**Econ 300 Spring 2020**

1. To implement Sharp RD, you run a linear regression:

, where S is the running variable subtracted the cutoff value, Above=1 if S 0, Above=0 if S<0. Y is outcome. Which parameter gives the causal impact in RD?

**a. **

b. 

c. 

d. 

2. You are studying the impact of class size on earnings. You know that the rule determining class size is 1) If there are 30 students in a grade, the class size is 30; 2) if there are 31 students in a grade, students will be split into 15, and 16. In this example, what is the running variable?

a. Class size

b. Earning

**c. Number of students in a grade**

d. Number of grade in a school

3. Which one of the following is **FALSE** about RD?

**a. When implementing RD, linear model always gives us the best fit**

b. Fuzzy RD is a type of RD

c. Sharp RD is a type of RD

d. The idea of RD is to mimic randomization around the cutoff

4. Suppose you are studying the impact of smoking on health. You know on 9/1/2009 there is an increase in Tabaco tax in California but not in other states. Which identification strategy is best fit for this analysis?

**a. Differences-in-difference**

b. Regression discontinuity

c. Randomized control trial

d. Simple OLS

5. Which interpretation about RD estimate is **FALSE**?

a. RD has great tests to explicitly check its identification assumption

b. RD’s identification assumption will fail if there is perfect manipulation

c. RD estimate is a LATE because it estimates the average treatment effect only for people around the cutoff value

d. RD estimate is a LATE because sometimes it estimates the average treatment effect for people who are late.

6. Which is **NOT** a test for the identification assumption?

a. Pre-trend test

b. Covariate smoothness test

c. Balance table

**d. F-test**

**Part II. Short questions.**

1. a. Use an example to explain how you use regression discontinuity method and why you can get credible causal estimate. (You can take an example of either a Sharp or Fuzzy RD)(5pts)

b. Explain the difference between Sharp and Fuzzy RD. (5pts)

**Answers:**

**a. Example: What is the effect of gaining legal access to drinking affect death rate. We compare the average death rate of people who just turned 21 and people who are a bit under 21. The reason we can get credible causal estimate is that we assume those two group of people are similar in all other ways than whether have legal access to drinking. This is an “apple to apple” comparison.**

**b. Sharp RD: the treatment status is completely determined by the running variable. Fuzzy RD: the treatment is only partially determined by the running variable**

2. Suppose you are interested in the research question: does raising minimum wage lead to lower employment? You know there is a minimum wage increase for the state of New York in 2012 but not for other states.

a. Can you compare the employment rate before and after this policy change in NY to get the causal effect of the minimum wage increase on employment? Explain.(5pts)

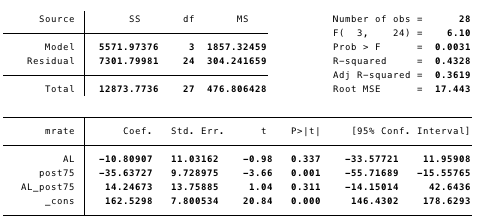
b. Can you design a research to get the causal effect? Explain your research design. (Don’t have to write the regression)(5pts)

**Answers:**

**a. No. A simple before-after comparison may give us a biased estimate because this effect may be confounded by other omitted variables such as general time trend.**

**b. We can find one or several comparison states that would have followed the same trends in employment rate with New York. We take the difference in employment before and after the policy kicks in for comparison group. Then subtract this second difference from the first difference. This is a differences-in-differences estimate, and it is the causal effect of the minimum wage law on employment rate.**

3. Recall in the class for DID, we examined the topic whether lowering Minimum Legal Drinking Age (MLDA) leads to more death. The state of AL lowered MLDA from 21 to 19 in 1975, while the state of AR didn’t. We specified a DID regression to get the causal estimate of this policy change on death rate. Below are the regression results. AL is a dummy variable whether state is AL. Post75 is a dummy whether the year is after 1975. AL\_post75 is the interaction term of the two dummy variables. Death rate is defined as number of death among 100,000 people.



a. What is the causal impact of lowering MLDA on death rate? Does the sign meet your expectation? (5pts)

b. Look at the standard error and t-stat of the causal estimate, what can you conclude? (5pts)

c. What does the coefficient of post75 (-35.63727) mean? (5pts)

**Answers:**

**a. The lowering of MLDA increase the number of death by 14.24673 per 100,000 people. Yes. It meets my expectation since drinking at a younger age may create irrational behaviors which possibly lead to deaths.**

**b. Standard error for the causal estimate is 13.75885. Its t-stat is 1.04<2, so this estimate is not statistically different from 0. This means this effect may simply be due to sampling variation and it is not true in the population.**

**c. The coefficient for post75 is the difference in employment before and after the policy change for the comparison group. (second difference) It is the general time effect.**

**Part III. Long questions.**

1. To study whether Medicare coverage contributes to better health status, Card et al use a regression discontinuity design. People who are 65 and older are qualified for Medicare and who are under 65 are generally not qualified, with the exception that they are receiving Social Security Disability Insurance.

a. What is the running, outcome and treatment variables in this example?

