Economics 472
Problem Set #6

Note: For all of the computer-based exercises, please include the regression output (you don’t need to include any graphs) with your completed problem set.

(1) Using the baseball data provided on the course website, run the multiple regression

\[ Wins_i = \beta_1 + \beta_2 ERA_i + \beta_3 BatAvg_i + u_i. \]

Verify that the coefficient estimates you obtain from this regression are the same as those presented in your lecture notes (The standard errors will be a bit different, however).

Note: To do this, I would first create a log file, load the data, etc., just as you did in problem set 4. To perform a multiple regression using stata, simply type

```
regress wins era batavg
```

(given, of course, that you have named the variables wins era and batavg in your “infile” command). In other words, to run a multiple regression using STATA simply enumerate all the explanatory variables that you want to include in the regression after the dependent variable (separated by spaces).

(2) Load the California schools data into STATA and perform the multiple regression

\[ TestScore_i = \beta_1 + \beta_2 STR_i + \beta_3 PctEl_i + u_i. \]

(Note: Use the Lab Outline provided on the course website to get the right ordering of the variables in the data set- see the “infile” command on the bottom of the first page of the lab outline. If you use the same names for the variables, “testscr” refers to Test Score, “str” refers to student-teacher ratio, and “elpct” refers to the percent of English learners.) Verify that the coefficients you get are the same as those reported on page 163 of your book.
[NOTE: The standard errors that STATA outputs will be different than those reported in your textbook! The reason is that the standard errors reported in the text are heteroscedasticity-corrected standard errors. (That is, they do not assume that the variance of the error term is constant and equal to $\sigma^2$ for all $i$). To obtain the heteroscedasticity-corrected standard errors in STATA, simply add on a “, robust” command to the end of your regress statement in STATA. That is, you could perform a robust regression for question (1) above by typing

\texttt{regress wins era batavg, robust}

You are not required to perform these robust regressions, but it would be worthwhile for you to try it and then verify that you get the same standard errors as those reported in your book!]

(2a) Can you explain the large reduction in the Str coefficient relative to the simple regression results reported on page 154? Use the omitted variables bias formula discussed in class to illustrate your argument, and also read the discussion related to this issue beginning on page 162.

(2b) Using the same data set, perform the multiple regression

$$\text{TestScore}_i = \beta_1 + \beta_2 \text{STR}_i + \beta_3 \text{Expn}_i + \beta_4 \text{PctEl}_i + u_i,$$

where \text{Expn} denotes expenditure per student (and is called \text{expenstu} in the infile command on the lab outline).

Verify that the estimates you get are the same as those reported on page 165. [NOTE: The book uses \text{Expn}/1,000 so the coefficient you get on \text{Expn} will be the the coefficient in the text divided by 1,000. Though you are not required to do so, you could do the same thing by first generating a new expenditure variable. Simply type

\texttt{generate expend = expenstu/1000}

Then, run your regression using \texttt{str expend} and \texttt{elpct} as the explanatory variables.]
Interpret what the coefficient estimates actually mean (keeping in mind that this is now a multiple regression).

(3) Consider the explanation of results on page 164 of your textbook:

“A 95 percent confidence interval for the population coefficient on STR is $-1.10 \pm 1.96(0.43) = (-1.95, -0.026)$; that is, we can be 95 percent confident that the true value of the coefficient is between -1.95 and -0.026...”

Do you agree with this interpretation of the confidence interval? If not, explain why you disagree.