Problem Set 6

Francisco Brady

19 Apr 2024

### Part I: Minimum legal drinking age and mortality (Regression Discontinuity)

This assignment is based on a research paper by Chris Carpenter and Carlos Dobkin, who investigate the effect of alcohol consumption on mortality: https://www.aeaweb.org/articles/pdf/doi/10.1257/app.1.1.164

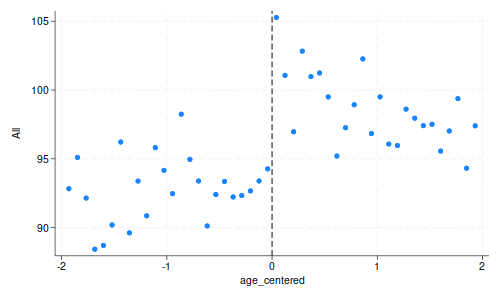
To investigate the causal relationship between alcohol consumption and mortality, the authors exploit a discontinuity in when people are legally able to purchase alcohol in the United States, at age 21. The authors compare the mortality rates of individuals leading up to their 21st birthday, to the mortality rates of individuals just past their 21st birthday. You will use data from age21mort.dta to investigate this question, which contains information from the National Center for Health Statistics, which contains mortality data due to a variety of causes, including internal, external, homicide, suicide, and motor vehicle accidents.

The idea is that the minimum legal drinking age (which is currently set at age 21) produces a sharp difference in access to alcohol for young adults. We assume that observed and unobserved determinants of alcohol consumption and mortality are likely smooth across the age-21 threshold, so the only difference between those just above 21 and just below 21 is their legal access to alcohol.

#### 1. How does alcohol consumption affect overall mortality?

##### a) Produce a scatterplot of mortality due to all causes and the running variable. Do you see evidence of a discontinuity at age 21?

file all\_age.png saved as PNG format



All Mortality and Age

Yes, there is a clear discontinuity as the age variable crosses the centered age threshold (21).

##### b) Write down a regression model to estimate this regression discontinuity, using mortality due to all causes as the outcome variable, and assuming a linear function for the running variable, with constant slopes on either side of the cut-point.

────────────────────────────────────────────────────────────────────────────────────  
 Reg. 1 Reg. 2 Reg. 3 Reg. 4   
   
────────────────────────────────────────────────────────────────────────────────────  
age\_centered -0.975 0.827 -0.831 -1.612   
 (0.66) (0.72) (2.85) (2.28)   
over21 7.663\*\*\* 7.663\*\*\* 9.548\*\*\* 9.753\*\*\*  
 (1.51) (1.27) (1.83) (1.90)   
over21=0 # age\_cen~d 0.000 0.000 0.000   
 (.) (.) (.)   
over21=1 # age\_cen~d -3.603\*\*\* -6.017 -3.289   
 (1.12) (4.53) (3.73)   
agesq -0.840   
 (1.54)   
over21=0 # agesq 0.000   
 (.)   
over21=1 # agesq 2.904   
 (2.26)   
\_cons 91.841\*\*\* 93.618\*\*\* 93.073\*\*\* 92.524\*\*\*  
 (0.71) (0.63) (0.78) (0.86)   
────────────────────────────────────────────────────────────────────────────────────  
Adj. R-Squared 0.577 0.645 0.644 0.674   
R-Squared 0.595 0.668 0.682 0.716   
Observations 48.000 48.000 48.000 24.000   
────────────────────────────────────────────────────────────────────────────────────  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

##### c) [Regression 1]: Now run the regression using Stata. Interpret the coefficient on the treatment variable. (Hint: In Stata, create a variable [call it “over21”] to indicate whether the estimate is above the cut-point, 21. This is your ‘treatment’ variable).

* The coefficient on treatment (over21) is 7.663.
* The estimate is significant at the p<0.01 level.
* This implies that being over 21 is associated with an increase of 7.663 in the all cause mortality rate per 100,000.

##### d) Now let’s allow the running variable to have a different “slope” on either side of the discontinuity. Write down a regression function that assumes a linear function for the running variable, but allows for a different slope on either side of the discontinuity.

##### e) [Regression 2]: Now run the regression specified in part d) in Stata and interpret the two slopes. Do the results suggest different slopes on age on either side of the discontinuity? Explain how you reached your conclusion.

* The results imply different slopes on either side of the cutoff.
* The interaction coefficient is -3.603, and it is significant. This implies that there is a different slope on either side of the age threshold.
* The inclusion of the interaction variable did not change the coefficient on treatment. The coefficient is significant, implying that crossing the over 21 threshold is associated with an increase of 7.663 in the all-cause mortality rate.

##### f) [Regression 3]: Now run the regression assuming a quadratic function for the running variable, with separate quadratic functions on either side of the discontinuity. Does the results suggest a quadratic fit is a better fit of the data? How can you tell? [Hint: You will have to create variables for the quadratic function of the running variable].