**Outcome: health status**

**Treatment: Medicare coverage**

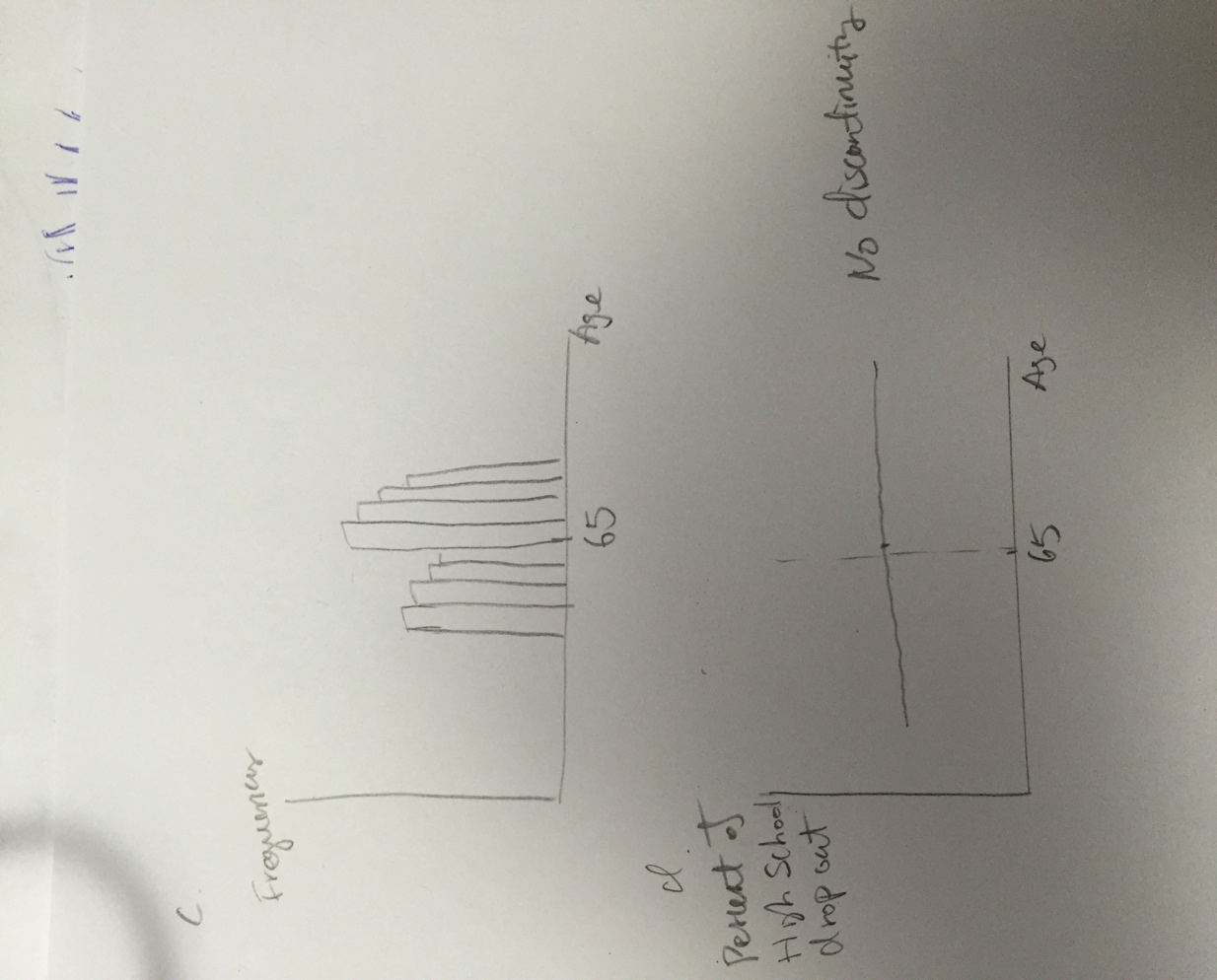
**Running: age**

b. What is the RD identification assumption in this example?

