

**Homework #3**  
**Economics 113**  
**Introduction to Econometrics**  
**Professor Spearot**  
**Due Monday, February 9th, 2008 – Beginning of class**

For the following examples, discuss whether each satisfy the four assumptions we use for linear regression. If not, which assumptions are violated, and why?

a) I wish to examine the relationship between wages and education for residents of California. To do so, I collect data from a random Econ 113 class on years of postsecondary education (*educ*), the most recent wage offer of each student (*wage*), and estimate the following equation:

$$wage = \beta_0 + \beta_1 educ + \varepsilon$$

*This linear regression violates the following assumptions:*

- (1) Linear in parameters: The relationship between wage and education is not linear. It is not reasonable that another year of education always increases the wage by the same amount, regardless of the initial level of education.*
- (2) Random sampling: The linear regression is aimed for testing the relationship between wages and education for residents of California. However, the sample is not selected at random from the population of interest but only from college students.*
- (3) Zero conditional mean: The error term  $\varepsilon$  might contain factors such as unobserved ability or effort, which affects earnings and is correlated with education.*

b) I run the same regression as in (a) but instead I decide to use only a sample of seniors, who all have been in school for four years.

*This will only worsen the violation of random sampling assumption and make the estimates even more biased. Further, and more critically, there is no variation in the variable “educ”, and thus we would not be able to estimate anything at all!!*

c) Finally, I run the following adjusted specification, where *score* is the score on the first midterm.

$$wage = \beta_0 + \beta_1 educ + \beta_2 score + \varepsilon$$

Does this solve any of the previous problems in (a) or (b)

*This may help a bit in solving the omitted variables problem. However, there may be remaining omitted variables which are biasing the results. For example, perhaps parental education is correlated with the students’ education level, and the ability level which impacts the score on the first midterm. Since parental education is also likely to influence wages, this is a problem.*

d.) What happens to the estimates in (c) if I measure education in months rather than years?

*This is a change of units of measurement of the independent variable. In this case, educ (in months) is defined as educ (in years)\*12. The estimate of  $\beta_1$  (the slope coefficient of education) is divided by 12. The estimates of  $\beta_2$  and  $\beta_0$  don't change.*