Using SPSS for Correlation

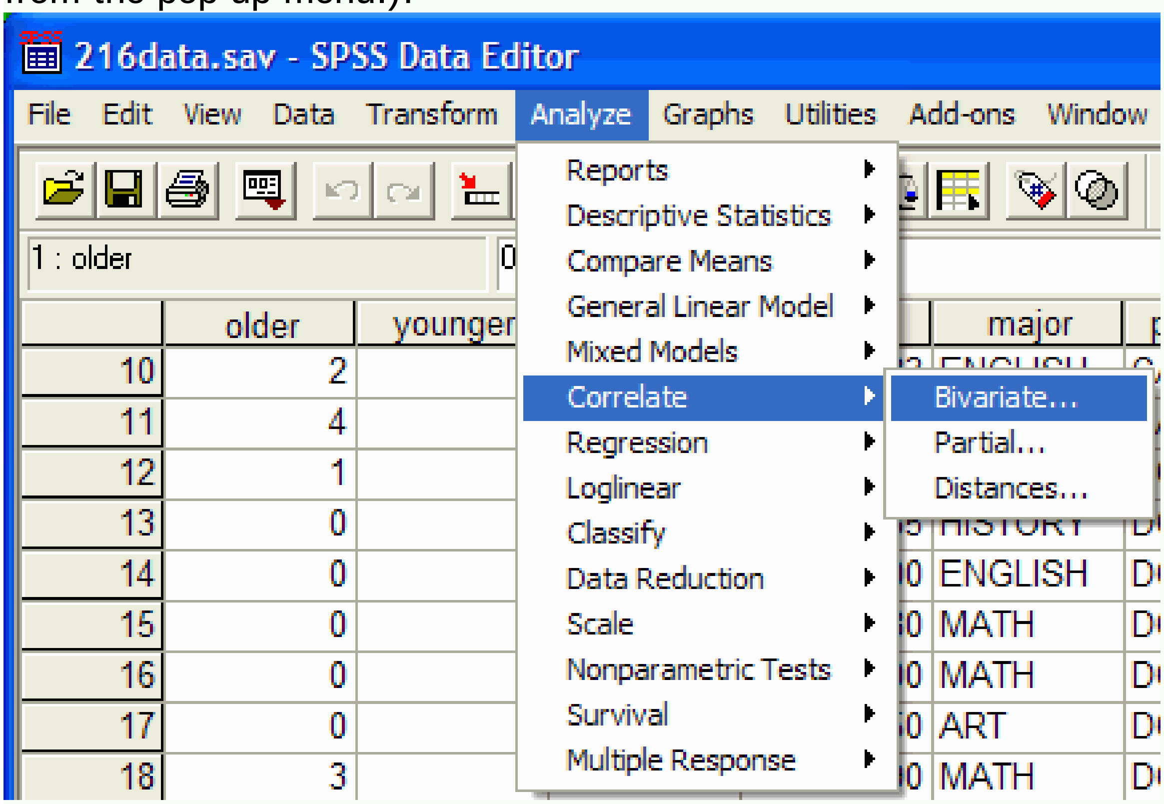
This tutorial will show you how to use SPSS version 12.0 to perform bivariate correlations. You will use SPSS to calculate Pearson's r.

This tutorial assumes that you have:

