Bloomsburg University: Department of Psychology
Applied Statistics for the Behavioral Sciences
Dr. Jeffrey D. Leitzel

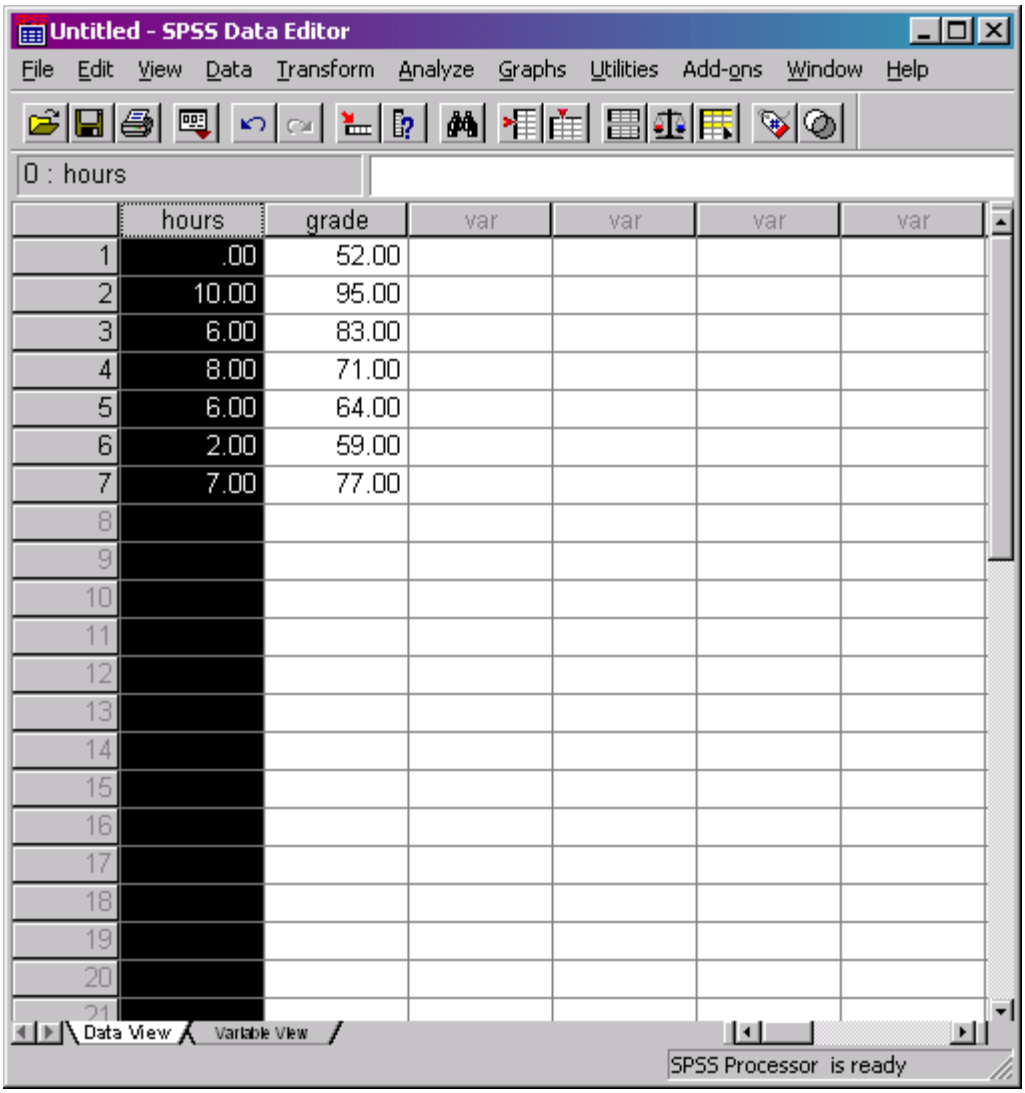
Setting up correlation and regression problems in SPSS

First, the raw data must be presented in this, or a similar form. The important feature is that the values of each variable belonging to each individual case can be paired with one another.

