**María Andrea Figueroa Suden - 201424648**

**Assignment 4: Due Sunday, June 14th 2020**

**Directions**: Please turn in your answers on separate paper, typed, and **beautifully written** with **beautiful tables** and **beautiful figures**.**[[1]](#footnote-1)**

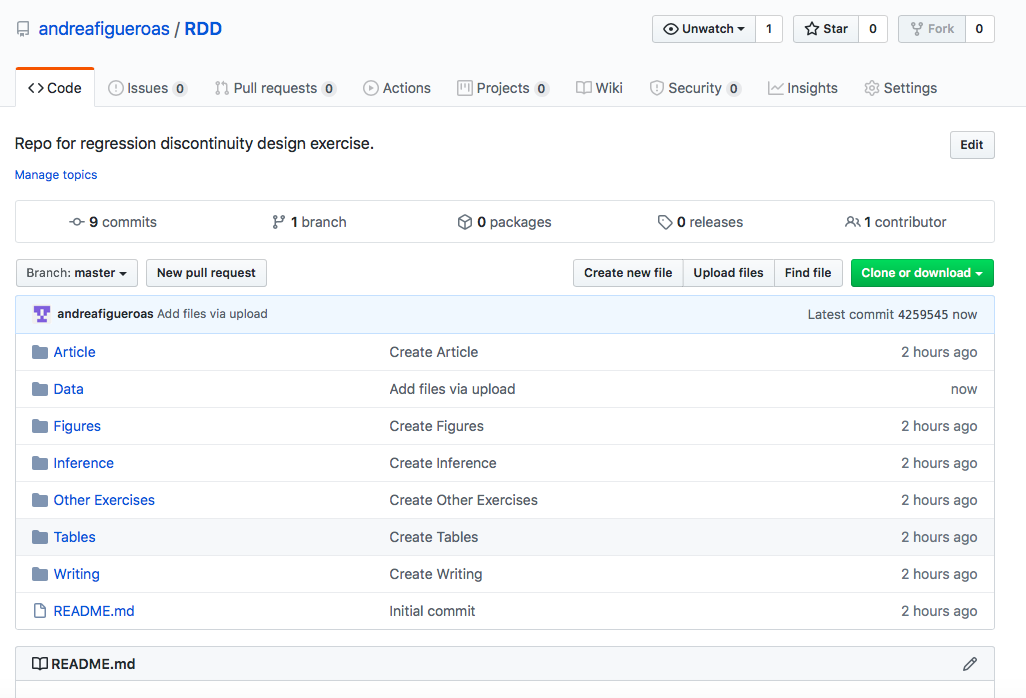
**Github repo and summary (worth 2 points)**

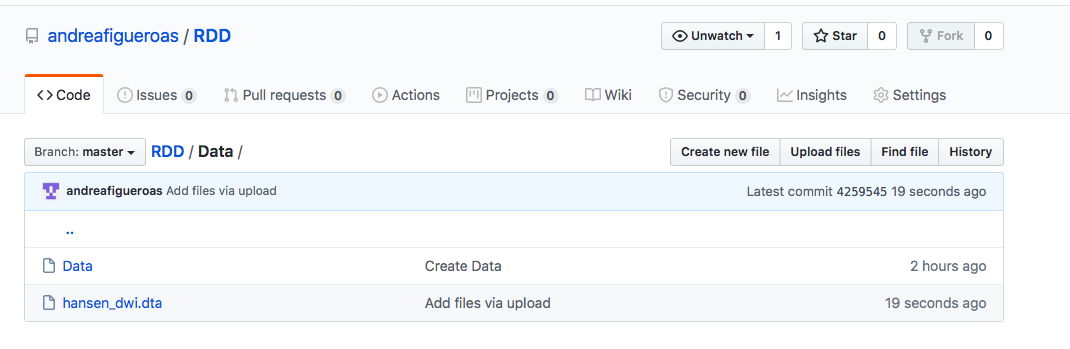
1. Download Hansen\_dwi.dta from github at the following address.

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.

Repolink: <https://github.com/andreafigueroas/RDD>





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?

The article titled “Punishment and Deterrence: Evidence from Drunk Driving” by Benjamin Hansen is frameworked in the literature regarding how the punishment severity might affect the commission of future crimes, phenomenon known as recividism. However, Hansen’s work revolves around a special kind of crime which facilitates testing the aforementioned effect: drunk driving. In this sense, the research question he pretends to solve is which is the effect of punishments and sanctions due to drunk driving on future drunk driving.

The research design chosen by the author is regression discontinuity design, a method that exploits the quantifiable nature of a running variable (in this case blood alcohol content) for which some fixed thresholds are established by law in order to determine punishment severity which can range from license suspension to jail. Furthermore, this identification strategy is built over the assumption that both observables and unobservables from drivers such as age, gender, race and prior offenses, among others, do not change at these thresholds, only the treatment status or the likelihood to be treated. The author presents evidence of these conditions being satisfied to support the consistency of his regression discontinuity estimates.

To answer his principal research question, Benjamin Hansen employs administrative records from the state of Washington. More specifically, he uses 512,964 events of stops due to DUI (Driving Under Influence) from 1999 to 2007. However, he further elaborates on the alternative mechanisms of deterrence, incapacitation and rehabilitation, which could provide likely theoretical reasons for changes in future drunk driving. For this part, he merges the breath test data with another database from the Washington State courts which keeps a centralized record of information such as: fines paid, jail time served, parole, home monitoring, court-ordered license suspensions, alcohol screenings, and other treatments mandated by the court.

