Lab 5:
OLS Live and Regression

Objectives:

1. Review OLS. What does OLS mean? What does it accomplish? What do we choose?
2. Create a spreadsheet including the data manipulations needed to estimate a regression line.
3. Given sample data, use Excel to estimate a regression line.

Key Terms:

1. ***Correlation***.
2. ***Regression, fitting a line, estimating*** $\hat{\beta}_0$ ***and*** $\hat{\beta}_1$**.**

Data: *Lab 5 – OLS Live and Regression.xls*.

Exercises:

- ***Regression Lines – OLS estimation.***

1. Open the worksheet titled: ***OLS – Live***; either from the ResEc 312 folder or from the course website. You will observe a table with 5 columns including: values for *Y* (weekly expenditures in \$), corresponding values for *X* (weekly income - \$), "fitted values" ($\hat{Y}_i$), errors (e_i), and errors squared (e_i^2). You will also observe a second table below the first and two graphs. The first graph, titled **OLS: Finding the line of "best fit"** shows a scatter diagram of the *X,Y* values. The second table has a few possible slope values in it. The values for the sum of the squared-errors are blank; you'll need to generate those values. Once you fill out that table, the relationship between the sum of the squared-errors and choices for the slope will appear. **Don't add any graphs or charts to this page. Just use the existing templates**.
2. To show how the sum of the squared-errors changes as you change the slope do the following:
   a) Enter a possible slope value for $\hat{\beta}_1$ in cell B17. (I've already done the possible value of 0.90 and copied the sum of the squared-errors value to the first cell in the second table.) Try additional values between 0.9 and 0.3. When you enter the value 0.80 in cell B17, Excel calculates new values in Table 1. You'll see changes in the fitted values, the errors and the squared-errors. Copy the new value that appears in cell F13 for the **sum of the squared-errors**. When pasting the values into the second table, you only want to paste the number, not the formula. To do this, always use **Paste Special** and **Values**.
   b) You should observe two things: (1) Changing the slope changes the line through the data, changes all the errors and changes the sum of the squared-errors. (2) Excel is plotting the **sum of the squared-errors** versus the different values of the slope in the second graph. With enough values for the slope, you should be able to create a "U-shaped curve" in that graph. What happened when you decreased the slope from 0.90 to 0.80? What happened to the red line in the X,Y scatter diagram? What happens to the errors in the table? What happens to the sum of the squared-errors?
   c) Complete the second table by determining sums of squared-errors for values between 0.9 and 0.3. Which value appears to give the line of best fit?
   d) You can play around with the values and try to find the absolutely 'best' slope. If you highlight two cells in this second table (the cells for a slope of 0.50 and the sum of squared-errors in the next column), then "right click" and choose "insert," Excel will allow you to shift the cells down and insert one extra row in the table. You can try additional values between the two values you think are best using this new cell. Experiment with different values until you find **the best slope**.

e) Now, think about what is happening here. You changed the slope, holding the intercept constant. When you did that, the ***errors*** changed, the ***squared-errors*** changed and the ***sum of the squared-errors*** changed. What was your "objective" as you changed the values for $\hat{\beta}_1$? Explain in a textbox on the worksheet. What happened to the ***sum of the errors***?

f) Find your best slope, then for that value of $\hat{\beta}_1$, change the value for $\hat{\beta}_0$ to, say 28.50. What happens to the sum of the squared-errors for this new value of $\hat{\beta}_0$?

♦ ***Regression in Excel..***

1. The worksheet **OLS Estimation** has the same data. Rather than searching for the best slope and intercept values, let Ordinary Least Squares (OLS) regression do it for you.
2. Use Excel's **Data Analysis ToolPak** to calculate the correlation coefficient for X and Y. How would you describe the relationship between X (income) and Y (expenditures) given the correlation coefficient?
3. Create a ***scatter diagram***. Does the diagram look like what you would expect given your estimate of correlation? Add a **trendline** to the scatter diagram and have Excel include the equation for the line. This is the OLS regression line.
4. Assume that expenditures and income are related according to the population regression equation:

$$Y = \beta_0 + \beta_1 X + u .$$

5. Use Excel's **Data Analysis**, **Regression**, to estimate this regression model. Excel asks you to input the Y range (the dependent variable) and the X range (the independent variable). Include the labels. ***Place the results in a separate worksheet titled OLS Regression***.
6. Compare your OLS regression results to your OLS Live worksheet. Can you find the slope estimate? The intercept estimate? How about the sum of the squared-errors? Your minimum sum of the squared-errors should match closely with the value that OLS provides in it's ANOVA Table. There are also two boxes shaded on the sheet labeled "beta-1 hat" and "beta-0 hat." Enter the following Excel functions in those cells: "=intercept(b2:b11,c2:c11)" and "=slope(b2:b11,c2:c11)". These two functions also give you the OLS estimates for the intercept and slope.

♦ Honda Accord Sample Regression Function:

1. You recall our population of used 2004 Honda Accords. We said that one reason for variation in prices may be that different cars have different mileage. We saw that there was a inverse relationship between price and mileage when we created a scatter diagram for the population data.
2. Estimate that population relationship by drawing a sample of n=30 used Honda Accords from the population. OWL has just such a sample ready for you. Go to the course website, then to **Personalized PRS Data.** Enter your 8-digit ID and a sample of n=30 will be provided. Copy, and paste that sample into today's spreadsheet. The data are separated by commas and will come in as a single column of data. Use Excel's **Text to Columns** wizard (on the Data ribbon) to parse the data into separate columns (variables.)
3. Use **Data Analysis**, **Regression** to estimate the relationship between price and miles:

$$Price_i = \beta_0 + \beta_1 Miles_i + u_i .$$

4. What is your estimate for the population parameter β_1? **Interpret the estimate**.
5. What is your estimate for the population parameter β_0? **Interpret the estimate**.