

This step-by-step example shows how to enter data into SPSS and conduct a simple regression analysis to develop an equation to predict Y from X. Our data are as follows:

ID	X	Y
1	3	8
2	5	10
3	6	4
4	7	16
5	9	12

Open SPSS (double-click the SPSS icon in Windows, select Type in data, OK) to display the SPSS data editor.

We begin by defining our variables. Click the Variable View tab at the bottom of the Data Editor window. Under Name, enter **ID** in the first row, **X** in the second row, and **Y** in the third row. For all three rows, under Decimals enter **0**. Variable labels and value labels could be entered, but in this example they are not needed.

Go to the Data View window by clicking the Data View tab at the bottom of the Data Editor window. Enter your raw data into the data editor window of SPSS. In general, place all data from any case in the same row, with each variable in a separate column. It is generally useful to include an ID code for each case to facilitate tracking cases, especially in large data sets. There are five rows of data, where the first column has the ID numbers 1, 2, 3, 4, 5; the second column has the X values of 3, 5, 6, 7, 9; and the third column has the Y values of 8, 10, 4, 16, 12.

Now we are ready to set up our regression analysis.

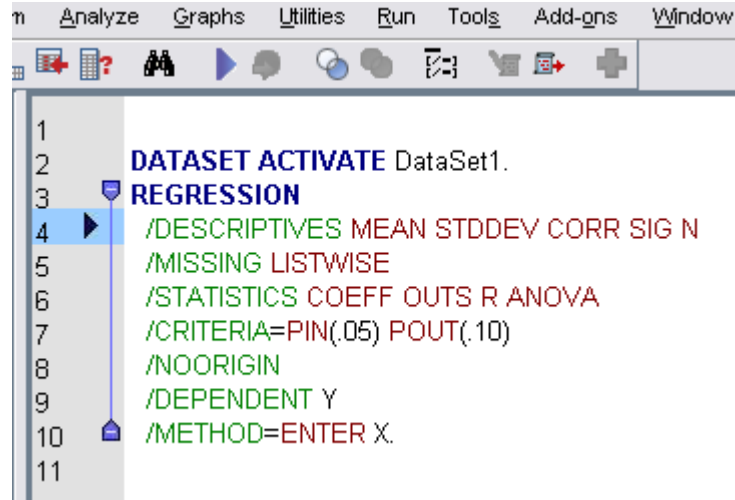
Click Analyze, Regression, Linear... Select Y, click the curved arrow to move **Y** into Dependent. Select X and click the curved arrow to move **X** into Independent(s). Click Statistics..., select Estimates, Model fit, Descriptives, Continue. Click Paste to display the syntax you have generated in the syntax window.

It is very useful to keep a copy of your syntax file, especially when you are working with multiple analyses or more complex analyses. Syntax provides a record of exactly what analyses were done. You can save the syntax file, edit it, and rerun it.

To locate the syntax file, click Window and select the syntax file.

SPSS provides many options regarding what you view when you open the syntax file. I recommend that you ask SPSS to display the Syntax Editor bar of icons. To display this bar, click View, Toolbars, and check Syntax Editor. This will open a bar that includes icons for many useful options, including a triangle that can be used to run the current command. You can also run the syntax file by clicking Run, All (or a subset).

Run the syntax file (move the cursor anywhere inside the REGRESSION command before the final period and click the triangle in the bar above, or click Run, All.



Below is selected output from the SPSS Statistics Viewer window.

Check descriptive statistics carefully to make sure the data went into the computer as you intended.

Regression

Descriptive Statistics

	Mean	Std. Deviation	N
y	10.00	4.472	5
x	6.00	2.236	5

An important first check is to verify that we have the right number of cases and that the means are as expected.

Correlations

		Y	X
Pearson Correlation	Y	1.000	.450
	X	.450	1.000
Sig. (1-tailed)	Y	.	.224
	X	.224	.
N	Y	5	5
	X	5	5

Eyeball calibration:

Note that the adjusted R square is negative!

This indicates that the observed $r = .450$ is a bit closer to zero than we would expect by chance with a sample of $n = 5$.

