MET MA 603: SAS Programming and Applications

Proc Corr

Statistical Correlation

In statistics, correlation measures the strength of the linear relationship between two variables.

Strong correlation indicates that the value of one variable is a good predictor of the value of the other.

Below is one way of describing the strength of correlations. Except for Perfect Correlation and Perfectly Uncorrelated, these are all subjective definitions, and do not need to be followed or memorized. Use only as a guide.

Range	Standard	Range	Standard
0	Perfectly Uncorrelated		
0 to 0.1	No or Weak Correlation	-0.1 to 0	No or Weak Negative Correlation
0.1 to 0.3	Weak Correlation	-0.3 to -0.1	Weak Negative Correlation
0.3 to 0.5	Moderate Correlation	-0.5 to -0.3	Moderate Negative Correlation
0.5 to 1	Strong Correlation	-1 to -0.5	Strong Negative Correlation
1	Perfect Correlatation	-1	Perfect Negative Correlation

The Correlation Procedure

The **Correlation Procedure** calculates the correlation between pairs of numeric variables in a SAS dataset.

```
proc corr data = data3.golf_total ;
run;
```

The Correlation Procedure (cont.)

By default, all numeric variables are included in the analysis. The VAR statement is used to list the variables to include in the analysis.

By default, all pairs of correlations are calculated and displayed in a 2-way table. The **WITH Statement** is used to specify the variables that appear down the side of the table.

The **PLOTS Option** creates plots in order to help visualize the relationship between the variables.


```
proc corr data = data3.golf_total
plots=matrix(histogram) ;
var  hole1 - hole3 ;
with total_score ;
run;
```

Practice

There is a theory in politics that when the Washington football team wins their last game before the election, that the incumbent party (i.e., the party in power) wins the election, and when Washington loses, the incumbent party loses. The dataset election.sas7bdat has the results from 1932 through 2016.

First use Proc Freq to check the data. Address any issues with the data. Then use Proc Corr to test the correlation between Washington's result and the election result.

Do you think this is a good theory?

Readings

- Textbook section 9.8