

# Using SPSS for Linear Regression

---

This tutorial will show you how to use SPSS version 12.0 to perform linear regression. You will use SPSS to determine the linear regression equation.

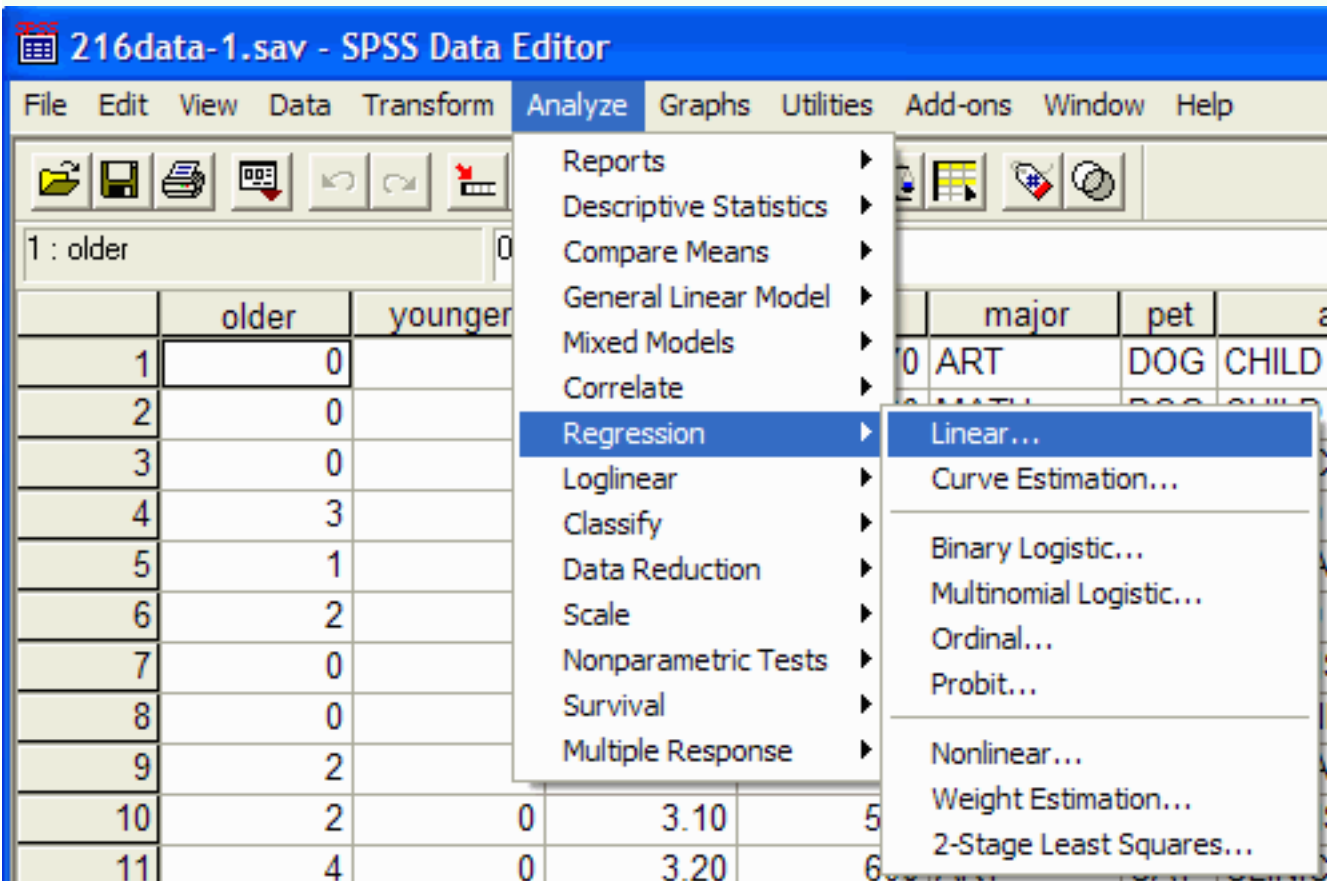
This tutorial assumes that you have:

- 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)