See below that t and F are both less than 1.00. If H_0 is true, the expected value for $F = 1.00$.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.450 ^a	.203	-.063	4.612

a. Predictors: (Constant), x

b. Dependent Variable: y

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16.200	1	16.200	.762	.447 ^a
	Residual	63.800	3	21.267		
	Total	80.000	4			

$$SS_{tot} = SS_{reg} + SS_{err}$$

$$80.00 = 16.2 + 63.8$$

- a. Predictors: (Constant), x
 b. Dependent Variable: y

The observed R of .450 is not statistically significant – we do not have convincing evidence that the population correlation is greater than zero.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.600	6.522		.705	.531
	X	.900	1.031	.450	.873	.447

- a. Dependent Variable: Y

The Unstandardized Coefficients are the B weights for the regression equation.

$$\hat{Y}_i = B_0 + B_1X_i ; \quad \hat{Y}_i = 4.600 + .900X_i$$

We could use this equation to generate a predicted value for Y for any value of X, including values of X that we haven't observed.

ID	X	Y	\hat{Y}
1	3	8	7.3
2	5	10	9.1
3	6	4	10.0
4	7	16	10.9
5	9	12	12.7

Practice Questions:

1. What is the predicted value for Y when X is at the mean, 6?
2. What is the predicted value for Y when X is zero?
3. What is the change in the predicted value of Y when X is increased one unit, say from 6 to 7?

Answers:

1. 10; When X is the mean for X, the predicted value for Y is the mean on Y.
2. 4.6; When X= 0, the predicted value for Y is the Constant.
3. .9; When X is increased by one unit, the predicted value of Y is increased by the coefficient for X, also called the 'slope' or the B weight. Here B = .9.

Scatterplot with SPSS

In general, it is important to examine plots of your data to assess how well the data satisfy the mathematical assumptions of statistical models. For regression analyses, we might begin with an examination of the univariate plots, but we also should examine the bivariate plots and residuals from our model. Are there errors? Extreme outliers? Departures from linearity? Heteroscedasticity?

To generate a Scatterplot with SPSS, click Graphs, Chart Builder to open the window. Under Gallery, select Scatter/Dot, drag the upper left box (Simple Scatter) into the Chart Builder window. Drag X from the Variables box into the horizontal axis and Y into the vertical axis. You can click OK to run the analysis or Paste to add new syntax to the syntax window.

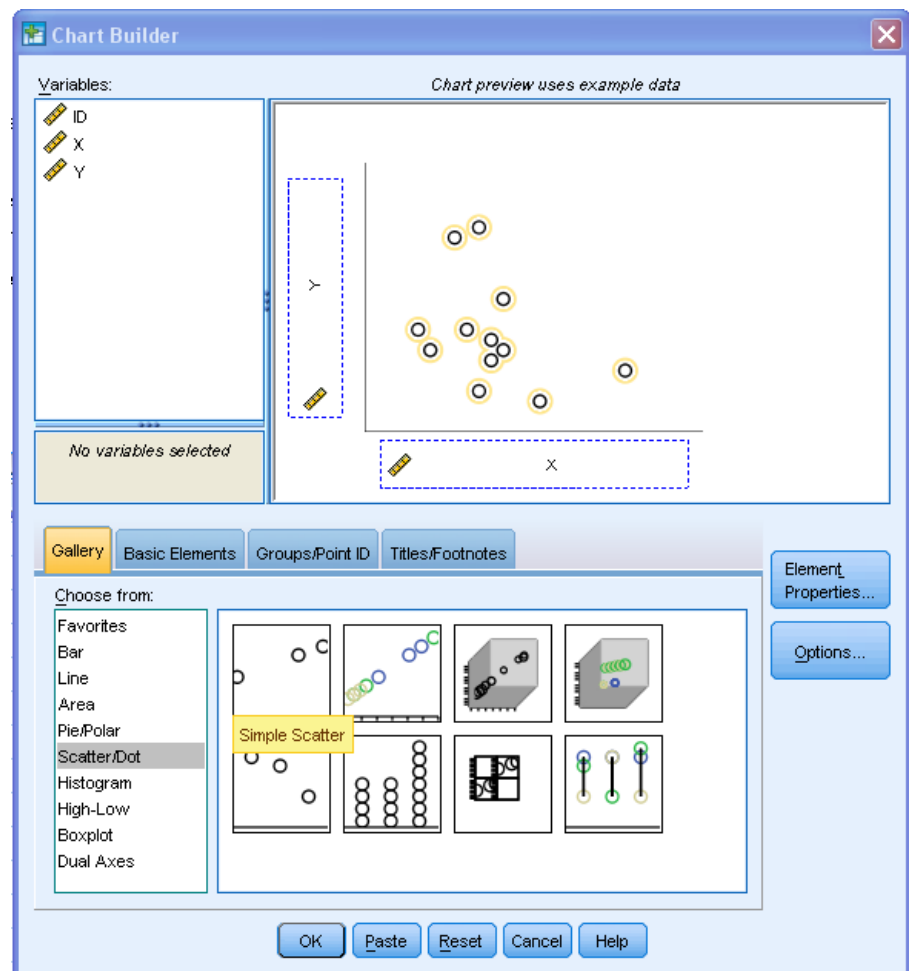
If you save the syntax you will see that your pointing and clicking generated two separate sets of commands, **GGRAPH** and **BEGIN GPL**. To run the analysis you will need to highlight both sets of commands and click the triangle (or click Run, Selection).

