Stata Lab 1

1. Open up dataset BWGHT.DTA, found in the Content folder on my website.

**Source:** J. Mullahy (1997), “Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behavior,” *Review of Economics and Statistics* 79, 596-593. Data obtained from the 1988 National Health Interview Survey.

1. Take a look at the variable definitions.
2. 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.
3. Type *sum* followed by *cigs faminc* *bwght*.
   1. How many observations are there in this dataset?
   2. What does the mean tell us about a variable?
   3. What does the standard deviation tell us about a variable?
   4. What is the highest family income in the sample?
   5. What is the maximum number of cigarettes smoked by an individual in the sample?
4. Type *tab* followed by *white*. How many people in the sample are white?
5. How many women are in the sample, and how many report smoking during pregnancy?
6. What is the average number of cigarettes smoked per day? Is the average a good measure of the “typical” woman in this case? Explain.
7. Among women who smoked during pregnancy, what is the average number of cigarettes smoked per day? (Type: *sum cigs if cigs>*0) How does this compare with your answer from part #7, and why?
8. Find the average of *fatheduc* in the sample. Why are only 1,033 observations used to compute this average?
9. Report the average family income and its standard deviation in dollars.
10. What is the correlation (*corr*) between the number of cigarettes pregnant women report smoking and the parents’ education level?
11. Open a new do.file by clicking on the “New Do-File Editor” button.

Type the following, replacing the word “filename” with the appropriate name of your dataset:

*clear*

*use “filename.dta”*

1. First let’s see if we can determine which factors contribute to whether or not a woman smokes during pregnancy (dependent variable *cigs*). Type the following in your do file and for each **interpret** the estimated coefficient(s) (the βhats) on each explanatory variable, e.g. “A $1 increase in family income is associated with βhat more cigarettes smoked.”

*reg cigs faminc*

*reg cigs fatheduc*

*reg cigs motheduc*

*reg cigs faminc fatheduc motheduc*

Holding constant father’s and mother’s education, what is the effect of family income on mean number of cigarettes smoked?

How do the coefficient estimates for the final regression compare with that in the first regression? Why are they different?

1. Next let’s see what factors affect birth weight of a child (dependent variable *bwght*). Come up with your own regression model by choosing the explanatory variables you think are relevant in explaining birth weight. Write the model in the space below. Justify the inclusion of each variable. Interpret your results.
2. Save your do.file as BWGHT.DO, print out a copy with your answers.