Econometrics: Problem Set 2

1. A researcher wants to understand the causal effect of police on crime rates. To do this he will take a random sample of US cities and regress the crime rate (crimes per capita) on the city's per capita police force size.

a. Explain why this regression will likely suffer from omitted variable bias. What possible variables do you think should be included to control for these issues?
b. Do you expect the true causal effect of police on crime to be positive or negative?
c. Use the OVB formula to determine if this researchers regressions will over-estimate or under-estimate the causal effect of police on crime rates.

2. A large survey was conducted of more than 1 million people aged 30-102. It tracked them for 4 years to see the length of their average nights sleep. It found that people who slept 6 or 7 hours had a significantly lower death rates (i.e. died during the 4 years) than people who slept 8 hours or more. Based on this would you tell people who sleep 9 hours a day to sleep less if they want to live longer? Why?

3. Use the data set 'college_data.csv'.
a. Run a regression of education on distance. What is the estimated effect? Interpret it.
b. No rerun the regression but also control for test, female, black, Hispanic, Hi-income, own home, fathers college, unemployment rate, and manufacturing wage. What is the estimated effect now?
c. Is the difference between the results of (a) and (b) large? Does it seem (a) suffers from important OVB?
d. Compare the fits of the regressions in (a) and (b) using SER and R^2. Why are the R^2 and adjusted R^2 in (b) so similar?
e. What does the coefficient on 'father's college' measure? Interpret the coefficient.
f. Why are 'unemployment rate' and 'manufacturing wage' in the regression? What would you have expected the sign of these to be? Why? Interpret these coefficients.
g. Are the coefficients on 'black' and 'Hispanic' similar? Test if these are different (recall you will need to load the library 'car').
h. Someone is a black man who's high school was 20 miles from the nearest college. His test score was 58, his family income in 1980 was $26,000 and his family owned their home. His mother attended college but not his father. The unemployment rate in his county is 7.5% and the average manufacturing wage is $9.75. What is his predicted years of schooling using your regression?
i. Create a table like those in the text or our notes with results from both of your regressions from (a) and (b).