us to confirm the presence of numeric variables suitable for correlation analysis.

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

ARABPSYCHOLOGY.COM

Example 1: Correlation Between Two Variables

Our first example focuses on calculating the linear relationship between only two variables: `Height` and `Width`. This is achieved by using the `VAR` statement within PROC CORR, specifying exactly which variables should be included in the calculation. This method is preferred when the analysis is targeted and resource efficiency is desired.

The following syntax instructs SAS to calculate the Pearson correlation coefficient using the `sashelp.fish` data, focusing exclusively on the `Height` and `Width` variables:

```
/*calculate correlation coefficient between Height and Width*/
proc corr data=sashelp.fish;
var Height Width;
```

run;

Executing this code yields detailed output tables that summarize the variables and present the final correlation statistics, which we will now examine in detail.

The CORR Procedure

2 Variables:	Height Width

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Height	159	8.97099	4.28621	1426	1.72840	18.95700
Width	159	4.41749	1.68580	702.38020	1.04760	8.14200

Pearson Correlation Coefficients, N = 159 Prob > |r| under H0: Rho=0		
	Height	Width
Height	1.00000	0.79288 <.0001
Width	0.79288 <.0001	1.00000

ARABPSYCHOLOGY.COM

Interpreting the Output of PROC CORR

The output generated by PROC CORR is typically divided into two main sections. The initial table provides descriptive summary statistics, such as the mean, standard deviation, minimum, and maximum values for both `Height` and `Width`. This is crucial for verifying data quality and distribution assumptions.

The second, and most critical, table displays the actual Pearson correlation coefficient between the two specified variables, alongside the associated p-value. The p-value is essential because it indicates whether the observed correlation is statistically significant, meaning we can reject the null hypothesis that the true population correlation is zero.

From the presented output for `Height` and `Width`, we observe the following key results:

Pearson correlation coefficient: **0.79288**
P-value: **<.0001**

This outcome leads to two important conclusions: First, the coefficient of 0.79288 indicates a **strong positive correlation** between Height and Width. Second, because the p-value is extremely

small (less than the standard significance level $alpha = .05$), we conclude that this correlation is **statistically significant**, strongly suggesting a genuine linear association between these two physical attributes.

Example 2: Correlation Between All Variables

While Example 1 focused on a specific pair, analysts often need to understand the relationships among all numeric variables in a dataset simultaneously. To calculate the correlation coefficient for every possible pairwise combination of variables, we simply omit the `VAR` statement when executing PROC CORR. This default setting instructs SAS to analyze all numeric variables found in the input dataset.

The required code is simplified substantially, making the process of generating a large-scale correlation analysis highly efficient:

```
/*calculate correlation coefficient between all pairwise combinations of variables*/
proc corr data=sashelp.fish;

run;
```

The resulting output is a comprehensive correlation matrix, which lists the Pearson correlation coefficients and corresponding p-values for each possible pairing of numeric variables within the dataset.

ARABPSYCHOLOGY.COM

The CORR Procedure

6 Variables:	Weight Length1 Length2 Length3 Height Width

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Weight	158	398.69557	359.08620	62994	0	1650
Length1	159	26.24717	9.99644	4173	7.50000	59.00000
Length2	159	28.41572	10.71633	4518	8.40000	63.40000
Length3	159	31.22704	11.61025	4965	8.80000	68.00000
Height	159	8.97099	4.28621	1426	1.72840	18.95700
Width	159	4.41749	1.68580	702.38020	1.04760	8.14200

Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Weight	Length1	Length2	Length3	Height	Width
Weight	1.00000 158	0.91644 <.0001 158	0.91937 <.0001 158	0.92447 <.0001 158	0.72869 <.0001 158	0.88741 <.0001 158
Length1	0.91644 <.0001 158	1.00000 159	0.99952 <.0001 159	0.99203 <.0001 159	0.62538 <.0001 159	0.86705 <.0001 159
Length2	0.91937 <.0001 158	0.99952 <.0001 159	1.00000 159	0.99410 <.0001 159	0.64044 <.0001 159	0.87355 <.0001 159
Length3	0.92447 <.0001 158	0.99203 <.0001 159	0.99410 <.0001 159	1.00000 159	0.70341 <.0001 159	0.87852 <.0001 159
Height	0.72869 <.0001 158	0.62538 <.0001 159	0.64044 <.0001 159	0.70341 <.0001 159	1.00000 159	0.79288 <.0001 159
Width	0.88741 <.0001 158	0.86705 <.0001 159	0.87355 <.0001 159	0.87852 <.0001 159	0.79288 <.0001 159	1.00000 159

ARABPSYCHOLOGY.COM

Reviewing the matrix allows for rapid identification of strong relationships. For example, examining the correlations involving `Weight` reveals exceptionally strong positive associations with length measurements:

The Pearson correlation coefficient between Weight and Length1 is **0.91644**
The Pearson correlation coefficient between Weight and Length2 is **0.91937**
The Pearson correlation coefficient between Weight and Length3 is **0.92447**

These coefficients confirm that as any measure of length increases, the weight of the fish also increases proportionally, indicating strong structural dependency between these variables.

