Lab 10: Data Manipulations and Non-Linear Models

Objectives:

It is not unusual to have to make some modifications to your data before estimation. In the first part of the lab, we'll use ***Minitab*** to do some data manipulations. The second part of the lab will introduce you to estimation of nonlinear models by again manipulating the data in ***Minitab*** and then estimating using *OLS*. There is a practice problem set posted on the website and attached below as well. Bring the practice problem to the lab; you should be able to complete these problems. We're not collecting all these answers, but you should be able to write answers to these problems following lab and lectures.

Data: ***CPS1985-NonLin.mtw*** and ***Lab 8 Exercises.xls***. or your Lab 9 minitab project.

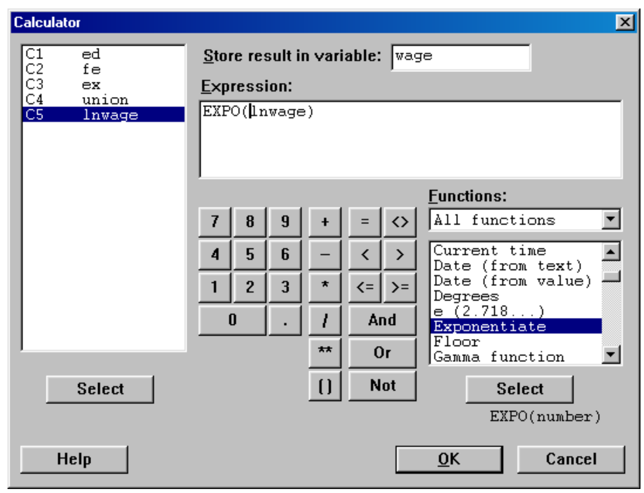
- ***More Dummy Variables and Nonlinear Models***

1. The data set `CPS1985-NonLin.MTW` is a ***Minitab*** Worksheet. Choose **File** and **Open Worksheet** in ***Minitab***. These data are a small sample from the 1985 Current Population Survey of the U.S. The data set contains a number of variables:

   lnwage = the natural logarithm of the individual's 1985 wage;
   ed = the individual's education in years;
   ex = the potential experience for an individual in years;

   as well as two dummy variables, ***fe*** indicates whether or not the individual was female and ***union*** indicates whether or not the individual was a member of a union.

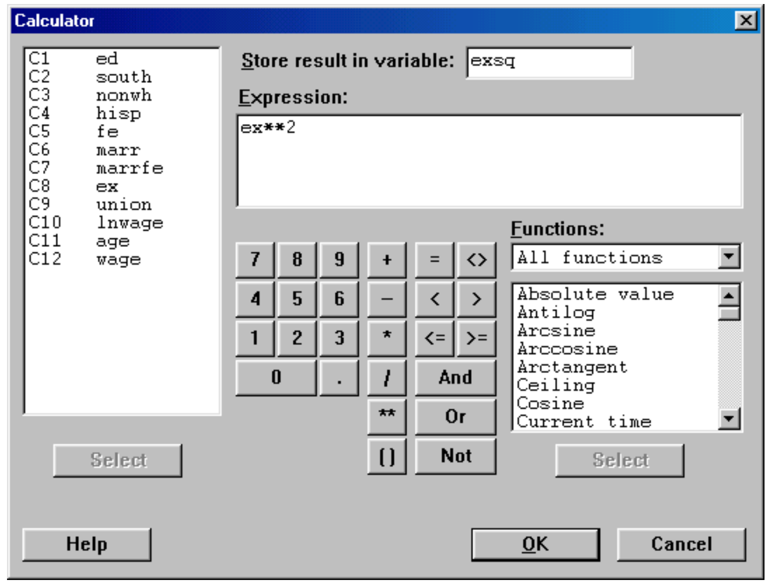
2. **Specify and estimate the two models** that are required to answer the question: ***Do women earn lower starting salaries and lower raises***? This question is asked in part *c* of the practice problem that was posted with this lab. It requires that you estimate two models and then use of an F-test to choose between the two models; you will determine whether women earn lower wages *and* lower raises than equally qualified males. The first model is given in the text for the problem set. It relates wages to two independent variables, education and experience. (Define a *raise* as the returns to an additional year of experience.) First, you must note that we do not have the individual's wage, but rather the natural log of wage. To solve that problem, create the variable ***wage*** by taking the exponent of ***lnwage***. Use the **Calculator** to create a new variable; the expression is ***expo(lnwage)***. (See the screen capture at the right.)