SPSS produces the GGRAPH shown on the next page. This is a good example of a case where you will most certainly need to change the SPSS defaults. Take a look at the graph and see if you notice any problems.

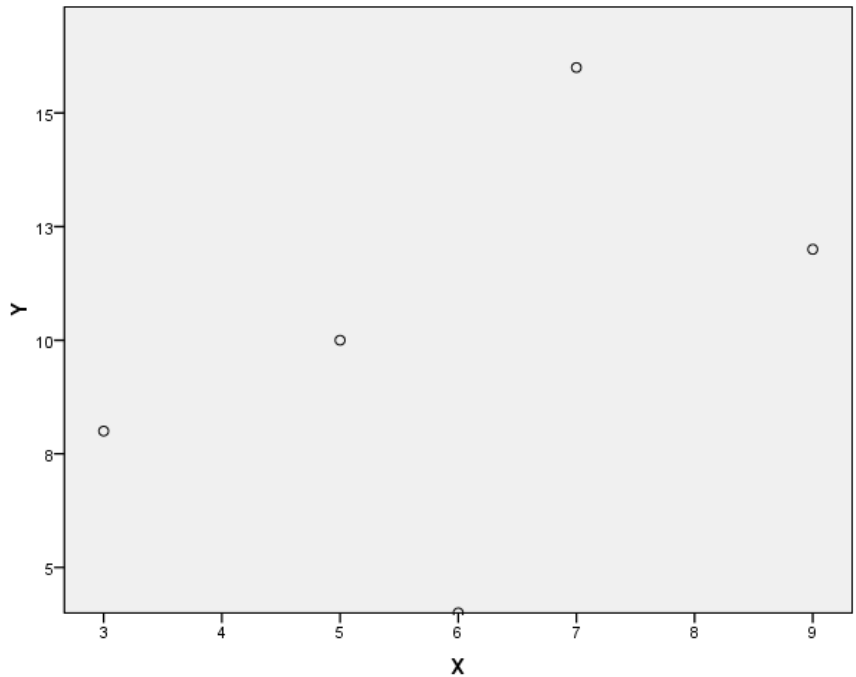
Hint: Check the location of the point where $X=3$, $Y=8$. (Whoa! Is SPSS that inaccurate?)

Then check the location of the point where $X=6$, $Y=4$ (The point is struggling to be seen).

Now, check the intervals between several values on the Y axis. Some intervals are 2 units and others are 3 units, but they are shown to be equal! (What is going on here???)



Some editing clearly is needed for this graph. The Y scale certainly needs our attention. SPSS has rounded off to whole numbers, but it scaled the intervals to be 2.5 points wide. Thus, the 8 is really 7.5. The lowest observed value on Y is 4. We could begin the scale at some value less than 4 so the lowest point will show more clearly. We also may wish to include a regression line.



Double-click on the scatterplot in the Viewer window to open a Chart Editor window. Click Elements, Fit Line at Total, select Linear, Close.

Double-click on the numbers on the Y axis to open a Properties window. Click Scale, under Custom, enter 2 as the Minimum and 2 as the Major Increment. Click Close.

This editing will produce the graph shown below. There are many additional options – explore them to see what they do.

Lesson: Don't trust SPSS defaults. You must take control to make sure that information is presented accurately and clearly.

