SSA 456 Problem Set--due 3/11

(selections)

1. **Test scores and tutoring: A fake randomized trial**. You are Secretary of Education in a small state. You are trying to raise achievement test scores and to understand urban-rural and potential gender disparities in the situation. The dataset tutoring\_RCT.csv includes pertinent data on 1,500 graduating students. Read the data using the attached program simulated\_RCT.rmd
   1. Plot the distribution of test scores by gender and by urbanicity.:
   2. Based on these graphs, what is your hypothesis on the relationship between these variables and test scores? One way to test your hypothesis is to run a basic regression that seeks to predict test score as a function of these two variables.
   3. Many students take test-prep courses that appear to raise their test scores. The Secretary is worried that this creates a socio-economic gap. The variable **dose** indicates the number of tutoring months a student has received. What happens if you include that dose variable in your regression?
   4. Interpret your results. In what ways is this a better approach than whan you did in (b)? In what ways is it potentially more biased?

After you did the analysis, you got a call from a University of Chicago prof. This is actually data from a randomized trial involving 1500 students. The treatment group (Z=1) got Uber or Lyft gift cards they could use in travel to tutoring sessions. The control group received a $25 Best Buy gift card.

* 1. Verify the experiment was properly randomized.

* 1. Perform an OLS regression analysis to examine how much assignment to the treatment group increased tutoring:

Also plot up the distribution of tutoring and scores in the treatment and the control groups.

1. Now let’s examine the impact of this intervention.
   1. Let’s start with the ITT estimate of the impact of treatment group assignment to the **offer** of tutoring.
   2. We started trying to understand the impact of the actual tutoring for kids who got it, measured in **dose** units. Compute the Wald estimate of this effect. The below code will work to get you the denominator and the numerator you will need:

aggregate(x = v\_data\_frame$dose,

by = list(v\_data\_frame$Z),

FUN = mean)

aggregate(x = v\_data\_frame$score,

by = list(v\_data\_frame$Z),

FUN = mean)

* 1. We can get slightly more precise estimates by deploying a multivariate analysis that considers the confounding variables woman and urbanicity. This code works. It considers treatment assignment Z to be an instrumental variable for increased dosage of tutoring.

ivreg\_rct = ivreg(score~dose+woman+urban | woman+urban+Z,data=v\_data\_frame)

summary(ivreg\_rct,diagnostics=TRUE)

Compare your answers in (b) and (c).

* 1. Give an intuitive explanation of how your results in (a), (b), and (c) are related to each other.
  2. Explain why the randomized trial seems to suggest different results from what you found in 1(b). What is a potential policy implication of this difference?