# LINEAR REGRESSION WITH MICROSOFT EXCEL

KELLIN RUMSEY

Before you begin, you will want to include the Microsoft Excel Add-In *Analysis ToolPack*. To do this, open up Excel and follow these easy steps.

(1) Select the *Options* tab under File.
(2) Choose *Add-Ins*.
(3) Select *Analysis ToolPack* (For our purposes, do NOT choose the VBA option).
(4) Press the *Go* button.

Once you have the *Analysis ToolPack* installed, you should be able to find a *Data Analysis* option underneath the Data tab.

To begin, you should include your data in the first two columns of your Excel Worksheet (X variable first). You can include headers if you like, but they are not necesarry. Here is an example, using 20 highly populated counties in the United States. I will try to fit a Regression Model for predicting % of the population below poverty level (Response) vs. % of the population with a Highschool Diploma (Explanatory).

FIGURE 1
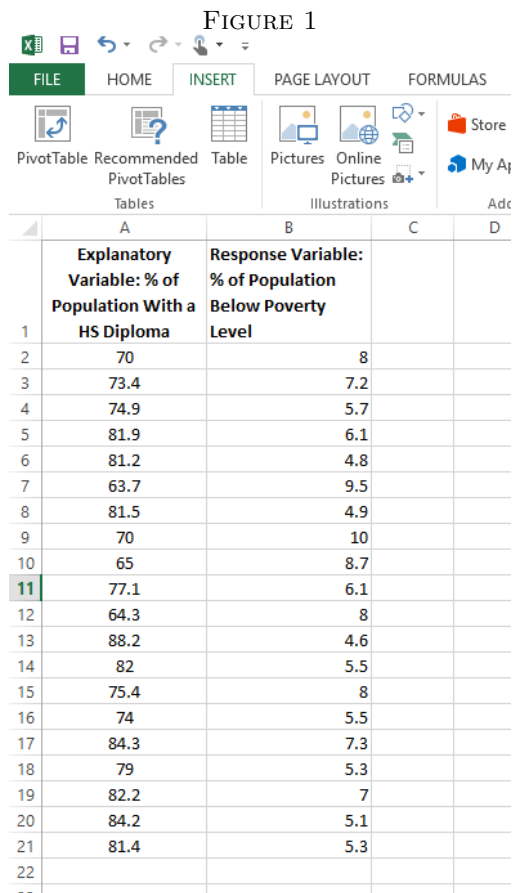
Once you have the data imported, it is a good idea to create a Scatterplot. Choose a Scatterplot from the *Insert* tab like so.
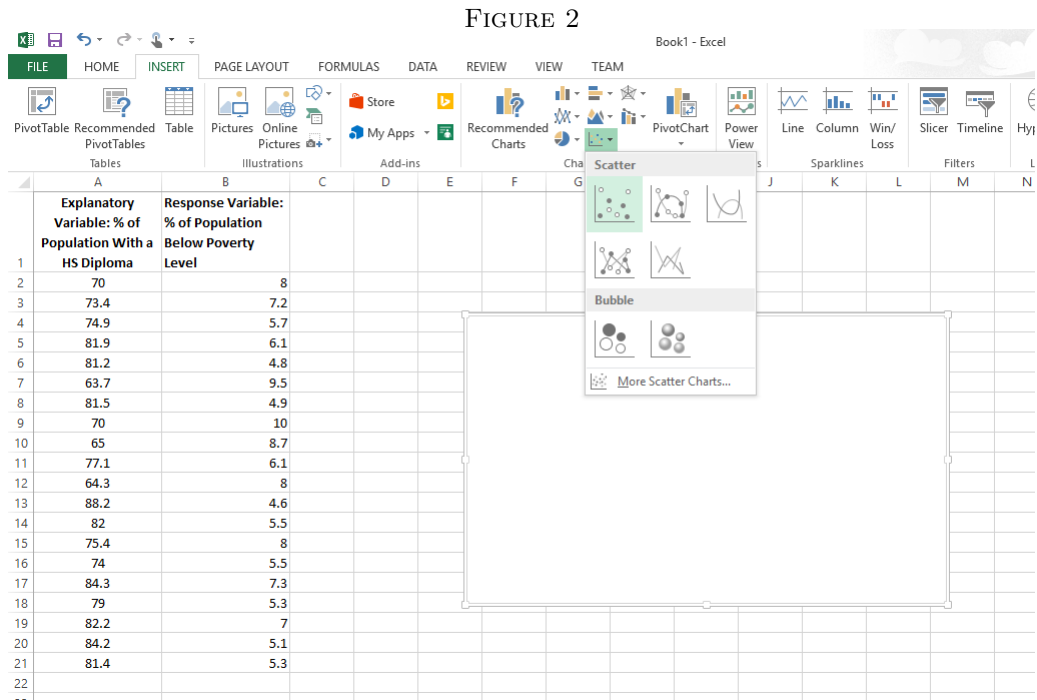
FIGURE 2

