

Economics 371
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) 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 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 σ^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

regress testscr str elpct, 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!]

(1a) 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.

(1b) Using the same data set, perform the multiple regression

$$TestScore_i = \beta_1 + \beta_2 STR_i + \beta_3 Expn_i + \beta_4 PctEl_i + u_i,$$

where *Expn* denotes expenditure per student (and is called *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 *Expn*/1,000 so the coefficient you get on *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

generate expend = expenstu/1000

Then, run your regression using *str* *expend* and *elpct* as the explanatory variables.

Interpret what the coefficient estimates actually mean (keeping in mind that this is now a multiple regression).

(2) Using the data on the website, *HPRICE1.dta*, perform the following regression:

$$Price_i = \beta_0 + \beta_1 sqrf_t_i + \beta_2 bdrms_i + u_i.$$

In the above, *Price_i* represents the sales price of a house (measured in thousands of dollars), *sqrf_t* is the square footage of the house, and *bdrms* is the number of bedrooms.

Note: This data set is already in STATA format. To load it into STATA, simply type “use a:HPRICE1” (assuming that you saved the file to the a: drive, and the title is capitalized). Note that all variables are named already, so there is no need for you to do this on your own.

(2a) Interpret the coefficient estimates from this regression. Are the results consistent with what you expect?

(2b) Holding square footage constant, how does adding a bedroom impact the sales price of the house?

(2c) What is the estimated increase in sales price when adding a bedroom which *also* adds 150 square feet to the home?

(2d) Calculate the p -value associated with the null hypothesis that the coefficient on *bdrms* equals zero. (Note: Do not simply report the p -value from the regression output. Discuss how you calculate this on your own, and verify that your result is similar to that reported by STATA).

(2e) The first house in the sample has $\text{sqrft} = 2,438$ and $\text{bdrms} = 4$. Find the predicted sales price for this house.

(2f) The actual sales price of this house was \$300,000. Find the estimated residual associated with this house. Based on your results, can you conclude that the buyer overpaid or underpaid for the house?

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

“A 95 percent confidence interval for the population coefficient on STR is $-1.10 \pm 1.96(.43) = (-1.95, -.026)$; that is, we can be 95 percent confident that the true value of the coefficient is between -1.95 and -.026...”

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