**People just under 65 and over 65 are similar in both observable and unobservable ways.**

**(other ways of expressing the identification assumption: apple to apple/potential outcome is continuous around the cutoff)**

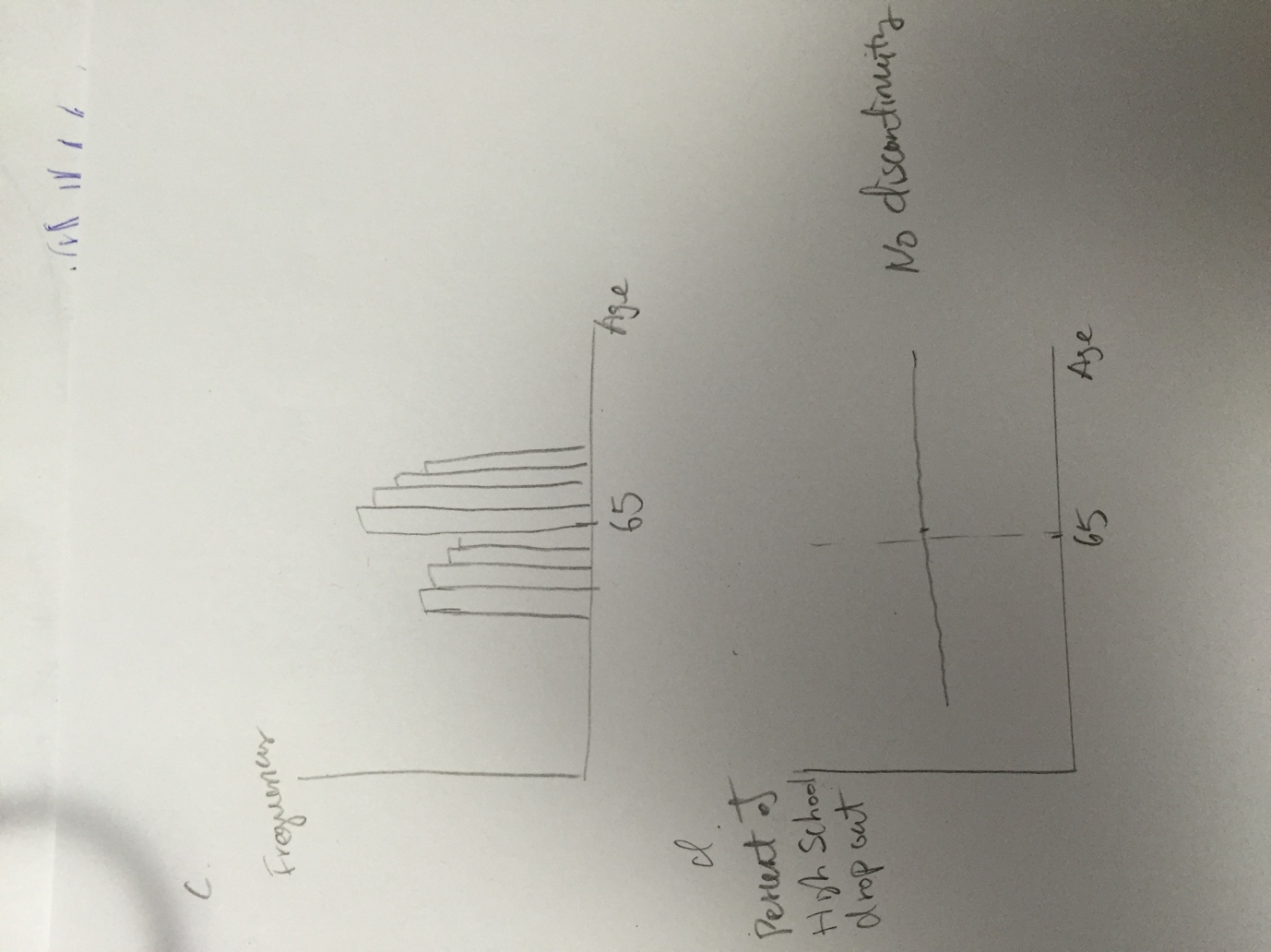
c. Draw a histogram to test the identification assumption under the condition that many 64 year olds lie about their age. (They would say they are 65 but in fact they are 64)



d. What does the histogram test in c) suggest about your identification assumption? Do you think RD is valid in this case?

**No. This indicates perfect manipulation around the cutoff. If those who lie about age also have worse health (They lie because they need more treatment). The difference in health of those above 65 and below 65 will be smaller than it should have been due to the fact that those above 65 are sicker than they should have been. This will create a downward bias.**

e. Draw a covariate test assuming the identification assumption is valid. Pick one covariate that you see fit



f. Is this an example of Sharp or Fuzzy RD? Explain.

