**Assignment 4: Due Saturday, June 13th 2020[[1]](#footnote-1) by Juan Carlos Urueña Mejia - Urosario**

**Directions**: Please turn in your answers on separate paper, typed, and **beautifully written** with **beautiful tables** and **beautiful figures**.

**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. 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. <https://github.com/juancauruena/RDD.git>
2. In the writing subdirectory, place your assignment. For the first part, read Hansen’s paper in the articles directory of the main class github entitled “Hansen AER”. Briefly summarize this paper.

What is his question? What data does he use?

What is the effect of harsher punishments and sanctions on driving under the infuence (DUI)?

Administrative records on 512,964 DUI stops from the state of Washington (WA)

What is his research design? What does he find.

Regression discontinuity derived estimates suggest that having a BAC above the DUI threshold reduces recidivism by up to 2 percentage points (17 percent).

**Replication (worth 6 points)**.[[2]](#footnote-2)

1. In the United States, an officer can arrest a driver if after giving them a blood alcohol content 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. Dummy’s name is “bac”
3. The first thing to do in any RDD is look at the raw data and see if there’s any evidence for manipulation. If people were capable of manipulating their blood alcohol content (bac1), describe the test we would use to check for this.

A) McCrary (2008) suggested examining the density of observations of the assignment variable. If there is a discontinuity in the density of the assignment variable at the threshold for treatment, then this may suggest that some agents were able to perfectly manipulate their treatment status.

In this assignment, I test If people were capable of manipulating their blood alcohol content: under the null the density should be continuous at the cutoff point, Under the alternative hypothesis, the density should increase at the kink.

Now evaluate whether you see this in these data? Recreate Figure 1 using the bac1 variable as your measure of blood alcohol content. Do you find evidence for sorting on the running variable?

Figure 1



The distribution of BAC shows little evidence of endogenous sorting of the threshold studied. The McCrary test implies p-values of 0.0276 at the 0.08 and if the p-value for the McCrary test is less than or equal to .05, then we can be 95% percent sure that there is manipulation going on in the data. However, after run conventional McCrary test there isn´t manipulation on data.

1. The second thing we need to do is check for covariate balance. Recreate Table 2 but only white male, age and accident (acc) as dependent variables. Use your equation 91) for this. Are the covariate balanced at the cutoff? It’s okay if they are not exactly the same as Hansen’s.

Table 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PANEL A | (1) | (2) | (3) | (4) |
|  | bac1 | bac1 | bac1 | bac1 |
| male | 0.000797\* (2.00) |  |  |  |
| white |  | 0.000717 (1.59) |  |  |
| aged |  |  | -0.000235\*\*\* (-17.57) |  |
| acc |  |  |  | -0.00925\*\*\* (-18.23) |
| Constant | 0.0546\*\*\* (154.54) | 0.0546\*\*\* (132.03) | 0.0635\*\*\* (126.72) | 0.0563\*\*\* (325.82) |
| Observations | 23010 | 23010 | 23010 | 23010 |
| PANEL B | (1) | (2) | (3) | (4) |
|  | bac1 | bac1 | bac1 | bac1 |
| male | -0.00122\*\*\* (-4.66) |  |  |  |
| white |  | 0.00315\*\*\* (10.16) |  |  |
| aged |  |  | 0.000450\*\*\* (48.59) |  |
| acc |  |  |  | 0.0189\*\*\* (64.27) |
| Constant | 0.153\*\*\* (658.83) | 0.149\*\*\* (517.98) | 0.136\*\*\* (400.91) | 0.149\*\*\* (1305.34) |
| Observations | 191548 | 191548 | 191548 | 191548 |

*t* statistics in parentheses

\* *p* < 0.05, \*\* *p* < 0.01, \*\*\* *p* < 0.001

1. 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.

Figure 2



1. 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, use heteroskedastic robust standard errors.
2. 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



* 1. Fit quadratic fit using only observations with less than 0.15 bac on the bac1



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. [↑](#footnote-ref-1)
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. [↑](#footnote-ref-2)