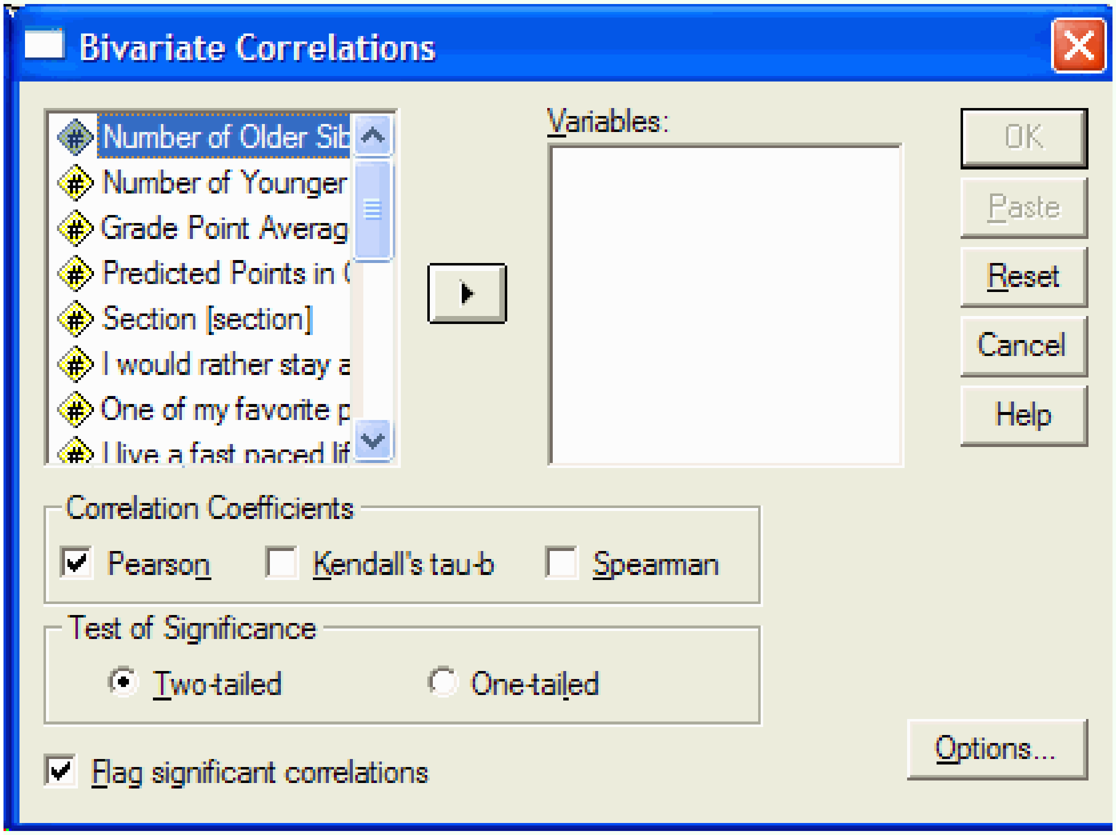
- Downloaded the standard class data set (click on the link and save the data file)
- Started SPSS (click on Start | Programs | SPSS for Windows | SPSS 12.0 for Windows)
- Loaded the standard data set

Bivariate Correlation

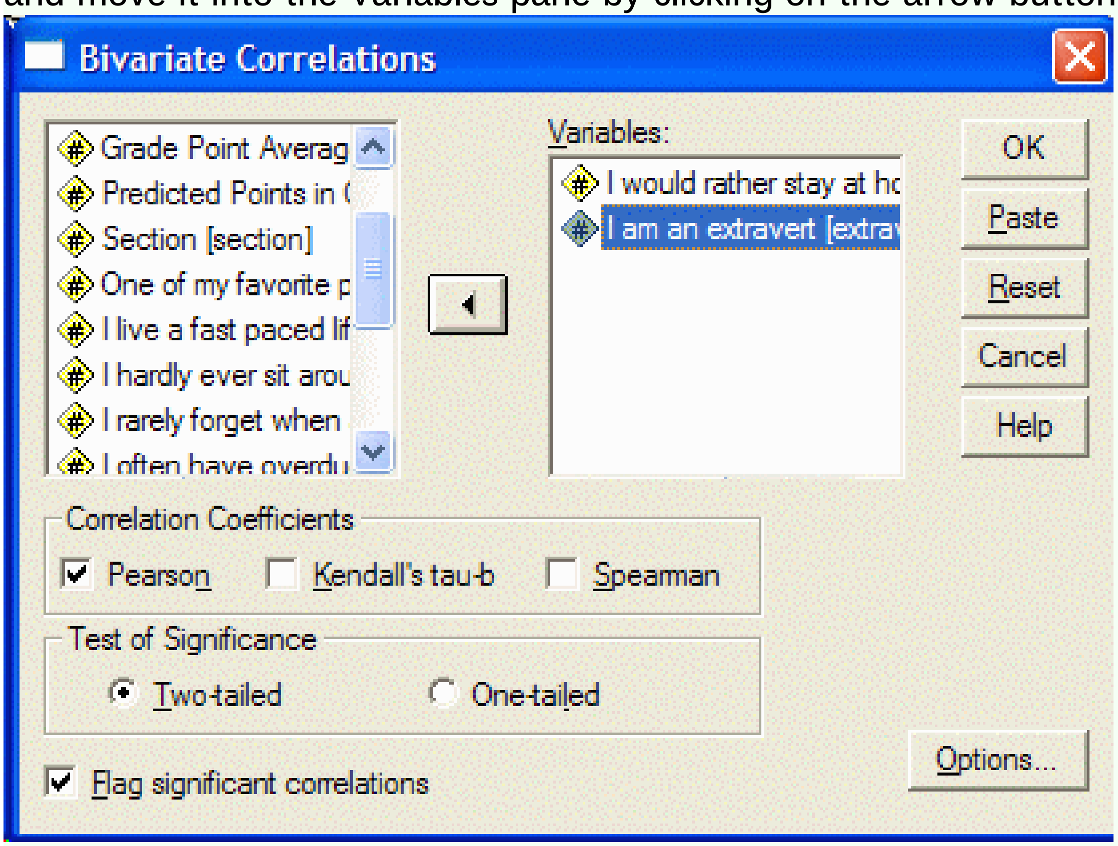
Bivariate correlation can be used to determine if two variables are linearly related to each other. Remember that you will want to perform a scatter plot before performing the correlation (to see if the assumptions have been met.) The command for correlation is found at Analyze | Correlate | Bivariate (this is shorthand for clicking on the Analyze menu item at the top of the window, and then clicking on Correlate from the drop down menu, and Bivariate from the pop up menu.):



