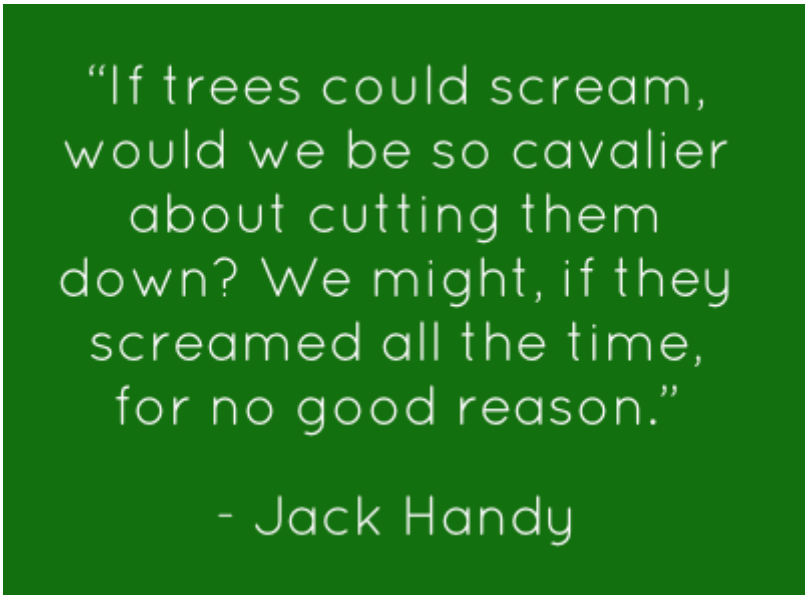
First Replication




Regression Discontinuity and Drunk Driving Recidivism

For this problem you will estimate a regression discontinuity design to test whether having a BAC over the legal limits. You can compare this to the paper by yours truly about DUI punishments we read earlier.

The file is bac.csv

First things first. You must test for non-random sorting in the dataset.

a. Create a histogram of the running variable, BAC. Make sure you do it allowing for discrete bins. Is there evidence of clear sorting at the threshold?

b. Get the R package rddensity. Perform a density test on the running variable. Is there evidence of sorting?

c. Next run a regression discontinuity model. To do so, create a dummy variable for a BAC over .08. Include that dummy variable, and the rescaled BAC (BAC-.08) as a control, and also include an interaction between that dummy variable and the running variable in model. First use age, gender, accident at the scene and race as outcomes. Do those factors shift at .08?

d. Now run a regression of recidivism on the same regression discontinuity design. What is your estimated effect using a bandwidth of .05, and a rectangular kernel (no weighting). Create a visualization of this by graphing the mean recidivism rate against the running variable. Show this for the whole BAC distribution, and the range from .03 to .13. Please include a fitted line.

e. Do the same thing as part c and d but for the aggravated threshold of .151.

f. Now run this model for every possible bandwidth between .01 and .07. Store both the point estimates and lower and upper confidence intervals. Create a scatter plot of the confidence interval and the point estimates. Are the estimates robust? Create a visualization of this.

g. Finally, I want you to reestimate your models using instead of the .08 threshold, every other BAC as the threshold, keeping a bandwidth of .03, between .03 and .12 (i.e. .03, .031, .032, etc). Create a scatter plot of both the point estimates (yaxis) against the potential RD thresholds (x axis). Now create a scatter plot of your test statistic on the null hypothesis of your point estimates (z-stat/t-stat) with the test statistic on the y axis, and the threshold on the x axis. What is the rank of (1 being the largest) of your point estimate estimated at .08? How many estimates did you do? The rank divided by the number of tests is called an empirical pvalue based on a form of permutation inference.