Homework Set 5

Due on May 22th, 2017

All questions are from the end-of-chapter exercises in Chapter 7. The question numbers refer to those in the book. I highly recommend you reading the textbook and lecture notes before completing the homework questions. When reading the textbook, please pay attention to the sections on how to interpret the estimated coefficients.

Exercises

7.7 Question 6.5 reported the following regression (where standard erros have been added):

$$\widehat{Price} = \underset{(23.9)}{119.2} + \underset{(2.61)}{0.485}BDR + \underset{(8.94)}{23.4}Bath + \underset{(0.011)}{0.156}Hsize + \underset{(0.00048)}{0.002}\ Lsize$$
$$+ \underset{(0.311)}{0.090}Age - \underset{(10.5)}{48.8}Poor,\ \bar{R}^2 = 0.72,\ SER = 41.5$$

a. is the coefficient on BDR statistically significantly different from zero?

b. Typically five-bedroom houses sell for much more than two-bedroom houses. Is this consistent with your answer to (a) and with the regression more generally?

c. A homeowner purchase 2000 square feet from an adjacent lot. Construct a 99% confident interval for the change in the value of her house.

d. Lot size is measured in square feet. do you think that another scale might be more appropriate? Why or why not?

e. The F-statistic fro omitting BDR and Age from the regression is $F = 0.08$. Are the coefficients on BDR and Age statistically different from zero at the 10% level?

7.8 Referring to Table 7.1 in the text:

a Construct and compute the R^2 from $\bar{R}^2$ for each of the regression.

b Construct the homoskedasticity-only F-statistic for testing $\beta_3 = \beta_4 = 0$ in the regression shown in column (5). Is the statistic significant at the 5% level?

c Construct a 99% confidence interval for β_1 for the regression in column (5).

7.9 Consider the regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Use Approach #2 from Section 7.3 to transform the regression so that you can use a t-statistic to test

a. $\beta_1 = \beta_2$

b. $\beta_1 + a\beta_2 = 0$, where a is a constant;

c. $\beta_1 + \beta_2 = 1$. (*Hint*: You must redefine the dependent variable in the regression)

7.11 A school district undertakes an experiment to estimate the effect of class size on test scores in second grade classes. The district assigns 50% of its previous year's first graders to small second-grade classes (18 students per classroom) and 50% to regular-size classes (21 students per classroom). Students new to the district are handled differently: 20% are randomly assigned to small classes and 80% to regular-size classes. At the end of the second-grade school year, each student is given a standardized exam. let Y_i denote the exam score for i^{th} student, X_{1i} denote a binary variable that equals 1 if the student is assigned to small class, and X_{2i} denote a binary variable that equals 1 if the student is newly enrolled. Let β_1 denote the class effect on test scores of reducing class size from regular to small.

a. Consider the regression $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$. Do you think that $E(u_i|X_{1i}) = 0$? Is the OLS estimator of β_1 unbiased and consistent? Explain.

b. Consider the regression $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$. Do you think that $E(u_i|X_{1i}, X_{2i})$ depends on X_1? Is the OLS estimator of β_1 unbiased and consistent? Explain. Do you think that $E(u_i|X_{1i}, X_{2i})$ depends on X_2? Will the OLS estimator of β_2 provide an unbiased and consistent estimate of the causal effect of transferring to a new school (that is, being a newly-enrolled student)? Explain.

Empirical Exercise

E7.1 use the data set **CPS08** described in Empirical Exercise 4.1 to answer the following questions.

a. Run a regression of average hourly earnings (*AHE*) on age (*Age*). What is the estimated intercept? What is the estimated slope?

b. Run a regression of *AHE* on *Age*, gender (*Female*), and education (*Bachelor*). What is the estimated effect of *Age* on earnings? Construct a 95% confidence interval for the coefficient on *Age* in the regression.

c. Are the results from the regression in (b) substantively different from the results in (a) regarding the effects of *Age* and *AHE*? Does the regression in (a) seem to suffer from omitted variable bias?

d. Bob is a 26-year-old male worker with a high school diploma. Predict Bob's earnings using the estimated regression in (b). Alexis is a 30-year-old female worker with a college degree. Predict Alexis's earnings using the regression.

e. Compare the fit of the regression in (a) and (b) using the regression standard errors, R^2 and $\bar{R}^2$. Why are the R^2 and $\bar{R}^2$ so similar in regression (b)?

f. Are gender and education determinants of earnings? Test the null hypothesis that *Female* can be deleted from the regression. Test the null hypothesis that

Bachelor can be deleted from the regression. Test the null hypothesis that both *Female* and *Bachelor* can be deleted from the regression.

g. A regression will suffer from omitted variable bias when two conditions hold. What are these two conditions? Do these conditions seem to hold here?