**RDD Replication**

**Due Friday, February 19th 2021.**

**Directions**: Please turn in your answers on separate paper, typed, and **beautifully written** with **beautiful tables** and **beautiful figures** using a Markdown format which will compile all of your work beautifully in either Stata or R. **[1]**

**Github repo and summary (worth 2 points)**

1. Download Hansen\_dwi.dta from github at the following address.

use https://github.com/scunning1975/causal-inference-class/raw/master/hansen\_dwi, clear

Create a new github repo named “RDD”.  Inside the RDD directory, put all the subdirectories we’ve discussed in class. Post the link to the repo so I can see it’s done as discussed in your assignment. Save the Hansen\_dwi.dta file into your new /data subdirectory.  Note: The outcome variable is “recidivism” or “recid” which is measuring whether the person showed back up in the data within 4 months.

1. In the writing subdirectory, place your assignment. For the first part of this assignment, read Hansen’s paper in the /articles directory of the main class github entitled “Hansen AER”.  **Briefly summarize this paper**.  What is his research question? What data does he use?  What is his research design, or “identification strategy”?  What are his conclusions?

 # RDD Replication

## Github repo and summary

1. https://github.com/EvelynCheng-Github/RDD

1. \*What is his research question? \*

His research question is that the effect of harsher punishments and sanctions on driving under the influence.

\*What data does he use? \*

He utilizes the administrative records of 512,964 drunk driving parking spots in Washington State. What’s more, blood alcohol content thresholds are important value to control drinking and driving.

\*What is his research design, or “identification strategy”?\*

This article provides quasi-experimental evidence on the impact of severity of punishment on future crimes. In order to provide evidence for these alternative mechanisms, this article examines the degree of change in sanctions and punishments in terms of thresholds, multiple time windows for recidivism, and alcohol-related alternative crimes.

\*What are his conclusions?\*

Conclusion is that the additional sanctions experienced by drunk drivers at BAC thresholds are effective in reducing repeat drunk driving.

**Reproducing somewhat Hansen’s results (but just follow directions) (worth 6 points)**.[2]

1. In the United States, an officer can arrest a driver if after giving them a blood alcohol content (BAC) test they learn the driver had a BAC of 0.08 or higher. We will only focus on the 0.08 BAC cutoff. We will be ignoring the 0.15 cutoff for all this analysis. Create a dummy equaling 1 if **bac1**>= 0.08 and 0 otherwise in your do file or R file.
2. The first thing to do in any RDD is look at the raw data and see if there’s any evidence for manipulation (“sorting on the running variable”). If people were capable of manipulating their blood alcohol content (bac1), describe the test we would use to check for this.  Now evaluate whether you see this in these data?  Either recreate Figure 1 using the bac1 variable as your measure of blood alcohol content or use your own density test from software.  Do you find evidence for sorting on the running variable? Explain your results.  Compare what you found to what Hansen found.
3. The second thing we need to do is check for covariate balance. Recreate Table 2 Panel A but only white male, age and accident (acc) as dependent variables.  Use your equation 1) for this. Are the covariates balanced at the cutoff?  It’s okay if they are not exactly the same as Hansen’s.
4. Recreate Figure 2 panel A-D. You can use the -cmogram- command in Stata to do this. Fit both linear and quadratic with confidence intervals. Discuss what you find and compare it with Hansen’s paper.
5. Estimate equation (1) with recidivism (recid) as the outcome. This corresponds to Table 3 column 1, but since I am missing some of his variables, your sample size will be the entire dataset of 214,558. Nevertheless, replicate Table 3, column 1, Panels A and B.  Note that these are local linear regressions and Panel A uses as its bandwidth 0.03 to 0.13.  But Panel B has a narrower bandwidth of 0.055 to 0.105.  Your table should have three columns and two A and B panels associated with the different bandwidths.:
   1. Column 1: control for the bac1 linearly
   2. Column 2: interact bac1 with cutoff linearly
   3. Column 3: interact bac1 with cutoff linearly and as a quadratic
   4. For all analysis, estimate uncertainty using heteroskedastic robust standard errors. [ed: But if you want to show off, use Kolesár and Rothe’s 2018 “honest” confidence intervals (only available in R).]
6. Recreate the top panel of Figure 3 according to the following rule:
   1. Fit linear fit using only observations with less than 0.15 bac on the bac1
   2. Fit quadratic fit using only observations with less than 0.15 bac on the bac1
7. Discuss what you learned from this exercise. What was the hypothesis you tested and what did you find?  How confident are you in Hansen’s original conclusion? Why/why not?

[1] Again, my preference is that you attempt to create automated tables and automated figures as much as you can.  I’ve placed a simple estout program called ols.do in the estout subdirectory.  You just need to edit.

[2] Much of this advice applies to Stata commands, but you can check the R files for lmb.r to see ways of doing the same in R.