How can I calculate correlation in SAS? Can you provide some examples?

Authored by

stats writer

July 1, 2024

SAS is a statistical software that allows users to calculate correlations between variables. To calculate the correlation in SAS, one can use the PROC CORR procedure. This procedure generates a correlation matrix that displays the relationships between all selected variables. Users can also specify the type of correlation coefficient they want to calculate, such as Pearson's or Spearman's. Additionally, SAS allows for the inclusion of categorical variables in the correlation analysis. For example, one can calculate the correlation between income and education level, while also considering gender as a categorical variable. Overall, SAS provides a flexible and robust method for calculating correlations and can be utilized for various research and data analysis purposes.

Calculate Correlation in SAS (With Examples)

One way to quantify the relationship between two variables is to use the , which measures the linear association between two variables.

It always takes on a value between -1 and 1 where:

-1 indicates a perfectly negative linear correlation between two variables0 indicates no linear correlation between two variables1 indicates a perfectly positive linear correlation between two variables

The further away the correlation coefficient is from zero, the stronger the relationship between the two variables.

The following examples show how to use proc corr in SAS to calculate the correlation coefficient between variables in the SAS built-in dataset called , 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 printdata=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

Example 1: Correlation Between Two Variables

We can use the following code to calculate the Pearson correlation coefficient between the variables Height and Width:

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

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

The first table displays summary statistics for both Height and Width.

The second table displays the Pearson correlation coefficient between the two variables, including a that tells us if the correlation is statistically significant.

From the output we can see:

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

This tells us that there is a strong positive correlation between Height and Width and that the correlation is statistically significant since the p-value is less than α = .05.

Example 2: Correlation Between All Variables

We can use the following code to calculate the Pearson correlation coefficient between all pairwise combinations of variables in the dataset:

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

run;
```

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

The output shows a , which contains the Pearson correlation coefficient and corresponding p-values for each pairwise combination of numeric variables in the dataset.

For example:

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

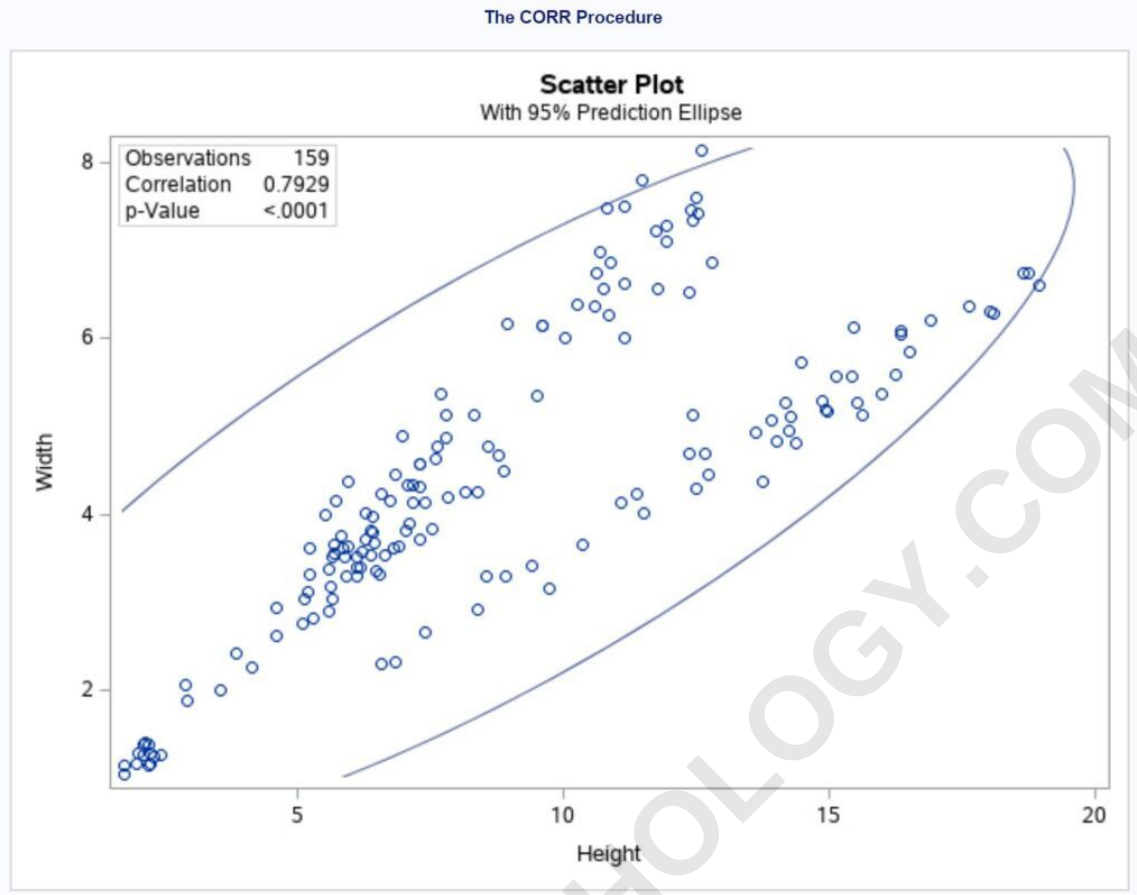
And so on.

Example 3: Visualize Correlation with a Scatterplot

We can also use the plots function to create a scatterplot to visualize the correlation between two variables:

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

run;
```



From the plot we can see the strong positive correlation between Height and Width. As Height increases, Width tends to increase as well.

In the top left corner of the plot we can also see the total observations used, the correlation coefficient, and the p-value for the correlation coefficient.

Additional Resources

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