Right click the empty Scatterplot, and choose *Select Data.* From here, you should select your data as depicted in the following figure.

FIGURE 3
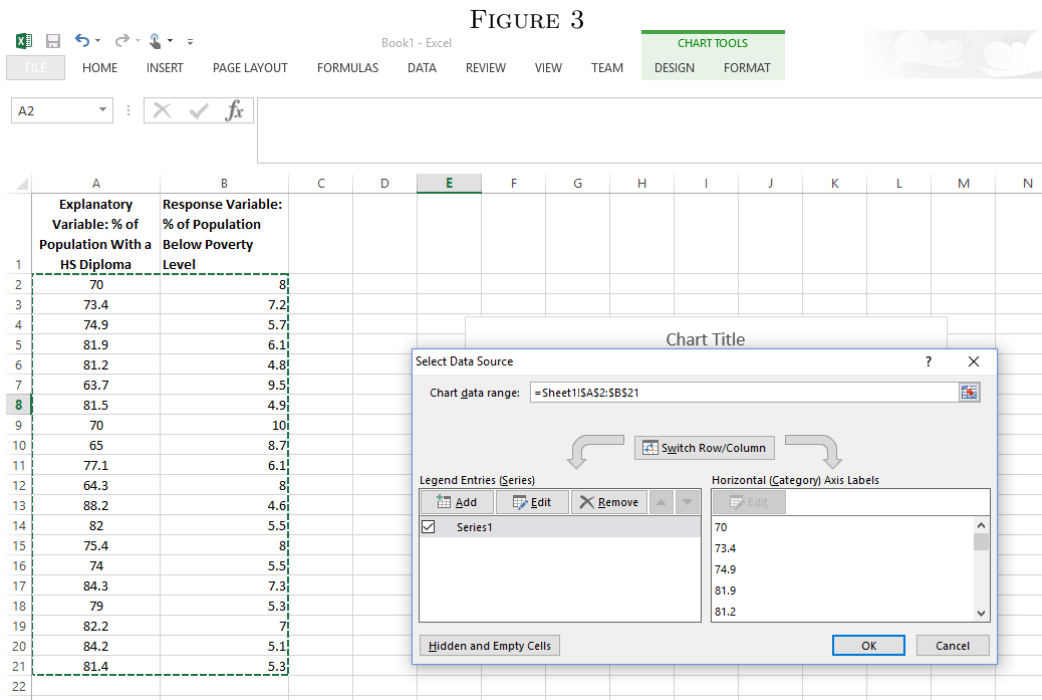
When you press okay, the Scatterplot should be created. Now you can mess around with your plot until you are happy.

You can do things like:

(1) Give your plot a title.
(2) Click chart elements (the green plus sign), to add axis titles.
(3) Right click on an Axis to choose specific values.
(4) Right click on the points to change their color.