3. Specify and estimate the **second model** needed to answer the question posed in the practice problem, part *c*. This model requires that you include a dummy variable (*fe*) and an interaction term (*feex*). How do you create the interaction term that is required? You should have no problems completing this given our work in lab last week. ***Copy the results for these two regressions into* Word *and write a paragraph interpreting the estimated coefficients for the variables fe and feex. Are these individual effects statistically different from zero? What hypothesis test would you specify to complete the test?***

4. Next, using the ANOVA tables from these two regressions, complete the F-test to determine whether women earn lower wages *and* lower raises than equally qualified males. Use Excel or your calculator to calculate the F-statistic needed to complete the test. (Hint: you'll need to draw from the two ANOVA tables in the two regression printouts.) Follow the steps we discussed in class to calculate the test statistics. Then draw a picture of the test (draw an F-distribution) and complete the test. What do you conclude?

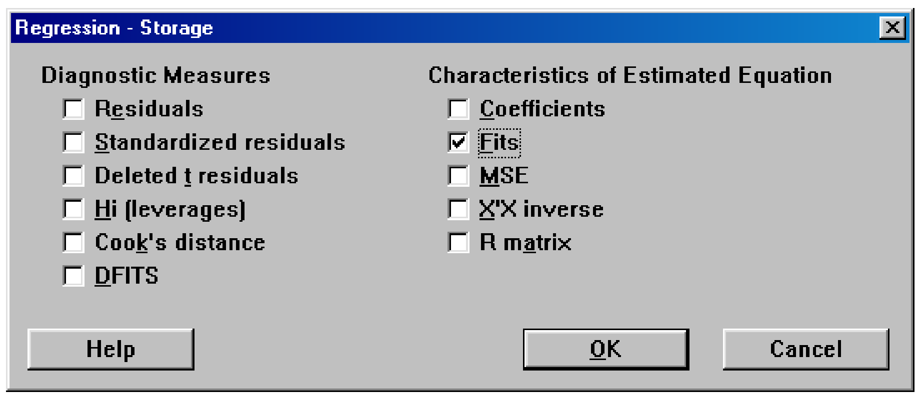
5. In the practice problem a non-linear model is also considered. A quadratic model allows the returns to experience to vary. To estimate a quadratic model, we need to create the variable, $exsq = ex^2$. To do this in ***Minitab***, we select **Calc** and then **Calculator.** Tell ***Minitab*** that you are creating the new variable ***exsq*** and then type the equation as shown in the ***Minitab*** screen capture at the right. The two asterisks (**) in the equation indicate that the variable ***ex*** is being raised to the second power (we *square* it). You could also just multiply ***ex*** by itself.



6. Estimate the regression model with both ***ex*** and the new variable ***exsq*** included (include the dummy variable ***fe*** and the interaction term ***feex*** as well):

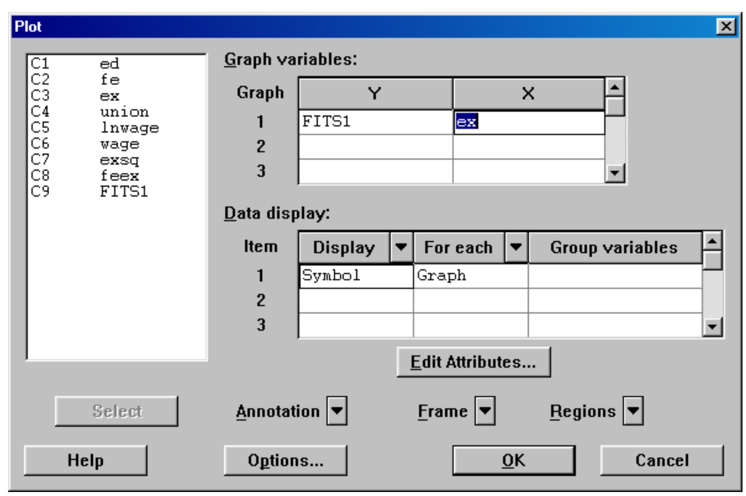
$$wage_i = \beta_0 + \beta_1 Ed_i + \beta_2 Ex_i + \beta_3 Exsq_i + \delta\, fe_i + \gamma\, feex_i + u_i$$

When you estimate using ***Minitab***, choose **Storage** and then check **Fits** as shown at the right. This will have ***Minitab*** save the fitted values as a variable (***FITS1)*** in the worksheet. The fitted values are predicted wages using the values for the independent variables for each observation (person).