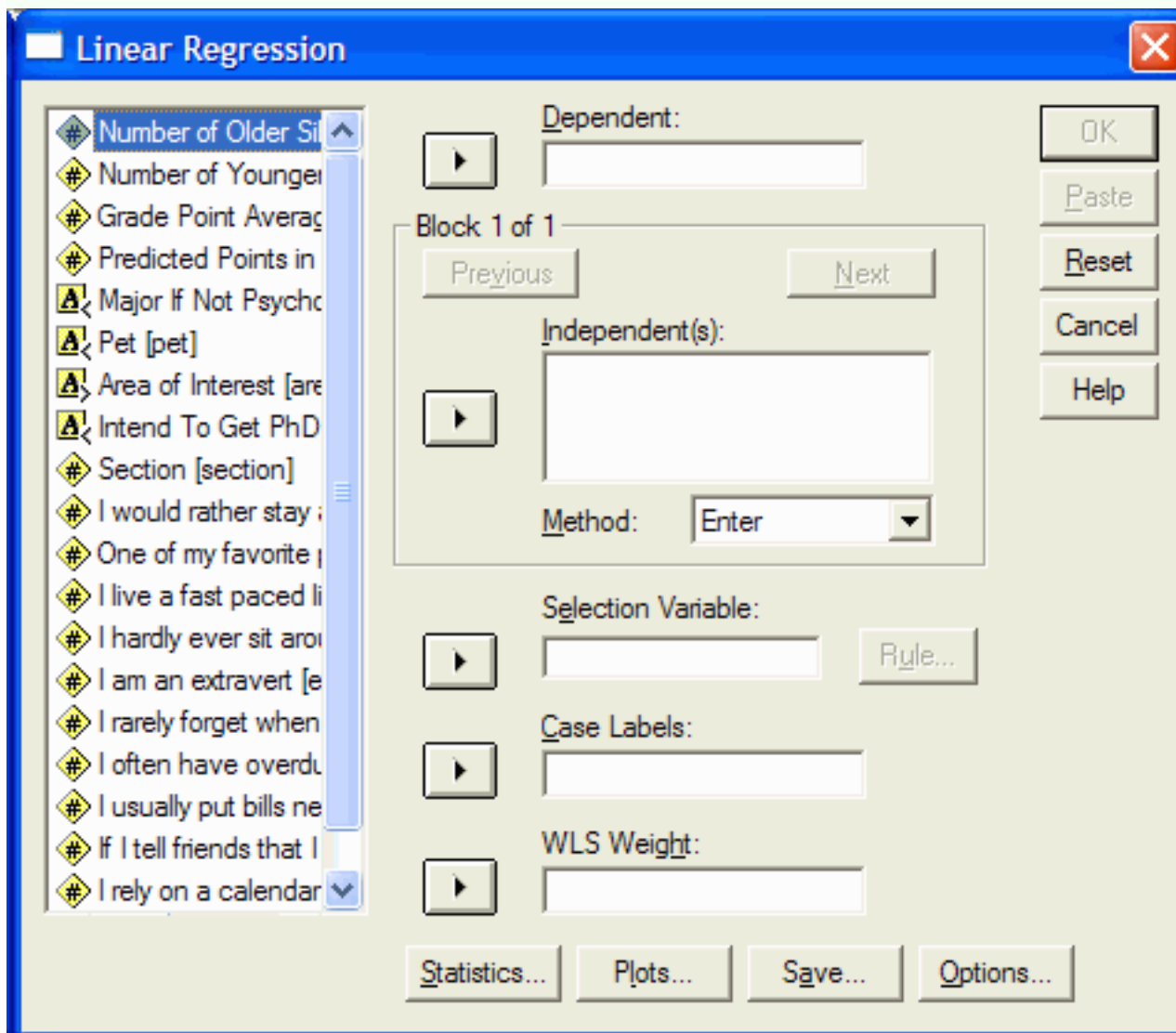
## Linear Regression

Linear regression is used to specify the nature of the relation between two variables. Another way of looking at it is, given the value of one variable (called the independent variable in SPSS), how can you predict the value of some other variable (called the dependent variable in SPSS)? Remember that you will want to perform a scatterplot and correlation before you perform the linear regression (to see if the assumptions have been met.)