Example 3: Visualize Correlation with a Scatterplot

While numerical correlation coefficients provide precise metrics, visualizing the relationship through a scatterplot offers immediate, intuitive understanding. PROC CORR supports graphical output through the use of the `PLOTS` option, allowing you to generate scatterplots directly alongside the statistical tables.

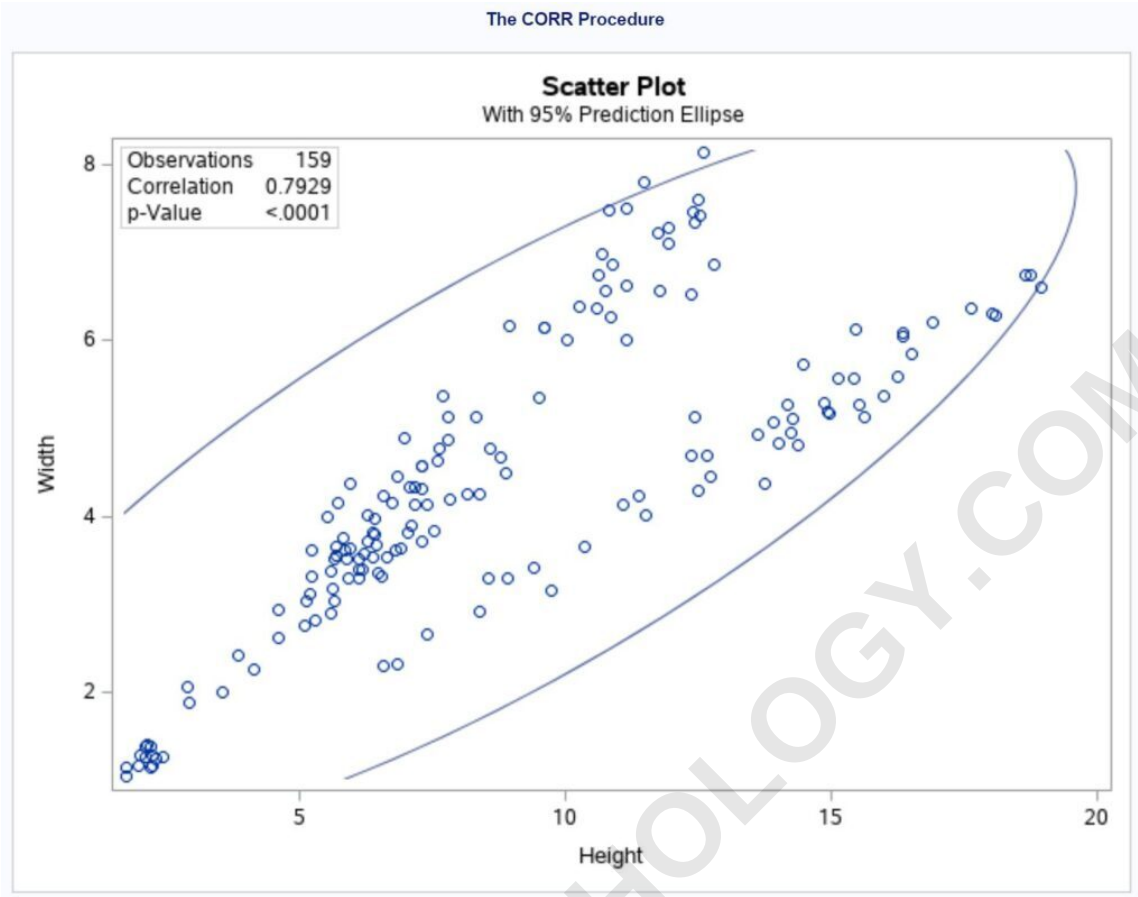
To visualize the strong correlation previously found between `Height` and `Width`, we append the `PLOTS=scatter(nvar=all)` option to the `PROC CORR` statement. The `nvar=all` sub-option ensures that scatterplots are generated for all specified pairs in the `VAR` statement, providing a clear visual representation of the linear trend:

```
/*visualize correlation between Height and Width*/
proc corr data=sashelp.fish plots=scatter(nvar=all);;
var Height Width;

run;
```

This visual confirmation is often critical, as it can help identify outliers, non-linear patterns, or clusters that might not be fully captured by a single Pearson coefficient.

ARABPSYCHOLOGY.COM



The resulting plot clearly displays the **strong positive correlation**. The data points cluster tightly around an upward sloping line, confirming that as the fish's `Height` increases, its `Width` tends to increase significantly. Furthermore, the plot includes key summary statistics in the top left corner, reiterating the total observations used, the calculated correlation coefficient, and the highly significant p-value.

Further SAS Operations and Analysis

Mastering correlation analysis using PROC CORR is just one step in leveraging the full power of SAS for statistical computing. Understanding these core procedures enables more complex modeling and hypothesis testing.

The following tutorials explain how to perform other common and essential data analysis operations in SAS:

ARABPSYCHOLOGY.COM