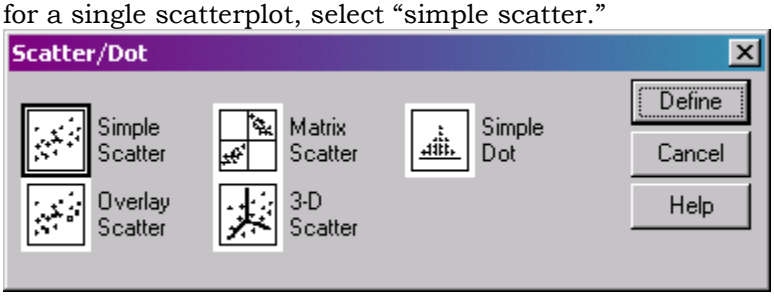
Table 1
Hours spent studying and test grades for seven students

Student	Hours Studied	Test Grade
1	0	52
2	10	95
3	6	83
4	8	71
5	6	64
6	2	59
7	7	77

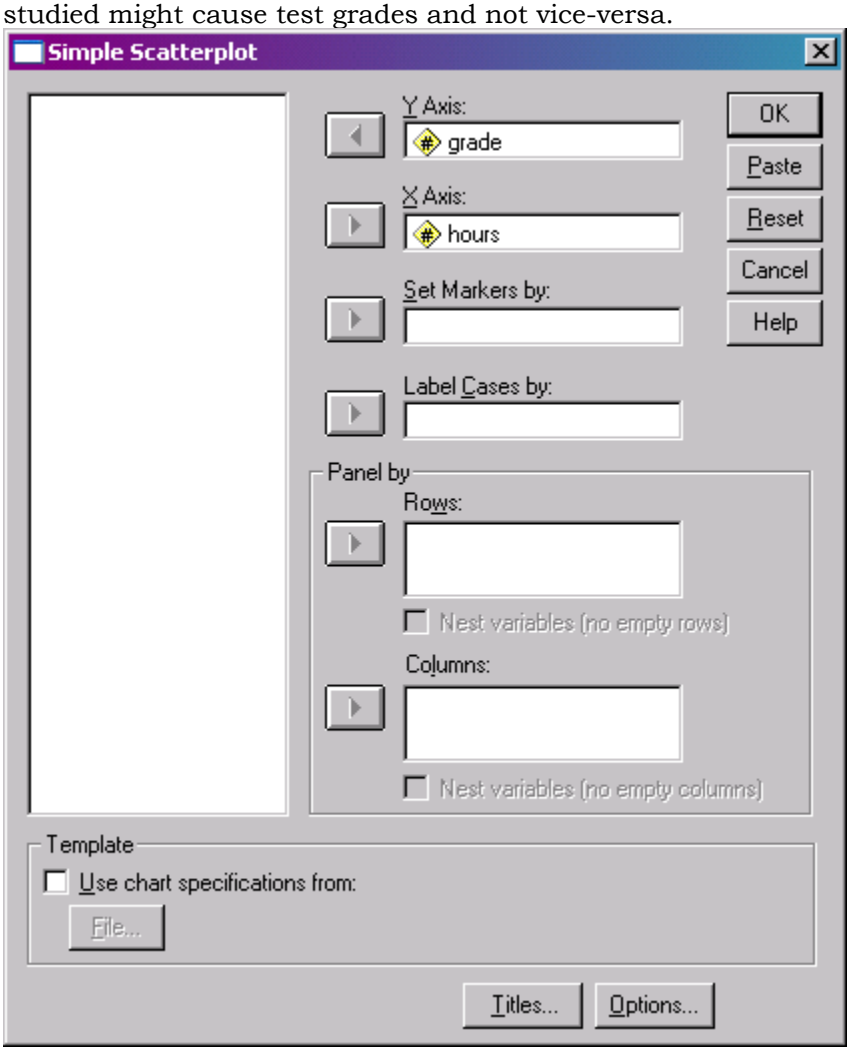
Setting the data up in SPSS simply involves maintaining the correspondence between the data elements for each case. Here is what the data editor should look like after the data above have been correctly entered.