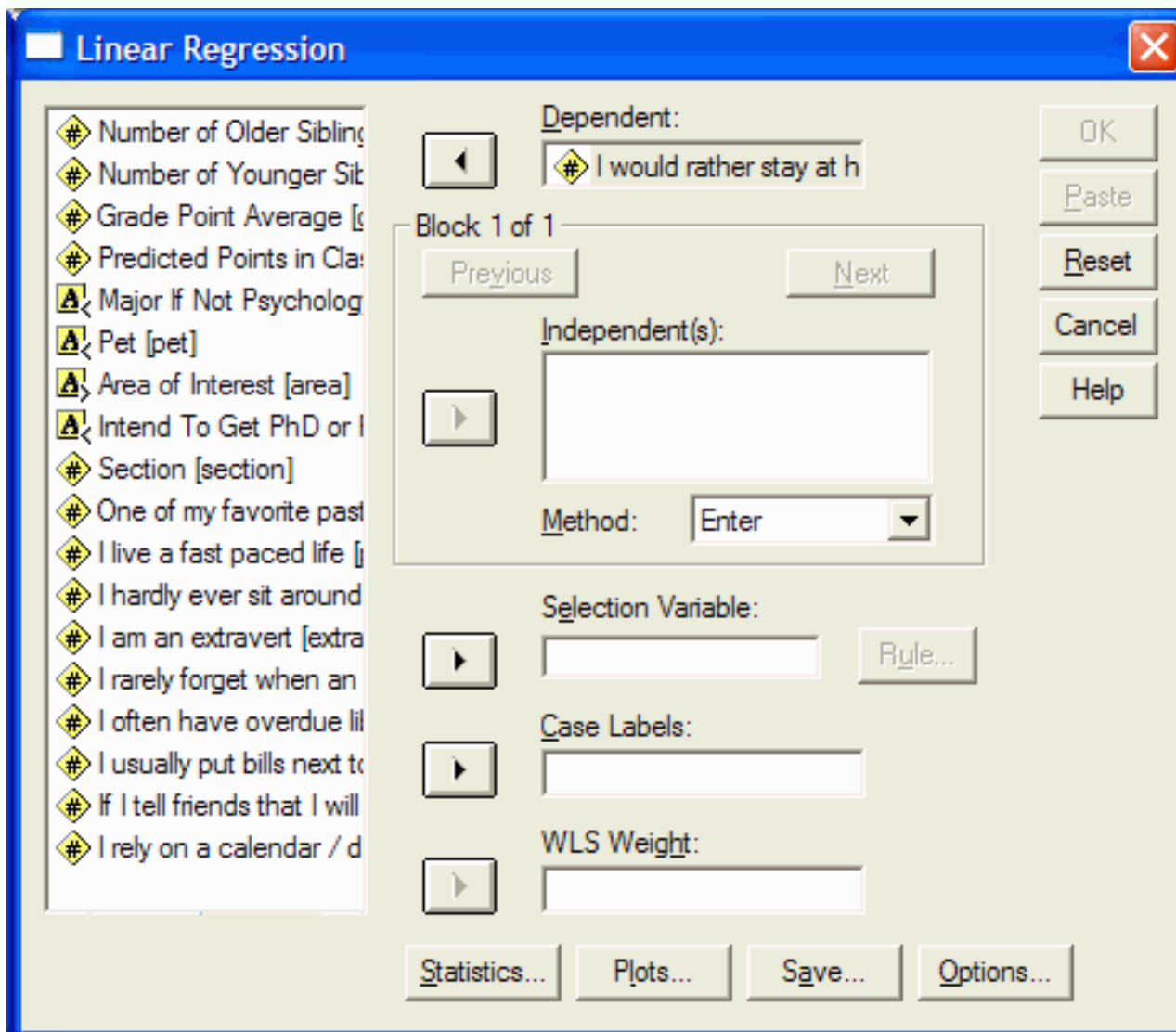
The linear regression command is found at Analyze | Regression | Linear (this is shorthand for clicking on the Analyze menu item at the top of the window, and then clicking on Regression from the drop down menu, and Linear from the pop up menu.):



