Problem Set 1

Prof. Kaltenberg
Econ 585: Applied Econometrics

Deadline: February 15, 2021 at 11pm

1 Problem 1

Suppose you are interested in estimating the effect of the number of digital ads (measured in minutes of exposure) on sales. The target population are 18-25 year olds.

1. You are given a grant to run a controlled experiment. Explain how you would structure the experiment in order to estimate the causal effect of ads on sales.
2. Consider the following model where sales and ads are randomly sampled from the population. The population model is:

$$sales = \beta_0 + \beta_1 ads + u \tag{1}$$

3. List at least two factors contained in u. Are these likely to have a positive or negative correlation with ads?
4. In the above equation, what should be the sign of β_1 if the ads are effective?
5. In the above equation, what is the interpretation of β_0.

2 Problem 2

Consider the savings function:

$$sav = \beta_0 + \beta_1 inc + u, u = \sqrt{inc \cdot e} \tag{2}$$

where e is a random vairable with $E(e) = 0$ and $Var(e) = \sigma_e^2$ Assume that e is independent of inc.

1. Show that $E(u|inc) = 0$, so that the key zero conditional mean assumption (SLR 4) is satisfied. [Hint: If e is independent of inc, then $E(e|inc) = E(e)$]
2. Show that $Var(u|inc) = \sigma_e^2$, so that the homoskedasticity Assumption 5 is violated. In particular, the variance of sav increases with inc. [Hint: $Var(e|inc) = Var(e)$ if e and inc are independent]
3. Provide a discussion that supports the assumption that the variance of savings increases with family income.

3 Problem 3

The following equations were estimated using the data in LAWSCH85:

$$\hat{lsalary} = 9.90 - .0041rank + .294GPA \tag{3}$$

(.24) (.0003) (.069)
n=142, $R^2 = .8238$

$$\hat{lsalary} = 9.86 - .0038rank + .295GPA + .00017age \tag{4}$$

(.29) (.0004) (.083) (.00036)
n=99, $R^2 = .8036$

How can it be that the R-squared is smaller when the variable age is added to the equation?

4 Problem 4

Consider the multiple regression model containing three independent variables, under assumptions MLR1-ML4:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = \beta_3 x_3 + u \tag{5}$$

You are interested in estimating the sum of the parameters on $x_!$ and x_2; call this $\theta_1 = \beta_1 + \beta_2$

1. Show that $\hat{\theta_1} = \hat{\beta_1} + \hat{\beta_2}$ is an unbiased estimator of θ_1

5 Problem 5: Application in R

Use the data in GPA1 to answer this question to answer this question. We can compare multiple regression estimates, where we control for student achievement and background variables. Be sure to include your R script with your submission.

1. In the simple regression equation

$$colGPA = \beta_0 + \beta_1 PC + u \tag{6}$$

   obtain $\hat{\beta_0}$ and $\hat{\beta_1}$. Interpret these estimates.

2. Now add the controls hsGPA and ACT. Does the coefficient of PC change from part i? Does $\hat{\beta_{hsGPA}}$ make sense?

3. What is worth more, owning PC or having 10 more points on the ACT score:

4. Add the binary indicators for the parents being college graduates by interacting the indicators for the parents with PC. Interpret the coefficients. Does this interaction make sense to include? Explain.

5. Suppose someone looking at your regression from part 4 says to you, "The variables hsGPA and ACT are probably pretty highly correlated, so you should drop one of them from the regression." How would you respond?