**Fuzzy RD. Because the treatment variable-Medicare coverage is also determined by whether a person is receiving Social Security Disability Insurance.**

g. Specify the reduced form RD regression, make sure to label things correctly.

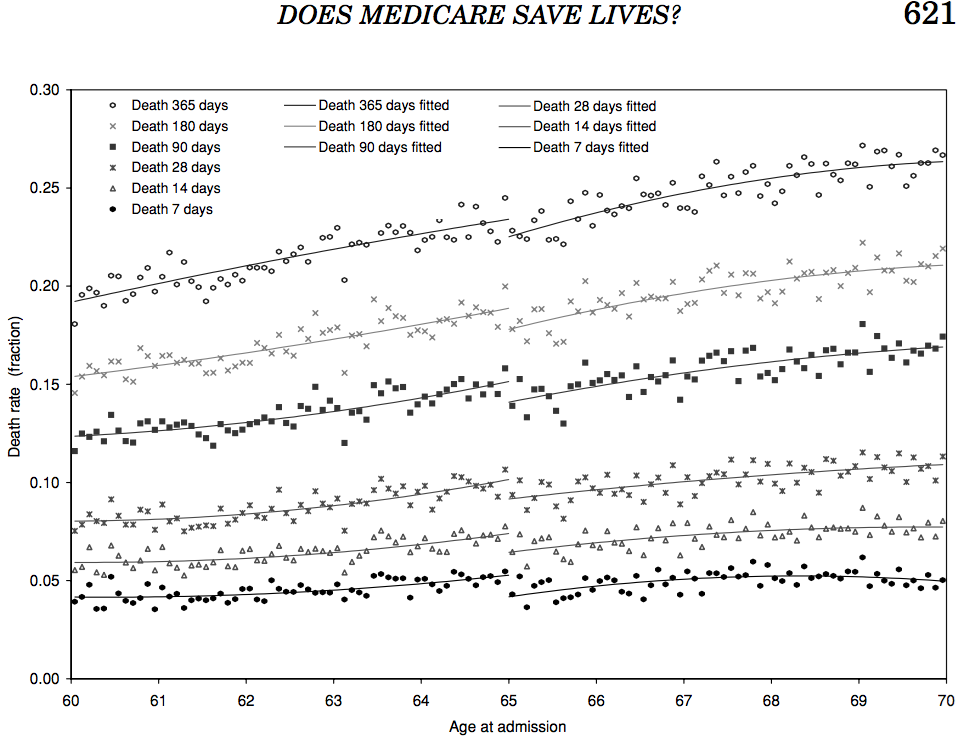
**Define S=age-65. Define above=1 if S>=0 and above=0 if S<0.**

**Health status=+above+S+(S\*above)+**

**is the causal estimate of getting over 65 on health status**

h. Look at the result graph below. What is your conclusion of the research?

**The research identified a decrease in death rate at the age 65. This is most likely due to the Medicare.**



**Part IV Quiet Study Areas- Regression Discontinuity [30 pts]**

In a one-time trial at a large urban university the school is offering a dedicated quiet study space for all individuals that live further than 15 miles from campus. If an individual lives 16 miles from campus they receive access to a study room in the library. If an individual lives 14 miles away from campus they do not have access to the study room. A regression discontinuity design can be used to measure the impact of having a study room on student’s GPA.

1. What is the running variable? Which individuals are treated? What is the cutoff for treatment?

**Running: Distance from campus. Individuals who live more than 15 miles away are treated. Cutoff is 15 miles.**

The RD regression estimates for a model with a window of within 5 miles of the cutoff is

1. What is the average GPA of non-treated people just near the cutoff?

**Value of running is almost zero near cut off => average GPA is 2.9 for people just to the left of cutoff.**

1. What is the average GPA of non-treated people just near the cutoff?

**Value of running is almost zero near cut off => average GPA is 2.9 + 0.08= 2.98 for people just to the right of cutoff.**

1. What is the impact of being assigned a study room on GPA?

**GPA goes up by 0.08.**

1. What is the assumption of a regression discontinuity design in this context?

**Students who live 16 miles away from campus are similar to students who live 14 miles away from campus on both observables and unobservables.**

**OR / AND**

**There is no manipulation of the cut-off.**

1. What is a way we can check if the assumptions of RD are likely to hold in this context?

**Histogram test to see if there is a jump in number of people around the cut-off AND/OR**

**Covariate Smoothness to check if people on either side of the cut-off are similar to each other.**