* The interaction term between over 21 and age-squared is not significant, indicating that there is not strong evidence of a non-linear relationship near the cutoff.
* Adding the age-squared variable increased the standard errors for the treatment variable, but it remains significant. This implies that the addition of the age squared variable and the interaction term increased noise.
* All of this suggests that the addition of the quadratic term and the interaction does not result in a better fit.

To convince ourselves, we can run a test of joint significance on age-squared and the interaction with the over21 variable. The results suggest that they are not jointly significant:

( 1) agesq = 0  
 ( 2) 1.over21#c.agesq = 0  
  
 F( 2, 42) = 0.93  
 Prob > F = 0.4018

##### g) [Regression 4]: Alternative bandwidth. Now run regression specification #2, limiting the sample to the age range 20-22. How did this narrower age range affect the coefficient on over21?

Compared to regression #2:  
- Limiting the bandwidth decreased the coefficient on over21 from 7.663 to 9.753, the estimate is still significant.  
- This suggests that for individuals within this age range, turning 21 is significantly associated with a larger increase in all-cause mortality.

#### 2. Specific Causes of Death

##### a). Now run regression #2 for the outcomes for motor vehicle deaths, homicides, and suicides, and make a separate table indicating the results for each outcome. Each column should represent the results from a separate regression.

────────────────────────────────────────────────────────────────────  
 MVA Homicide Suicide   
   
────────────────────────────────────────────────────────────────────  
over21 4.534\*\*\* 0.104 1.794\*\*\*  
 (0.68) (0.37) (0.49)   
age\_centered -2.568\*\*\* 0.795\*\*\* 0.029   
 (0.38) (0.26) (0.18)   
over21=0 # age\_cen~d 0.000 0.000 0.000   
 (.) (.) (.)   
over21=1 # age\_cen~d -1.162\*\* -1.145\*\*\* -0.420   
 (0.58) (0.33) (0.42)   
\_cons 29.929\*\*\* 17.425\*\*\* 11.662\*\*\*  
 (0.33) (0.27) (0.21)   
────────────────────────────────────────────────────────────────────  
Adj. R-Squared 0.703 0.332 0.453   
R-Squared 0.722 0.375 0.488   
Observations 48.000 48.000 48.000   
────────────────────────────────────────────────────────────────────  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

##### b). Which of the specific types of deaths are most affected by alcohol consumption? Do these results fit with your expectations?

* The models indicate that motor vehicle accidents and suicide deaths are the most impacted by being over 21 years old. The intuition of these results makes sense to me, those over 21 will have access to situations involving both cars, which could lead to more deaths. The results on the suicide model make some sense, as suicidal ideation affects people of all ages, although those over 21 might be more likely to be successful in their attempts. The effect of the treatment is minimal for deaths by homicide.

#### 3. Placebo test: Does alcohol consumption affect mortality due to internal causes? If alcohol-related deaths are causing the increase in mortality observed for those over the age of 21, we would expect that there would not be increases in mortality due to internal causes, such as heart disease or diabetes. These conditions typically occur later in life, and should not necessarily be affected by acute alcohol consumption.

##### a). Run the same set of regressions in question 1 (regressions#1-4) using mortality due to internal causes (internal) as the outcome variable, and put the results in a third table.

────────────────────────────────────────────────────────────────────────────────────  
 Reg. 1 Reg. 2 Reg. 3 Reg. 4   
   
────────────────────────────────────────────────────────────────────────────────────  
over21 0.392 0.392 1.073 1.692\*\*   
 (0.54) (0.55) (0.80) (0.76)   
age\_centered 1.600\*\*\* 1.618\*\*\* 1.500 0.164   
 (0.25) (0.34) (1.34) (0.90)   
over21=0 # age\_cen~d 0.000 0.000 0.000   
 (.) (.) (.)   
over21=1 # age\_cen~d -0.036 -1.870 -0.348   
 (0.50) (2.01) (1.42)   
agesq -0.060   
 (0.69)   
over21=0 # agesq 0.000   
 (.)   
over21=1 # agesq 1.050   
 (0.99)   
\_cons 20.089\*\*\* 20.107\*\*\* 20.068\*\*\* 19.574\*\*\*  
 (0.27) (0.32) (0.49) (0.44)   
────────────────────────────────────────────────────────────────────────────────────  
Adj. R-Squared 0.790 0.786 0.785 0.420   
R-Squared 0.799 0.799 0.808 0.496   
Observations 48.000 48.000 48.000 24.000   
────────────────────────────────────────────────────────────────────────────────────  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

##### b). In a few sentences, discuss whether this analysis is consistent with the hypothesis that the increase in mortality among those over the age of 21 is driven by alcohol-related incidents.

In models #1 and #2, the coefficient on age is positive and significant, implying that as you get older, rates of mortality from internal causes increase, which is consistent with causes being heart disease or diabetes, or mostly things that affect older people. The coefficient for over 21 in the first two models is not statistically significant.  
In the third model, the estimate on over 21 is larger but not significant, and the standard errors have increased.  
For the fourth model, which uses a restricted dataset around the cutoff, only the over21 coefficient is significant, implying that being over 21 is associated with changes in the mortality rate due to internal causes. This lends supports a conclusion that for those between 20 and 22 years old, at least some deaths can be attributed to internal causes, not legal access to alcohol. This undercuts the idea that the initial results in the preferred specification are driven solely by alcohol deaths.

