Problem Set 3
Econometrics 410
Prof. Taber
Due: Thurs Feb. 17

Problem 1. Suppose you have data on health coverage in the 50 states. You know life expectancy (in years) in the state and you know the fraction of the population in the state that has health insurance. Suppose now that you run a regression of the Life Expectancy (L_i) on the fraction that have health insurance (H_i) and find that

$$L_i = 65 + 7H_i + \hat{u}_i.$$

Note that health insurance is a fraction so it can vary from 0 to 1.

a) Thinking about the model in terms of a descriptive manner, explain what it means that the slope coefficient is 7.

b) Thinking about the model in terms of a causal manner, explain what it means that the slope coefficient is 7.

c) Thinking about the model in terms of a forecasting model, explain what it means that the slope coefficient is 7.

Problem 2. Consider running a regression of y on x and also a regression of y on z where we have changed the units so that

$$z = 10x.$$

Use the formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$ to show what happens to the estimated regression coefficients in the two regressions.

Problem 3. In the second set of lecture notes we talked about the effects of cigarette smoking on birthweight. We defined a model

$$E(O \mid C = c) = \beta_0^1 + \beta_1^1 c$$

where O is birthweight in ounces and C is cigarettes smoked per day.

Suppose instead we used the model

$$E(L \mid P = p) = \beta_0^4 + \beta_1^4 p.$$

where L is birthweight in pounds and P is packs of cigarettes smoked per day (there are 20 cigarettes in a pack). What is the relationship between β_0^1 and β_0^4 and between β_1^1 and β_1^4?

Problem 4. Take the data set consump.dta from Wooldridge. Our goal is to use per capital real disposable income (variable y) to forecast consumption (variable c). The last year in the data is 1995.

a) Suppose that you knew that income in 1996 was $19,000, what is your forecast of consumption in 1996?

b) Explore with different models of consumption and income (i.e. log-log, log-linear, etc.). Which model has the highest R^2?

Problem 4. Wooldridge problem C2.3. However in answering both (i) and (ii) first answer as if this is a "descriptive" relationship and then answer as if this is a "causal" effect.