Problem set 10: week 14

This problem set is to be completed prior to your Stata class. The data set used, crime2.dta, is available on the EC203 website.

1. Open a do file in Stata. All the commands we use in this problem set will be copied into here. This is so you can recall what we have done, and the analysis can be repeated. It will also be useful for you to annotate the do file as you go along.

2. Load the crime.dta dataset into Stata.

3. Open a log file to record the output from Stata. This should be written into either your memory stick or H: drive.

4. Run the following regression: $narr86_i = \alpha + \beta_1 ptime86_i + \beta_2 qemp86_i + \beta_3 pcnv_i + \beta_4 avgsen_i + \beta_5 black_i + \beta_6 hispanic_i + \epsilon_i$. Interpret the new coefficients.

5. Generate a new variable: $arr_i = 1$ if individual has been arrested in 1986 and zero otherwise. What proportion of individuals have been arrested? Run the following regression: $arr_i = \alpha + \beta_1 ptime86_i + \beta_2 qemp86_i + \beta_3 pcnv_i + \beta_4 avgsen_i + \beta_5 black_i + \beta_6 hispanic_i + \epsilon_i$.

6. Interpret each of the coefficients.

7. How many of the predicted values lie outside of the $[0, 1]$ interval? Is this a big concern for our regression?

8. How are the residuals distributed? Do they have a normal distribution? Is this a big concern for our regression?

9. Statistically test for the presence of heteroskedasticity. That is square the estimated residuals and regress them on the X's in your model. Do your squared residuals vary with the X's in the model? Carry out the same test using the **estat hettest** command (with the fs option). What do you conclude? What are the consequences for our regression?

10. Re-run the regression in 5, this time using robust standard errors. Compare the coefficients and standard errors between the robust and non-robust regression. Are there any practical consequences?