From here, we could skip directly to the Regression Line, but it is good to remember how we calculate the Regression Line. We should start by finding $\bar{x}, S_x, \bar{y}, S_y$ and r. Microsoft Excel has built in commands for mean (average(...)), standard deviation (stdev.s(...)), and correlation (correl(...,...)). The following figures illustrate how to use them; don't forget to start with an '=' sign!
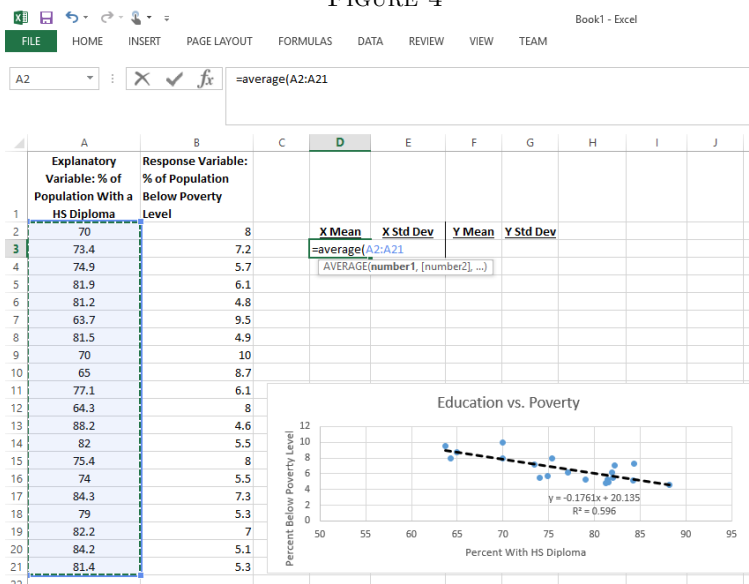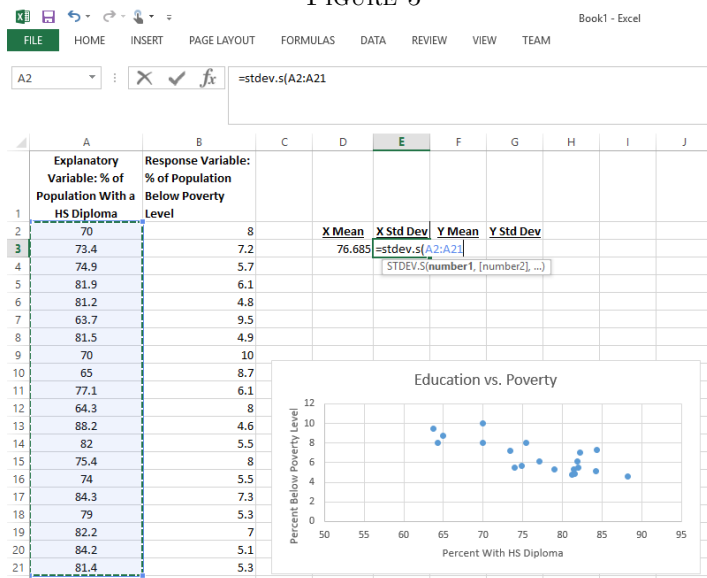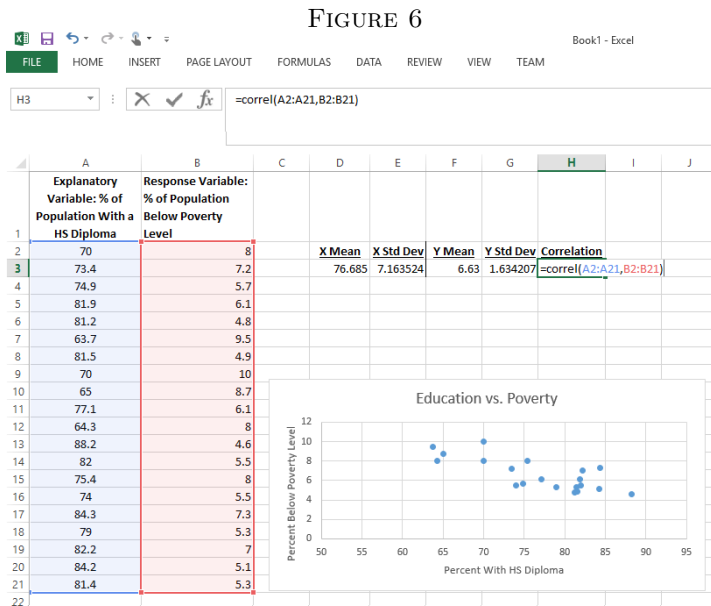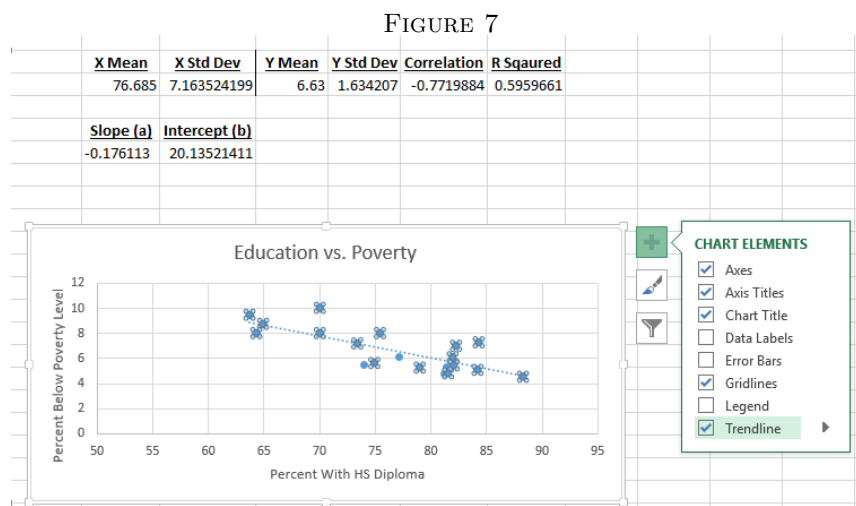
FIGURE 4

