Stata Lab 2

1. Download the dataset WAGE.DTA from my website. Create a new do.file to track your work on this problem set.

**Source:** M. Blackburn and D. Neumark (1992), “Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials,” *Quarterly Journal of Economics* 107, 1421-1436. (We are using only the 1980 data.)

1. Click on the “Data Editor (Browse)” button to view the data directly. Is this dataset cross-sectional, time series, or panel? Close out the data view.
2. Provide some descriptive statistics of the dataset.
	1. What is the sample size?
	2. What is the average IQ among individuals in the sample? Minimum? Maximum?
	3. What is the average wage among individuals in the sample? Minimum? Maximum?
	4. What is the correlation between IQ and wage?
3. Estimate a simple regression model where a one-point increase in *IQ* changes *wage* by a constant dollar amount.
	1. Write your model here:
	2. By how much will a 1 point increase in *IQ* affect *wage* on average?
	3. How does the estimated coefficient compare with the correlation you found in #3d?
	4. By how much will a 10 point increase in *IQ* affect *wage* on average?
	5. What is the R2 for this regression? Would you say the model does a good job of estimating wages? Why or why not?
	6. Type *predict wagefitted* right after running the regression (*wagefitted* can be replaced with anything you want; it’s just something to call the new variable you are asking Stata to generate). This will generate a column/variable in the dataset that gives each fitted value of wage, or *yhat*.
	7. Let’s see visually how your estimated regression line fits the data. Type:

 *graph twoway (line wagefitted IQ) (scatter wage IQ)*

Sketch your graph below and label the sample regression function, SRF.

1. Now estimate a model where each one-point increase in *IQ* has a constant percentage effect on *wage*.
	1. Write your model here:
	2. If *IQ* increases by 10 points, what is the approximate percentage increase in predicted *wage*?
2. Let’s expand the regression model to include multiple explanatory variables.
	1. Estimate a model that looks at the effects of IQ, education, experience, and race on wage. Write the model here:
	2. Interpret each coefficient (the estimate of each β).
	3. How does the R2 for this regression compare to the R2 in #4e above? What does this suggest about this model compared to the one in #4?
3. Run a regression of wage on education and experience.

Let’s take a look at the correlation between education and experience in the sample.

*corr educ exper*

Since there is some correlation between education and experience, it is possible that the regression of wage on education and experience is missing some overlapping effect of education and experience on wage. Let’s create a variable that accounts for this overlapping component; we call it an “interaction term”.

*gen expeduc=exper\*educ*

(*gen* is the command for generating a new variable. You can call the new variable whatever you want. Here I’ve called it *expeduc*.)

Now regress wage on education, experience, and on this new interaction term. How do the parameters on *educ* and *exper* change across the two regression models? Interpret the effect of education on wage. Interpret the effect of experience on wage.

1. Run the regression again but include the possibility for a nonlinear effect of age on wage. First generate the quadratic term for age:

*gen agesq=age^2* (agesq can be anything you want to call it)

What are the signs of the coefficients on *age* and *agesq*? What does this mean about the relationship between wage and age? Explain in words and graphically.

1. Save your do.file and email it to me at rhiller@fairfield.edu.