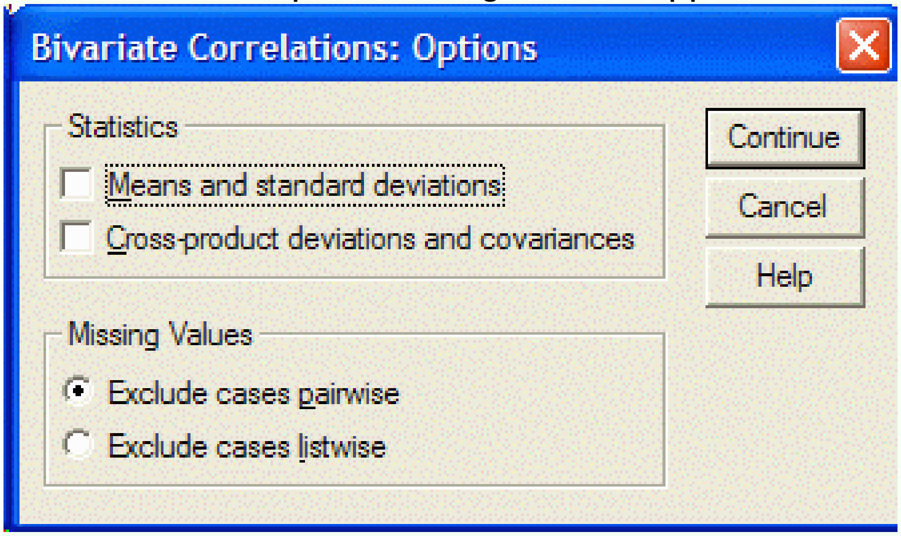
The Bivariate Correlations dialog box will appear:



Select one of the variables that you want to correlate by clicking on it in the left hand pane of the Bivariate Correlations dialog box. Then click on the arrow button to move the variable into the Variables pane. Click on the other variable that you want to correlate in the left hand pane and move it into the Variables pane by clicking on the arrow button:



Specify whether the test of significance should be one-tailed or two-tailed. (We won't get to this topic for quite a while. For now, select the one-tailed test by clicking on the circle to the left of "one-tailed".) You can click on the Options button to have some descriptive statistics calculated. The Options dialog box will appear:



From the Options dialog box, click on "Means and standard deviations" to get some common descriptive statistics. Click on the Continue button in the Options dialog box. Click on OK in the Bivariate Correlations dialog box. The SPSS Output Viewer will appear.

In the SPSS Output Viewer, you will see a table with the requested descriptive statistics and correlations. This is what the Bivariate Correlations output looks like:

Descriptive Statistics

	Mean	Std. Deviation	N
I would rather stay at home and read than go out with my friends	4.11	.900	46
I am an extravert	2.52	1.005	46

Correlations

		I would rather stay at home and read than go out with my friends	I am an extravert
I would rather stay at home and read than go out with my friends	Pearson Correlation	1	-.310*
	Sig. (1-tailed)	.	.018
	N	46	46
I am an extravert	Pearson Correlation	-.310*	1
	Sig. (1-tailed)	.018	.
	N	46	46

*. Correlation is significant at the 0.05 level (1-tailed).

The Descriptive Statistics section gives the mean, standard deviation, and number of observations (N) for each of the variables that you specified. For example, the mean of the extravert variable is 2.52, the standard deviation of the rather stay at home variable is 0.900, and there were 46 observations (N) for each of the two variables.

The Correlations section gives the values of the specified correlation tests, in this case, Pearson's r. Each row of the table corresponds to one of the variables. Each column also corresponds to one of the variables. In this example, the cell at the bottom row of the right column represents the correlation of extravert with extravert. Kind of silly, isn't it! This correlation must always be 1.0 (why?). Likewise the cell at the middle row of the middle column represents the correlation of rather stay at home with rather stay at home. It too, must always be 1.0. The cell at middle row and right column (or equivalently, the bottom row at the middle column) is more interesting. This cell represents the correlation of extravert and rather stay at home (or rather stay at home with extravert -- it doesn't matter. Why?) There are three numbers in these cells. The top number is the correlation coefficient. The correlation coefficient in this example is -0.310. The middle number is the significance of this correlation; in this case, it is .018. (The significance basically tells us whether we would expect a correlation that was this large purely due to chance factors and not due to an actual relation.

In this case, it is improbable that we would get an r this big if there was not a relation between the variables.) The bottom number, 46 in this example, is the number of observations that were used to calculate the correlation coefficient.