**Hansen DWI Replication**

**Directions:** Download Hansen\_dwi.dta from github at the following address. Note these data are not exactly the same as his because of confidentiality issues (so he couldn’t share all of it).

<https://github.com/scunning1975/causal-inference-class/raw/master/hansen_dwi>

The outcome variable is “recidivism” or “recid” which is measuring whether the person showed back up in the data within 4 months. Use this data to answer the following questions.

*Placebos, density tests*

1. We will only focus on the 0.08 BAC cutoff; not the 0.15 cutoff. Take the following steps.
   1. Create a treatment variable (dui) equaling 1 if **bac1**>= 0.08 and 0 otherwise in your do file or R file.
   2. Replicate Hansen’s **Figure 1** examining whether there is any evidence for manipulation on the running variable. Produce a raw histogram using bac1, then use the density test in Cattaneo, Titunik and Farrell’s rddensity package. Can you find any evidence for manipulation? What about heaping?
2. 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? Use two separate bandwidths (0.03 to 0.13; 0.055 to 0.105) for estimation.
3. Recreate **Figure 2 panel A-D.** Fit a picture using linear and separately quadratic with confidence intervals.

*Main results*

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, 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).]
2. Repeat but drop units in the close vicinity of 0.08 (i.e., the “donut hole” regression).
3. 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
   3. Use rdplot to do the same.
4. Estimate local polynomial regressions with triangular kernel and bias correction using rdrobust. Experiment with other kernels and polynomials.