FIGURE 5

FIGURE 6

Now that we have calculated the means, standard deviations and the correlation, we can compute the equation for the regression line using the following equations.

$$\hat{y} = a + bx$$

$$b = r\frac{S_y}{S_y}$$

$$a = \bar{y} - b\bar{x}$$

Microsoft Excel also offers a built in Regression Line feature when doing a Scatterplot. If we have done everything correctly, then the line calculated by Excel should have the same formula as the one we just calculated! Click on the scatterplot, and select *Chart Elements* (the green plus sign). Simply select *Trendline.*

FIGURE 7

If you right click on the trendline, you can play around with some options. Make sure to select *Linear.* You can also opt to have the program display the R^2 value and the equation for the Regression Line on the plot.

FIGURE 8
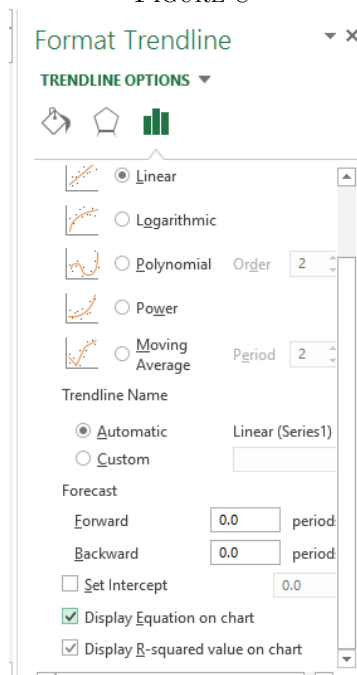
Check to see if the equation matches the one you calculated. Finally, you are left with a plot and a Regression Model to be proud of!

FIGURE 9