

# How to Create a Correlation Matrix in SAS (With Example)

Authored by  
**stats writer**

November 19, 2025

## RECOMMENDED CITATION

stats writer (2025). *How to Create a Correlation Matrix in SAS (With Example)*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=96979>

A correlation matrix in SAS is a statistical technique used to measure the strength and direction of relationships between two or more variables. To create a correlation matrix in SAS, you can use the PROC CORR procedure. This procedure requires the user to specify the variables that will be included in the analysis and the type of correlation to be calculated. An example of using the PROC CORR procedure is provided, demonstrating how to calculate a Pearson correlation coefficient between two numeric variables.

A is a square table that shows the between variables in a dataset.

It offers a quick way to understand the strength of the linear relationships that exist between variables in a dataset.

You can use the **PROC CORR** statement in SAS to create a correlation matrix for a given dataset:

```
/*create correlation matrix using all numeric variables in my_data*/  
proc corr data=my_data;  
run;
```

By default, this will create a matrix that displays the correlation coefficients between all numeric variables in the dataset.

To only include specific variables in the correlation matrix, you can use the **VAR** statement:

```
/*create correlation matrix using only var1, var2 and var3 in my_data*/  
proc corr data=my_data;  
var var1, var2, var3;  
run;
```

The following example shows how to create a correlation matrix in SAS.

## Example: Creating a Correlation Matrix in SAS

Suppose we have the following dataset in SAS that contains information about various basketball players:

```
/*create dataset*/  
data my_data;  
input team $ assists rebounds points;  
datalines;  
A 4 12 22
```

```
A 5 14 24  
A 5 13 26  
A 6 7 26  
B 7 8 29  
B 8 8 32  
B 8 9 20  
B 10 13 14
```

```
;
```

```
run;
```

```
/*view dataset*/
```

```
proc print data=my_data;
```

Obs	team	assists	rebounds	points
1	A	4	12	22
2	A	5	14	24
3	A	5	13	26
4	A	6	7	26
5	B	7	8	29
6	B	8	8	32
7	B	8	9	20
8	B	10	13	14

We can use the **PROC CORR** statement to create a correlation matrix that includes each numeric variable in the dataset by default:

```
/*create correlation matrix using all numeric variables in my_data*/
```

```
proc corr data=my_data;
```

```
run;
```

## The CORR Procedure

3 Variables: assists rebounds points

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
assists	8	6.62500	1.99553	53.00000	4.00000	10.00000
rebounds	8	10.50000	2.77746	84.00000	7.00000	14.00000
points	8	24.12500	5.56616	193.00000	14.00000	32.00000

Pearson Correlation Coefficients, N = 8 Prob >  r  under H0: Rho=0			
	assists	rebounds	points
assists	1.00000	-0.24486 0.5589	-0.32957 0.4253
rebounds	-0.24486 0.5589	1.00000	-0.52209 0.1844
points	-0.32957 0.4253	-0.52209 0.1844	1.00000

The output displays summary statistics of the numeric variables in the first table along with a correlation matrix.

Here is how to interpret the values in the correlation matrix:

(1) The Pearson correlation coefficient ( $r$ ) between **assists** and **rebounds** is **-0.24486**. The corresponding p-value is **0.5589**.

Since  $r$  is less than zero, this tells us that there is a negative linear association between these two variables. However, the p-value is not less than .05 so this correlation is not statistically significant.

(2) The Pearson correlation coefficient ( $r$ ) between **assists** and **points** is **-0.32957**. The corresponding p-value is **0.4253**.

There is a negative linear association between these two variables but it is not statistically significant.

(3) The Pearson correlation coefficient ( $r$ ) between **rebounds** and **points** is **-0.52209**. The corresponding p-value is **0.1844**.

There is a negative linear association between these two variables but it is not statistically significant.

Note that we could also use the **VAR** statement to only include specific numeric variables in the correlation matrix:

```
/*create correlation matrix using only assists and rebounds variables*/
proc corr data=my_data;
var assists rebounds;
run;
```

**The CORR Procedure**

2 Variables: assists rebounds

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
assists	8	6.62500	1.99553	53.00000	4.00000	10.00000
rebounds	8	10.50000	2.77746	84.00000	7.00000	14.00000

  

Pearson Correlation Coefficients, N = 8 Prob >  r  under H0: Rho=0		
	assists	rebounds
assists	1.00000	-0.24486 0.5589
rebounds	-0.24486 0.5589	1.00000

Notice that only the **assists** and **rebounds** variables were included in this correlation matrix.

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