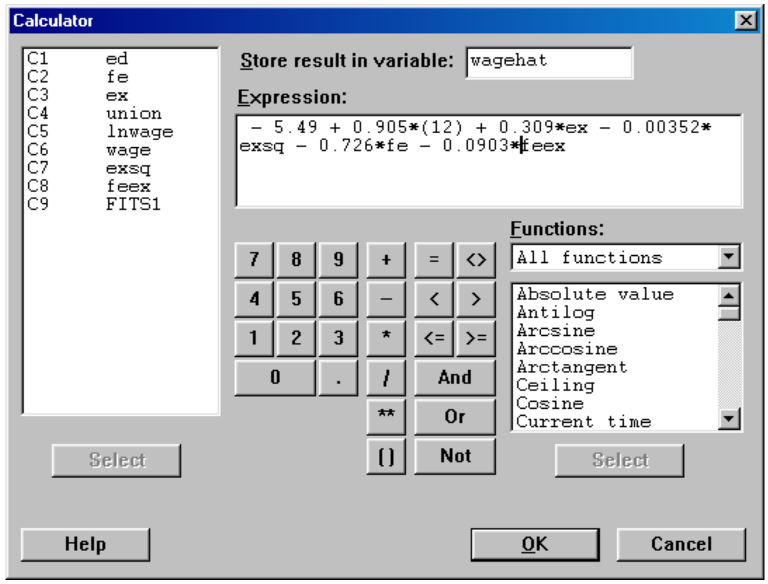
You can edit the names or column heading in ***Minitab***. Just click on the title (***FITS1***) and change it to ***Wagehat1***, or whatever you wish (it should make sense).

7. Select **Graph** from the main menu and choose **Plot**. You will see the screen below; just select ***FITS1*** as the Y variable (or whatever name you chose) and plot the estimated or fitted values versus the variable ***ex***.



Take a look at the plot. Pretty odd; what went wrong? Why don't we see nice lines like Dan shows in class? What the plot shows is the relationship between experience and wage, ***but education was not being held constant*** in this two-dimensional graph making it difficult to see the true relationship between ***ex*** and ***wage***.

8. To solve the problem, copy the SRF into **Calc** / **Calculator** to calculate your own predicted values. (See the screen capture at the right.) Select a new variable name (like ***wagehat***), and hold education constant by assigning a single value to all individuals. I've chosen to hold education constant at ***ed*** = 12 for all observations and I've left ***ex*** free to vary. After creating this new variable, create a new graph by plotting the new predicted values you created holding education constant, ***wagehat***, versus ***ex***. Now we can see the effect of ***ex*** on ***wagehat***. Looking at this graph, you can see that the partial effect of experience on wages changes as experience changes. You can see the same thing by taking the partial derivative of this equation and evaluating it at different levels of experience.



9. You should be able to click on the graph, and choose **Copy graph**, and the paste the graph into your **Word** document.

10. **Your Word file should now have three regressions (the two required for the F-test and the non-linear regression) and one graph showing the estimated non-linear relationship between wages and experience.**

11. One last analysis of these data. Recall that we assume the dependent variable for our population regression equation is normally distributed. Check the ***wage*** variable using Minitab's Histogram tool. What shape do you observe? Can these wages be assumed normally distributed? Next, check the distribution for the variable ***lnwage.*** What do you conclude? We often find that a natural log transformation can act to normalize data that are skewed right or left.

♦ ***Log Transformations: Nonlinear Log-Log Model***

1. Log transformations can be used to estimate the following non-linear model: $Y = A\, X_1^{\beta_1} X_2^{\beta_2} \exp^{u}$. This puppy is clearly a nonlinear combination of parameters and variables – a real problem. But, if we take the natural logarithm of this model, we get:

$$\ln Y = \ln A + \beta_1 \ln X_1 + \beta_2 \ln X_2 + u\text{; or } \ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + u$$

Now the model looks "linear in the parameters" so we can estimate using OLS.