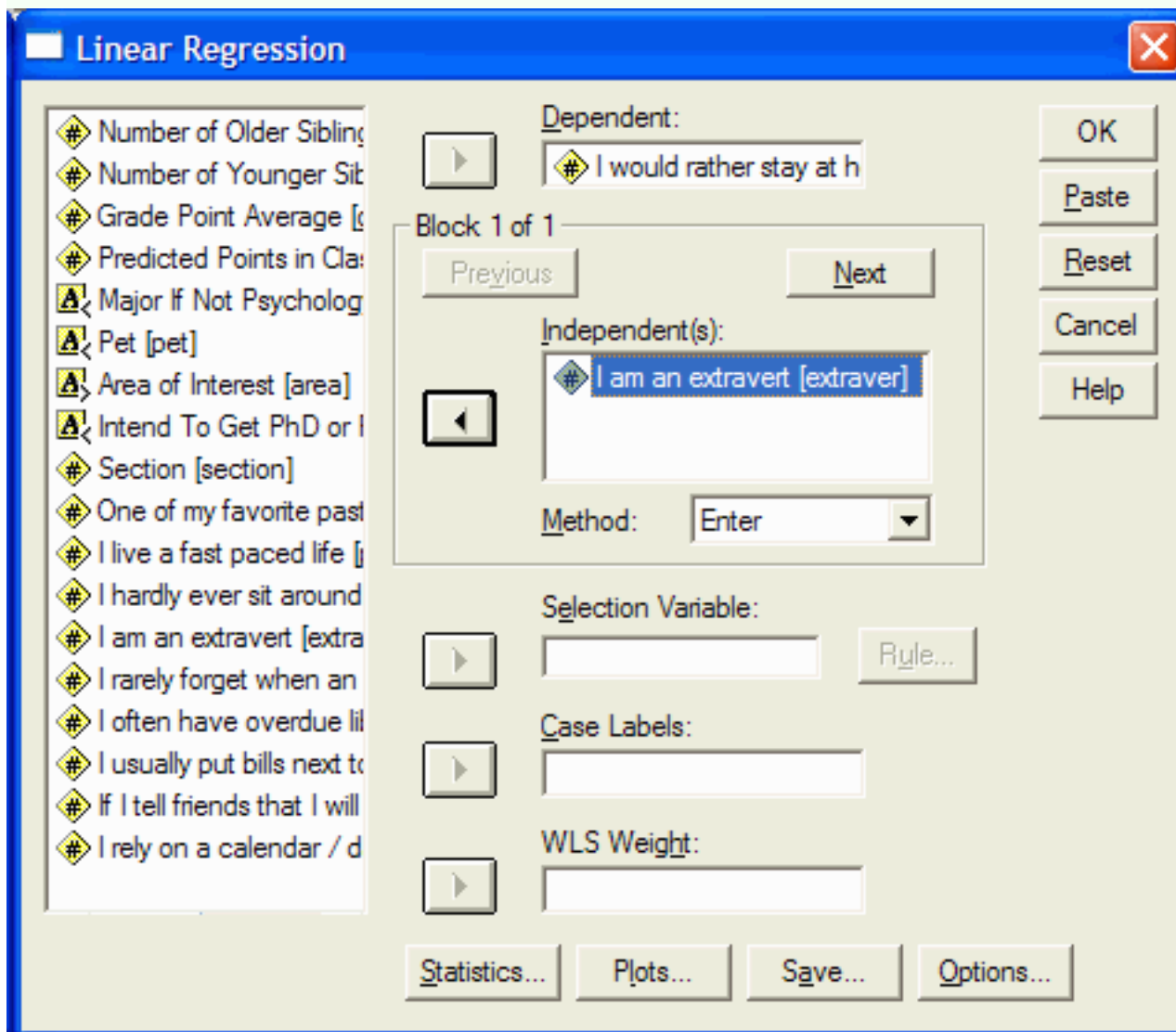
The Linear Regression dialog box will appear:



Select the variable that you want to predict by clicking on it in the left hand pane of the Linear Regression dialog box. Then click on the top arrow button to move the variable into the Dependent box:



Select the **single** variable that you want the prediction based on by clicking on it in the left hand pane of the Linear Regression dialog box. (If you move more than one variable into the Independent box, then you will be performing multiple regression. While this is a very useful statistical procedure, it is usually reserved for graduate classes.) Then click on the arrow button next to the Independent(s) box:



In this example, we are predicting the value of the "I'd rather stay at home than go out with my friends" variable given the value of the extravert variable. You can request SPSS to print descriptive statistics of the independent and dependent variables by clicking on the Statistics button. This will cause the Statistics Dialog box to appear:

**Linear Regression: Statistics**

Regression Coefficients
  Model fit

Estimates
 R squared change

Confidence intervals
  Descriptives

Covariance matrix
  Part and partial correlations

Collinearity diagnostics

Residuals

Durbin-Watson

Casewise diagnostics

Otliers outside:  standard deviations

All cases

Click in the box next to Descriptives to select it. Click on the Continue button. In the Linear Regression dialog box, click on OK to perform the regression. The SPSS Output Viewer will appear with the output:

### 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

The Descriptive Statistics part of the output gives the mean, standard deviation, and observation count (N) for each of the dependent and independent variables. For example, the "I'd rather stay at home than go out with my friends" variable has a mean value of 4.11.