Hansen’s results point out that having a blood alcohol content over the established thresholds effectively reduces the likelihood of recidivism within four years of the initial traffic stop among drunk drivers. Also he finds more consistent evidence of the deterrence effect -rather than the incapacitation or rehabilitation effects- sanctions and punishments might have becoming the primary explanation of the reduction in drunk driving.

**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 (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.

gen DUI=1 if bac1>=0.08

replace DUI=0 if DUI==.

1. 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?

The test used to evaluate the manipulation of the sorting variable is known as the McCrary density test. This test checks for bunching at the cutoff (which could mean that the units are being able to set at one of the sides of the cutoff) by means of partitioning the running variable into bins and calculating the frequency (number of observations) of each of these bins. Under the null hypothesis, the density is continuous at the cutoff point, which means there is no manipulation. Meanwhile, under the alternative hypothesis, the density increases at the kink rejecting the no manipulation assumption.

Regarding Hansen’s database, the following Figure 1 shows the blood alcohol content histogram.



According to this graph, there appears to be a slight significant kink (bunching) at the 0.08 cutoff showing kind of an evidence for manipulation. When the McCrary density test is runned we arrive to the following conclusions:

T (statistic) = 2.2032

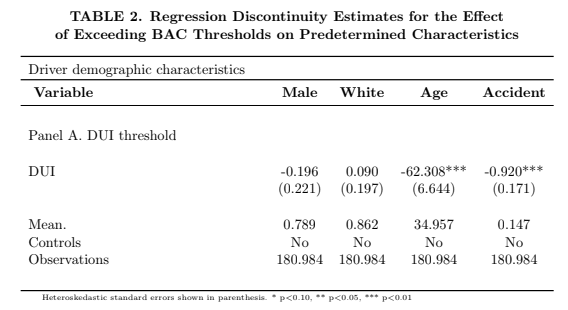
P>|T|= 0.0276

P-value < 0.05 🡪 Reject H0

As shown above in the test results’, the null hypothesis is rejected at a 95% level of confidence meaning there is manipulation of the running variable.

Running placebo cutoffs (different from 0.08), we found out that there is only evidence for manipulation exactly at the 0.08 cutoff, nowhere else. Thus, this leads to ambiguous conclusions about the manipulation of blood alcohol content which works in this research study as the sorting variable.

1. 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.



The recreation of Table 2 was done limiting the sample to the BAC range between 0.03 and 0.20, just as Hansen did since in this range falls 90% of the sample observations.

According to the results from this table, there is evidence that among the covariates only male and white are balanced at the cutoff due to the fact that the regression discontinuity estimates of the effect of having a BAC above the threshold (0.08) lack significance in these two cases. The meaning of this is that both characteristics (male and white) are unaffected by the running variable.

However, there is evidence that the characteristics age and accident are not balanced at the cutoff since the coefficient estimates of exceeding BAC thresholds are significant, in both cases at a 99% level of confidence. This allows to infer that having a BAC above the threshold is negatively related to age and to the fact that an accident was committed at the scene.

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.

Panel A. Accident at scene

1. Linear fit



1. Quadratic fit



Panel B. Male

1. Linear fit



1. Quadratic fit



Panel C. Age

1. Linear fit



1. Quadratic fit



Panel D. White

1. Linear fit



1. Quadratic fit

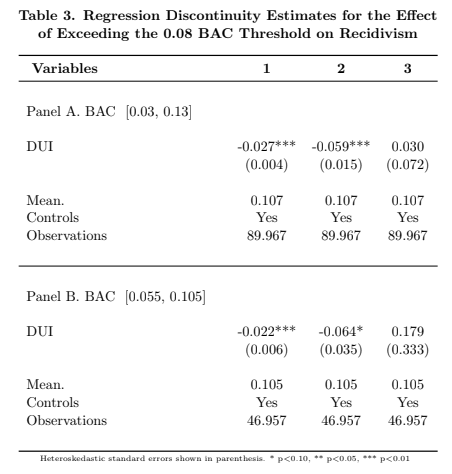


The results obtained support graphically the idea that the predetermined characteristics white and male are not affected by the blood alcohol content threshold since the distribution is almost continuous in the cutoff, in line with the estimation results obtained earlier.

In the case of age, it is clear that the quadratic fit is more suitable in this case because the linear fit does not show the continuous distribution at the cutoff (there is a jump after the threshold). Regarding the predetermined characteristic accident, the graphical evidence suggests it is balanced at the cutoff, contrary to the estimation results obtained earlier.

Comparing these results with Hansens’, it is evident that he performs only the linear fit but in some cases the quadratic fit, for example in age, is a more suitable option. Moreover, although the estimation results from Hansen suggest there is covariate balance at the cutoff, the male characteristic graph from his article seems to have a non-continuous distribution at the cutoff which is not evident in my graph. In addition to this, Hansen employs only 90% of the sample (the observations that fall between the BAC range of 0.03-0.20) and a rectangular kernel, while for the purposes of these graphs I kept the sample complete and did not determine the type of kernel. Therefore, our graphs might not be the same.

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.



The table shows the results for local linear regressions, in panel B using a narrower bandwidth than in panel A, therefore the reduction in the sample size. Regarding the coefficients of interest, I find a reduction in the outcome variable (recividism), controlling for the predetermined characteristics, in line with Hansen`s results.

1. 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)