Problem Set 6

This problem set will not be collected. It is for extra practice to help you prepare for the final. Solutions to the problem set are available on Smartsite.

1. Suppose that the size of the squirrel population in Davis depends on the amount of rain Davis receives and the amount of nuts the trees produce. Squirrels dislike rain, so more rain means fewer squirrels. Squirrels like nuts, so more nuts means more squirrels. The true relationship between the squirrel population and rainfall and amount of nuts is given by:

$$P = 1000 - 20R + 4N + \varepsilon \tag{1}$$

   where P is the number of squirrels, R is the average monthly rainfall in inches, N is the average number of nuts produced by a tree in Davis, and ε is an error term that satisfies all of our assumptions. The number of nuts each tree produces depends on how much rain Davis gets. The true relationship between rain and nuts is given by:

$$N = 100 + 2R + \nu \tag{2}$$

   where ν satisfies all or our assumptions.

   (a) Suppose that we run a regression with squirrel population as the dependent variable and rainfall as the independent variable. Will the errors be correlated with the regressor? If so, will they be positively or negatively correlated?

   (b) What is the expected value of the estimated slope coefficient? What would the expected value of the slope coefficient be if the correlation between N and R was zero?

   (c) One way to get a precise value for N is to count the number of nuts produced by every tree in Davis and then divide by the number of trees. Suppose that to save time we decided to randomly sample just a few trees (say 10 trees), count the number of total nuts produced and divide by the number of trees in the sample to get $\tilde{N}$. How does the expected value of $\tilde{N}$ compare to the true value of N? If we used $\tilde{N}$ instead of N to estimate equation (1), would the expected value of the slope coefficient on $\tilde{N}$ be greater than, less than or equal to 4?

   (d) Suppose we were only interested in the relationship between nuts and rainfall. If we used $\tilde{N}$ instead of N to estimate equation (2), how would the estimated slope coefficient be affected?

2. For each scenario below, determine whether the estimated coefficient will be biased. If there will be a bias, determine the sign of the bias. In every scenario, the researcher is running a regression with SAT score as the dependent variable and a dummy variable for whether a person takes an SAT prep course as the independent variable to test whether taking a prep course is associated with a higher SAT score. The dummy variable is equal to one if a person takes a prep course and zero otherwise. Treat each scenario separately (in other words, when answering one part, ignore any of the omitted variables mentioned in the other parts). There may be multiple correct answers for a part, the key is to be able to explain the economic intuition behind your chosen answer.

   (a) Individuals who take prep courses are more likely to have parents that will buy them other test prep materials which help improve scores.

   (b) Individuals who take prep courses do so because they tend to do poorly on standardized tests and need extra help.

   (c) All students sign up for prep courses but due to space limitations, only a random subset of students are admitted into the prep course.

   (d) All students sign up for prep courses but due to space limitations, only those students willing to pay the most are admitted into the prep course.

3. Suppose that the number of individuals who smoke (N) is related to the the number of ads for cigarettes (A), the number of studies about the link between smoking and cancer (S) and a random error term (ε). When the number of studies is held constant at 100, one extra ad is associated with a 5 percent increase in the number of smokers. When the number of studies is held constant at 200, one extra ad is associated with a 1 percent increase in the number of smokers. When there are 100 ads for cigarettes, an additional study is associated with a 5 percent decrease in the number of smokers. If there are no ads or studies, the expected number of smokers is 10,000. Write down an equation that captures the relationship between N, A and S. N should be your dependent variable.