Problem Set No. 4
Introduction to Econometrics

Professor Adriana Kugler

Due on November 19, 2009

1. The questions below refer to the following table of estimated regressions of the *ln*(*Price*), i.e., the natural log of sales prices, on various variables:

Regressor	(1)	(2)	(3)
Size	0.00042		
	(0.000038)		
ln(*Size*)		0.69	0.57
		(0.054)	(2.03)
ln(*Size*)^2			0.0078
			(0.14)
Pool	0.082	0.071	0.071
	(0.032)	(0.034)	(0.036)
View	0.037	0.027	0.027
	(0.029)	(0.028)	(0.029)
Condition	0.13	0.12	0.12
	(0.045)	(0.035)	(0.036)
Intercept	10.97	6.60	7.02
	(0.069)	(0.039)	(7.50)
SER	0.102	0.098	0.099
R^2	0.72	0.74	0.73

where the variables are defined as follows:

Size =house size (in squared feet)

Pool =binary variable (1if house has swiming pool,0 otherwise)

View =binary variable (1 if house has a nice view, 0 otherwise)

Condition =binary variable (1 if house is in excellent condition, 0 otherwise)

a. Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to the house? Construct a 95% confidence interval for the percentage change in price.
b. Comparing columns (1) and (2), is it better to use *Size* or *ln*(*Size*) to explain house prices?
c. Using the regression results in column (2), what is the estimated effect of pool on price? (Hint: make sure you get the right units.)
d. Is the quadratic term $ln(Size)^2$ important?

2. On my website (http://www.uh.edu/~adkugler/ProblemSets.html) you will find a file called *CPS*04.
   a. Run a regression of hourly earnings (*AHE*) on age (*AGE*), gender (*Female*) and education (*Bachelor*). If *Age* increases from 25 to 26, how are earnings expected to change? If Age increases from 33 to 34, how are earnings expected to change?
   b. Run a regression of the logarithm average hourly earnings, *ln*(*AHE*), on *AGE*, *Female*, and *Bachelor*. If Age increases from 25 to 26, how are earnings expected to change? If Age increases from 33 to 34, how are earnings expected to change?
   c. Run a regression of the logarithm average hourly earnings, *ln*(*AHE*), on *AGE*, AGE^2, *Female*, and *Bachelor*. Do you prefer the regression in c. or b.?
   d. Run a regression of the logarithm average hourly earnings, *ln*(*AHE*), on *AGE*, *Female*, and *Bachelor*, and the interaction term *Female* × *Bachelor*. What does the coefficient on the interaction term measure?

Documentation for CPS04 Data

Each month the Bureau of Labor Statistics in the U.S. Department of Labor conducts the "Current Population Survey" (CPS), which provides data on labor force characteristics of the population, including the level of employment, unemployment, and earnings. Approximately 65,000 randomly selected U.S. households are surveyed each month. The sample is chosen by randomly selecting addresses from a database comprised of addresses from the most recent decennial census augmented with data on new housing units constructed after the last census. The exact random sampling scheme is rather complicated (first small geographical areas are randomly selected, then housing units within these areas randomly selected); details can be found in the *Handbook of Labor Statistics* and is described on the Bureau of Labor Statistics website (www.bls.gov).

The survey conducted each March is more detailed than in other months and asks questions about earnings during the previous year. The file **CPS04** contains the data for 2004 (from the March 2005 survey). These data are for full-time workers, defined as workers employed more than 35 hours per week for at least 48 weeks in the previous year. Data are provided for workers whose highest educational achievement is (1) a high school diploma, and (2) a bachelor's degree.

Series in Data Set:

FEMALE: 1 if female; 0 if male

YEAR: Year

AHE : Average Hourly Earnings

BACHELOR: 1 if worker has a bachelor's degree; 0 if worker has a high school degree