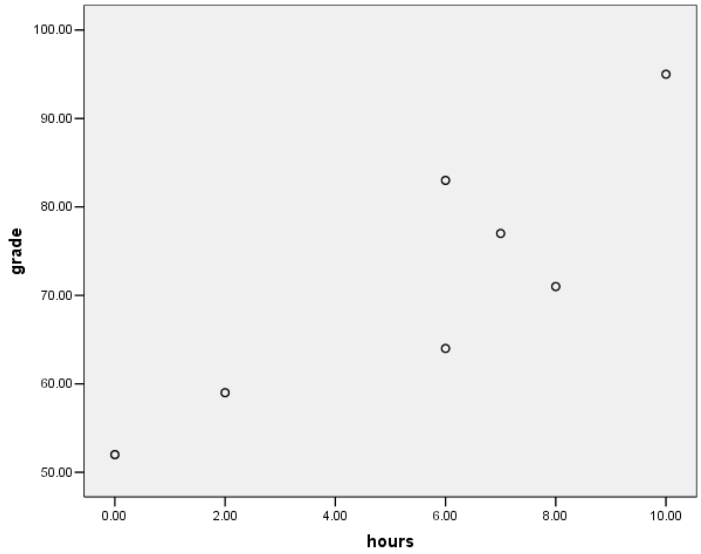
To generate a scatterplot select the Graphs->Legacy Dialogs -> Scatter/Dot option. In the menu that results, for a single scatterplot, select “simple scatter.”



When a distinction can be made, convention is to place the variable that is more likely to be a “cause” on the X axis and the “effect” variable on the Y axis. In this case, it is clearly reasonable to believe that hours studied might cause test grades and not vice-versa.

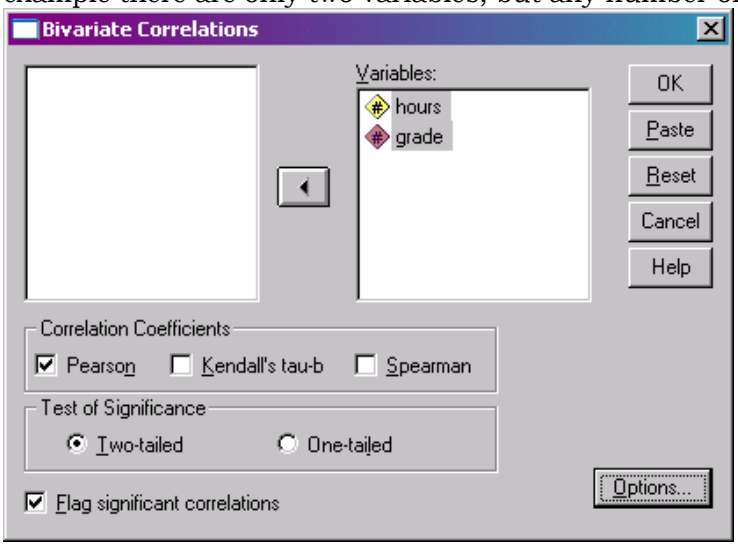


The following graph results:



To obtain the correlation coefficients select the following menus:
Analyze->Correlate->Bivariate

Then move the variables that you need to have correlations calculated for into the variables window, in this example there are only two variables, but any number of variables can be run simultaneously:



And the results:

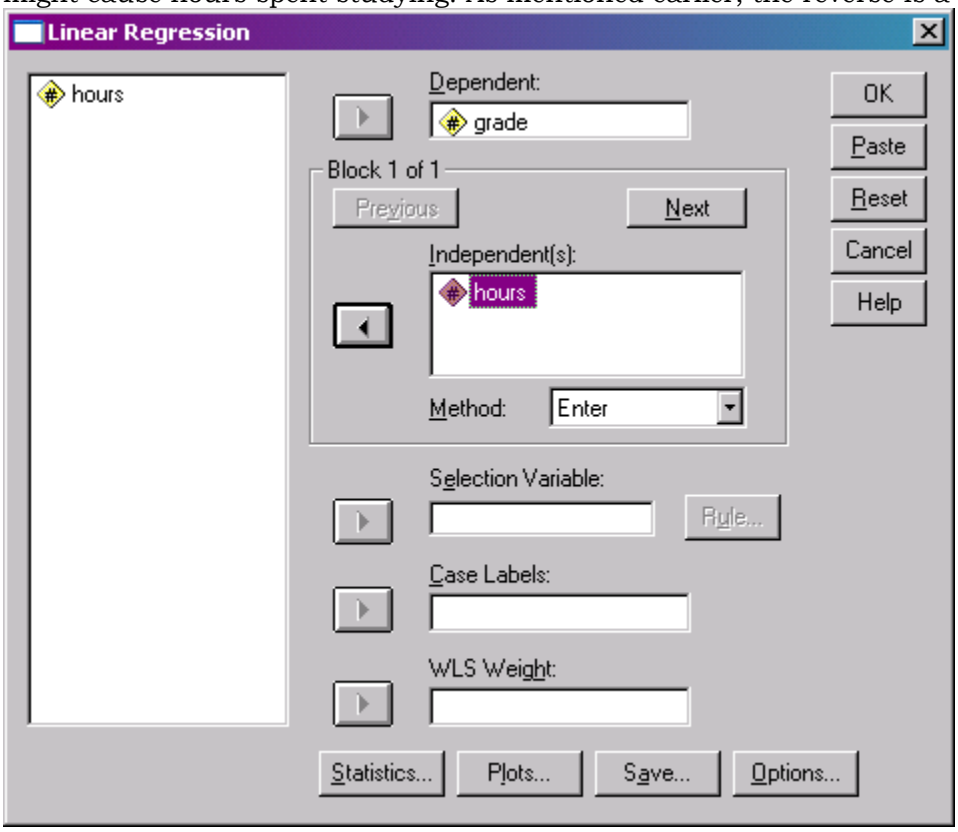
Correlations

		hours	grade
hours	Pearson Correlation	1	.868(*)
	Sig. (2-tailed)		.011
	N	7	7
grade	Pearson Correlation	.868(*)	1
	Sig. (2-tailed)	.011	
	N	7	7

* Correlation is significant at the 0.05 level (2-tailed).

Next, to obtain our regression coefficients, we use the following menus:
Analyze->Regression->Linear

Unlike when we calculate correlation coefficients, it does matter which variable we consider our dependent (Y) and which variable is our independent (X) variable. In this case it makes no sense to think that test grades might cause hours spent studying. As mentioned earlier, the reverse is a very reasonable theory of causation.



And the results:

The parts of the output that you need to concern yourself with for now will be indicated by boxes like this one.

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	hours[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: grade

R is the same correlation coefficient we calculated earlier

R^2 is the coefficient of determination (proportion of shared variability between the two variables

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.868[a]	.753	.703	8.03984

a. Predictors: (Constant), hours

For now, you do not need to pay any attention to this table. We'll get to the joys of ANOVA later! ☺

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	984.519	1	984.519	15.231	.011[a]
	Residual	323.195	5	64.639		
	Total	1307.714	6			

a. Predictors: (Constant), hours

b. Dependent Variable: grade

This is the table where you will find your regression coefficients

Y-intercept (a)

Coefficients[a]