2. The parameters of this model are ***elasticities***. For this expression:

$$\frac{\partial \ln Y}{\partial \ln X_1} = \beta_1 = \frac{\partial Y / Y}{\partial X_1 / X_1} = \frac{\text{\% change in Y}}{\text{\% change in X}_1}$$

So, when we estimate any model where both the dependent and independent variables are transformed into natural logs, we interpret the estimated parameter as an elasticity.

- ***Data Manipulations – the Log-Log nonlinear model for Roses Demand.***

1. Open your ***Minitab*** Roses project from lab 8 and/or 9. You should recall that we used the linear model estimates to calculate elasticities. In the linear model, the slope is constant but the elasticities change because the values for Y and X change. We calculated elasticities at the sample means, and for each observation.

2. Suppose we believe that the true specification for the demand for roses is as follows:

   (**1**) $$Sales_t = \exp^{\beta_0}\ Prose^{\beta_1} Pcarn^{\beta_2} Disinc^{\beta_3}\ \exp^{u}$$

3. **Transform** the dependent variable (***Sales***) and the three independent variables (***Prose***, ***Pcarn***, ***Disinc***) by taking the natural logarithms of each. The ***Minitab*** function that we use to take the natural logarithm is "loge" and is found in the **Calc** menu.

4. We're going to need **descriptive statistics** of the variable ***lnSales***, ***lnProse***, ***lnPcarn*** and ***lnDisinc***.

5. Estimate the following model:

   (**2**) $$ln\ Sales_t\ =\ \beta_0 +\ \beta_1\ ln\ Prose\ +\ \ \beta_2\ ln\ Pcarn\ +\ \beta_3\ ln\ Disinc\ +\ u$$

6. Review the regression results. How do you interpret the estimated parameters? Do they seem reasonable in sign and magnitude? Is the demand for roses price elastic or price inelastic? Is the demand for roses income elastic or income inelastic?

7. Calculate fitted values for the dependent variable ***lnSales***. Use the estimated Sample Regression Function to predicted the values for ***lnSales***, holding ***lnPcarn*** and ***lnDisinc*** constant at their sample means. Once you have these predicted or fitted values, transform them back into natural units (pounds) by using the ***Minitab*** transformation "EXPO." You should now have predicted values for Sales, in pounds, for each and every observation. If you already have a ***Saleshat*** variable in your Minitab project, just use ***Saleshat2***.

8. Plot these values versus the one variable we did not hold constant, Prose. You should see a nonlinear relationship between ***Saleshat*** and ***Prose***. While the relationship between ***lnSales*** and ***lnProse*** is linear, the relationship between the variable ***Saleshat*** and ***Prose*** brings us back to the original demand function specified in equation (1) above.

~~Exam 2~~ Practice Problems **Intro. Econometrics**
Res. Economics 312

1. Using the data in the file *CPS1985-NonLin.mtw* estimate the model(s) required to address the following questions. These are actual survey data collected in 1985 from 534 U.S. citizens. In answering the questions below, let us all assume that an individual's wage ($/hour) is determined by her/his education (in years of schooling) and experience (in years):

$$Wage_i = \beta_0 + \beta_1 Ed_i + \beta_2 Ex_i + u$$

a. Define each of the elements of the model above; i.e., what do *Wage, Ed, Ex, u,* β_0, β_1, and β_2 all represent? Explain what each is and what each means.

b. Estimate the above model and interpret all parameter estimates.

c. Do women earn lower starting salaries *and* lower raises? Specify a model that would allow you to address these two questions. Interpret all estimated parameters. In doing so, be sure to indicate which estimates are statistically different from zero.

d. The question in part c is a joint hypothesis test. Specify the joint hypothesis and complete the test. What conclusion did you reach?

e. Empirical evidence suggests that an individual's returns to experience (raises) are positive early in their career, but eventually decline. Thus, a graph that depicts one's wage profile over time would look like a "hill." Specify a model that will capture this **nonlinear effect of experience on wages**. Explain your model specification and what estimated parameter values you expect. Estimate the model and include a copy of your printout.

f. Interpret the results. How do you interpret the effects of experience on wage?

g. At what age are the returns to an additional year of experience the greatest?