#### 4. In 3-4 sentences, discuss whether you can conclude a causal relationship between alcohol consumption and mortality from this regression discontinuity analysis. Why or why not? Structure your discussion in terms of internal validity and external validity. Be sure to discuss the main threats to internal validity for a regression discontinuity design, and how your analysis addresses those threats.

I think I can accept the results as causal. Because RDs assume that people on either side of the threshold are similar on observables, and in the restricted model the treatment term retained it’s significance, I find the results slightly believable. Another assumption for sharp RDs is that the population has strict compliance. In reality I think compliance might not be perfect, so perhaps a fuzzy RD specification would be more appropriate (fake ids!). Another threat to internal validity is the results of the placebo test on internal death rates, especially with the restricted model, something is going on there. Another assumption of RDs is that there are no other major policy changes at the cutoff, which is true in this case, 18, 21, and 25 are the big years for youth in the United States, but there are no other policies that I can think of where the cutoff is 21.  
As far as external validity, these results may not translate to a context outside the United States, where legal drinking ages are lower.

### Part II: Internal and External Validity

Listen to one of the following podcasts from Probably Causation podcast, by Jennifer Doleac, Associate Professor at Texas A&M University and (Optional) read/skim the accompanying paper (all posted on blackboard, under Lectures 25 and 26) and answer the questions below:

1. [Analisa Packham, on syringe exchange programs and opioid use (Difference in differences)](https://www.probablecausation.com/podcasts/episode-3-analisa-packham)

#### 1. What is the research question and why is this question important?

How does the introduction of Syringe Exchange Programs (SEPs) impact the incidence of blood-borne illnesses and other drug-related health outcomes. This is important because public health officials have promoted SEPs as cheap and effective programs that can positively affect blood-borne illnesses, however most of the evidence has been observational and not causal. This paper also looks at other outcomes which can shed light on unintended consequences of SEP programs.

#### 2. What are some of the mechanisms through which the “treatment” may affect the outcomes of interest?

SEPs provide clean needles for people who use injection drugs. This can impact blood-borne illnesses through preventing the reuse and sharing of used needles. This is the first order outcome to measure, but there may be other impacts as well. It’s possible that SEP sites could help people access options for treatment, and in that way improve health outcomes, such as drug-related hospitalizations and mortality. For the crime-related outcomes, SEPs could affect drug-related arrests and thefts because many health interventions are done in partnership with local law enforcement, and so may reflect changing responses to people struggling with addiction.

#### 3. How did the authors use an RCT/RD/DiD study to answer their research question? (An intuitive explanation is appropriate here – no need for equations or nitty gritty details.)

To estimate the effects of new SEPs, the study compares counties that open new SEPs within the last 10 years with counties that already had existing SEPs and did not open any new ones during the study period. With this specification, they are able to isolate the change in outcomes for the treated counties while accounting for any unrelated trends that affected both counties similarly over time.

#### 4. What were the main research findings?

Four main outcomes: HIV Rates, Opioid and drug-related mortality, opioid-related hospital visits, drug-related arrests.  
- HIV Rates: The new SEPs led to significant reductions in HIV diagnoses, with around a 25% decrease in treated counties. The effect grows over time to about 30%. This translates to 6000 cases/year when extrapolated nationwide.  
- Mortality: New SEPs led to an increase in drug-related mortality, with the effects being driven by heroin deaths. The estimates indicate a 13.5% increase in drug-related mortality. The effects are being driven by rural areas.  
- Hospitalizations: SEPs are associated with an 8.9% increase in opioid-related emergency department visits.  
- Arrests for opioid-posession increases 16.4% after a new syringe exchange program is opened.

#### 5. What were some of the challenges mentioned by the author, or brought up by the host (Jen Doleac) threatening the ability to claim a causal relationship between the treatment of interest and the outcomes (e.g. what are the threats to internal validity)?

* It’s possible that the effect is really just measuring some preexisting trends that are in the data.
* Opening of SEPs is not randomly assigned, so those counties that decide to open might look very different that other counties.
* Opioid mortality might be trending up in those counties.
* Counties that decide to open SEPs might experience a new wave of drug use.
* Any changes could be attributed to changes in the population (new injection drug users moving into counties)

#### 6. What are the threats to external validity? Could the findings be applied elsewhere? Discuss specific contexts in which the findings could be applied more broadly.

* These findings are US specific. Packham mentions that in some medical studies, US outcomes for SEPs are compared with other countries, which can be problematic because those countries have fundementally different health systems.
* There are important rural and urban differences which may affect the generalizability of the results.

#### 7. Anything else you want to share about the paper or research topic.

I didn’t like the part where they mentioned that as economists, they were somehow able to deal with this issue more dispassionately. It came off as rude to other disciplines. This is a problem in economics more generally, where subject matter expertise is written off as less useful than econometric rigor.