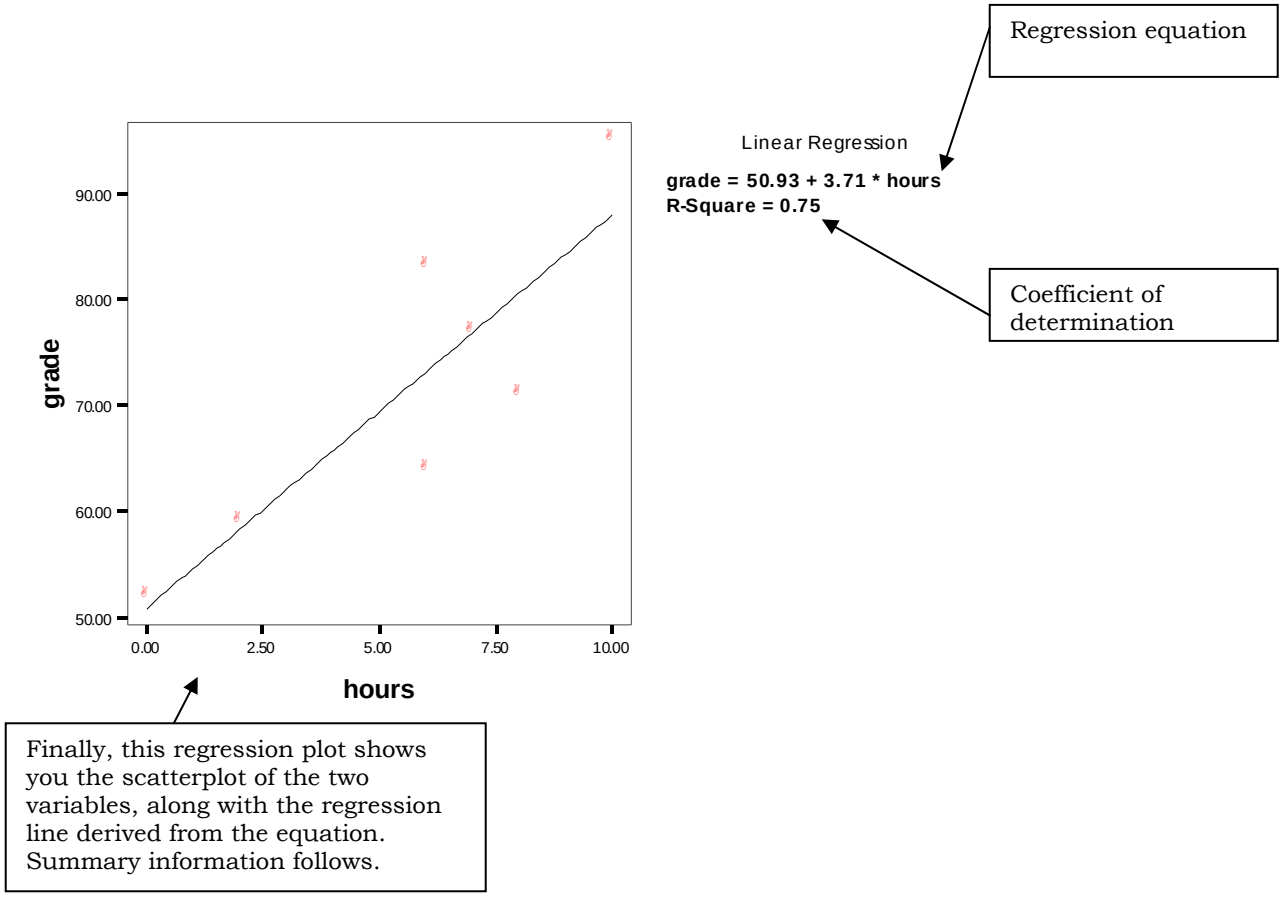
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	50.928	6.100		8.349	.000
	hours	3.705	.949	.868	3.903	.011

a. Dependent Variable: grade

Slope of regression line (b)

All you have to worry about in this table is where to find a (Y-intercept) and b (slope) these are listed in the unstandardized coefficients column.

This graph was generated using the Graphs->Interactive->Scatterplot menu (Not available in new version of SPSS)



The following shows the set up in the interactive graphs procedure to get the plot above