## Correlations

		I would rather stay at home and read than go out with my friends	I am an extravert
Pearson Correlation	I would rather stay at home and read than go out with my friends	1.000	-.310
	I am an extravert	-.310	1.000
Sig. (1-tailed)	I would rather stay at home and read than go out with my friends	.	.018
	I am an extravert	.018	.
N	I would rather stay at home and read than go out with my friends	46	46
	I am an extravert	46	46

The Correlations part of the output shows the correlation coefficients. This output is organized differently than the output from the correlation procedure. The first row gives the correlations between the independent and dependent variables. As before, the correlation between "I'd rather stay at home than go out with my friends" and itself and between extravert and extravert is 1, as it must be. The correlation between "I'd rather stay at home than go out with my friends" and extravert is  $-.310$ , which is the same value as we found from the correlation procedure.

The next row gives the significance of the correlation coefficients. See the discussion in the correlation tutorial to interpret this. As before, it is unlikely that we would observe correlation coefficients this large if there were no linear relation between rather stay at home and extravert.

The last row gives the number of observations for each of the variables, and the number of observations that have values for all the independent and dependent variables.

**Variables Entered/Removed<sup>b</sup>**

Model	Variables Entered	Variables Removed	Method
1	I am an extravert <sup>a</sup>	.	Enter

a. All requested variables entered.

b. Dependent Variable: I would rather stay at home and read than go out with my friends

The Variables Entered/Removed part of the output simply states which independent variables are part of the equation (extravert in this example) and what the dependent variable is ("I'd rather stay at home than go out with my friends" in this example.) Check this to make sure that this is what you want (that is, that you want to predict the "I'd rather stay at home than go out with my friends" score given the extravert score.)

**Model Summary**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.310 <sup>a</sup>	.096	.075	.866

a. Predictors: (Constant), I am an extravert

The Model Summary part of the output is most useful when you are performing multiple regression (which we are NOT doing.) Capital R is the multiple correlation coefficient that tells us how strongly the multiple independent variables are related to the dependent variable. In the simple bivariate case (what we are doing)  $R = |r|$  (multiple correlation equals the absolute value of the bivariate correlation.) R square is useful as it gives us the coefficient of determination.

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.496	1	3.496	4.667	.036 <sup>a</sup>
	Residual	32.961	44	.749		
	Total	36.457	45			

a. Predictors: (Constant), I am an extravert

b. Dependent Variable: I would rather stay at home and read than go out with my friends

The ANOVA part of the output is not very useful for our purposes. It basically tells us whether the regression equation is explaining a statistically significant portion of the variability in the



dependent variable from variability in the independent variables.

Coefficients<sup>a</sup>

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.808	.348		13.820	.000
	I am an extravert	-.277	.128	-.310	-2.160	.036

a. Dependent Variable: I would rather stay at home and read than go out with my friends

The Coefficients part of the output gives us the values that we need in order to write the regression equation. The regression equation will take the form:

Predicted variable (dependent variable) = slope \* independent variable + intercept

The slope is how steep the line regression line is. A slope of 0 is a horizontal line, a slope of 1 is a diagonal line from the lower left to the upper right, and a vertical line has an infinite slope. The intercept is where the regression line strikes the Y axis when the independent variable has a value of 0.

The predicted variable is the dependent variable given under the boxed table. In this case it is "I'd rather stay at home than go out with my friends.". The slope is found at the intersection of the line labeled with the independent variable (in this case extravert) and the column labeled B. In this example, the slope equals -0.277. The independent variable was extravert (we specified that when we set up the regression.) The intercept is found at the intersection of the line labeled (Constant) and the column labeled B. In this example, the intercept is 4.808. Putting it all together, the regression equation is:

Predicted value of "I'd rather stay at home than go out with my friends" = -0.277 X value of extravert + 4.808

That is, if a person has a extravert score of 2, we would estimate that their "I'd rather stay at home than go out with my friends" score would be  $-0.277 \times 2 + 4.808 = 4.254$ . Thus, we would predict that a person who agrees with the statement that they are extraverted (2 on the extravert question) would probably disagree with the statement that they would rather stay at home and read than go out with their friends (4 [~4.254] on the "I would rather stay at home..." question.) Given the small value of  $r$ , our prediction will, in general, not be very accurate.