BT2101 Week9 Tutorial

1. Data were collected from a random sample of 220 home sales from a community in 2003. Let ***price*** denote the selling price (in $1,000), ***BDR*** denote the number of bedrooms, ***bath*** denote the number of bathrooms, ***Hsize*** denote the size of the house (in square feet), ***Lsize*** denote the lot size (in square feet), ***Age*** denote the age of the house (in years), and ***Poor*** denote a binary variable that is equal to 1 if the condition of the house is reported as "poor". An estimated regression yields

$$\widehat{Price} = 119.2 + 0.485 * BDR + 23.4 * Bath + 0.156 * Hsize + 0.002 * Lsize + 0.090 * Age - 48.8 * Poor, \qquad \overline{R^2} = 0.72, SER = 41.5$$

a) Suppose that a homeowner converts part of an existing family room in her house into a new bathroom. What is the expected increase in the value of the house?

b) Suppose that a homeowner adds a new bathroom to her house, which increases the size of the house by 100 square feet. What is the expected increase in the value of the house?

c) What is the loss in value if a homeowner lets' his house run down so that its condition becomes poor?

2. Critique each of the following proposed research plans. Your critique should explain any problems with the proposed research and describe how the research plan might be improved. Include a discussion of any additional data that need to be collected and the appropriate statistical techniques for analyzing the data.

a) You as a consultant are interested in determining whether a large semi-conductor firm is guilty of gender bias in setting wages. To determine potential bias, the consultant collects salary and gender information for all of the firm's engineers. The consultant then plans to conduct a "difference in means" test to determine whether the average salary for women is significantly less than the average salary for men.

b) You as a student are interested in determining whether time spent in post-graduate program has a permanent effect on a person's wage rate. You collect data on a

random sample of NUS alumni (post-graduate). You collect similar data on a random sample of people who only finished their undergraduate program. The data set includes information on each person's current wage, age, ethnicity, gender, tenure (time in current job), occupation, and union status, as well as whether the person's previous education history. You plan to estimate the effect of education history on wages by regressing wages on an indicator variable for education history, including in the regression the other potential determinants of wages (occupation, tenure, union status, and so on).

3. (Y_i, X_{1i}, X_{2i}) satisfy the assumptions in multicollinearity. You are interested in β_1 , the causal effect of X_1 on Y. Suppose that $X_1$and X_2 are uncorrelated. You estimate β_1 by regressing Y onto X_1 (so that X_2 is not included in the regression. Does this estimator suffer from omitted variable bias? Explain?

4. Suppose that a company collects data on houses that have sold in a particular neighborhood over the past year and obtains the regression results in the table below.

Dependent Variable (DV): ln(Price)					
Independent Variable (IV)	Model 1	Model 2	Model 3	Model 4	Model 5
Size	0.00042 (0.000038)	-	-	-	-
ln(Size)	-	0.69 (0.054)	0.68 (0.087)	0.57 (2.03)	0.69 (0.055)
$\boldsymbol{ln(Size)^2}$	-	-	-	0.0078 (0.14)	-
Bedrooms	-	-	0.0036 (0.037)	-	-
Pool	0.082 (0.032)	0.071 (0.034)	0.071 (0.034)	0.071 (0.036)	0.071 (0.035)
View	0.037 (0.029)	0.027 (0.028)	0.026 (0.026)	0.027 (0.029)	0.027 (0.03)
Pool X View	-	-	-	-	0.0022 (0.10)
Condition	0.13 (0.045)	0.12 (0.035)	0.12 (0.035)	0.12 (0.036)	0.12 (0.035)
Intercept	0..97 (0.069)	6.60 (0.39)	6.63 (0.53)	7.02 (7.50)	6.60 (0.40)
Summary Statistics					
SER	0.102	0.098	0.099	0.099	0.099
$\bar{R}^2$	0.72	0.74	0.73	0.73	0.73

a. Using the results in column (1), what is the expected change in price of building a 500-square-foot addition to a 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 column (2), what is the estimated effect of pool on price? Construct a 95% confidence interval for this effect.
d. The regression in column (3) adds the number of bedrooms to the regression. How large is the estimated effect of an additional bedroom? Is the effect statistically significant? Why do you think the estimated effect is so small? (Hint: Which other variables are being held constant?)
e. Is the quadratic term ln $(size)^2$ important?
f. Use the regression in column (5) to compute the expected change to price when a pool is added to a house without a view. Repeat the exercise for a house with a view. Is there a large difference? Is the differences statistically significant?