**Please provide me your R code Abdul.**

Vinish Shrestha

Abdullah K Almansour

EXAM (1)

ECON 339

11/7/2021

**Section 1**

Table

Description automatically generated

**Annual Income**

1. Plot a histogram for income in dataframe datainc.

Chart, histogram

Description automatically generated

2. In this question, we would want to merge two files **data** (birthweight file) and **datainc** (income file) by using person id. Note that each person has a unique id in both files and they can be merged together using *person\_id* variable. For example, if person\_id 1 refers to Maya in data file then person\_id of 1 will also refer to Maya in income file. We want to bring Maya’s birthweight observation and her income observation together. To do so use:

Table

Description automatically generated

3. Run a univariate regression of the form: birthweight = α + βmother\_education + ε. Report the coefficient on β. Does this show a causal relationship between mother’s education and infant health. **Note that you should delete all observations with mother’s education of 99 (not reported values) before you proceed with this.**

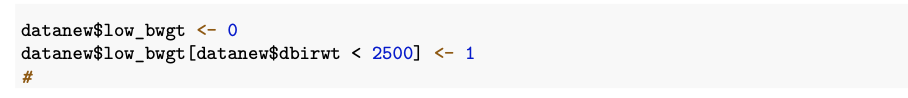
Text

Description automatically generated

Text

Description automatically generated

1. Next, generate an indicator variable for low birthweight called “low\_bwgt” if birthweight < 2,500 grams.



1. Calculate the proportion of infants with low birthweight by mother’s education. For instance, what proportion of infants belonging to mothers with 12 years of schooling have low birthweight. Do this for all education values.

Table

Description automatically generated with medium confidence

1. Using your calculations from 3 plot the relationship between mother’s years of schooling and proportion of infants with low birthweight. *hint:* type help(plot) in the console.

Chart, scatter chart

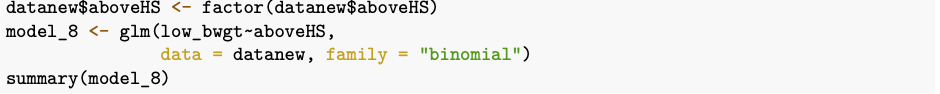
Description automatically generated

1. Now construct an indicator that represents mothers with more than high school education; “aboveHS” by using the following hint.



1. Now run the regression of the model specification: low\_birthweight = α + βabove\_HighSchool + ε.

Comment on the coefficient of β.



Text

Description automatically generatedText

Description automatically generated with low confidence

From the logit model, the dependent variable is represented as the log odds, ln(odds) = ln(p/(1-p)) = a\*x1 + b\*x2 + . + z\*xn. Since we have the aboveHS1 as a dummy variable, being aboveHS1 reduces the log odds of being under category low birth weight by 0.4045996.

1. Next, generate an indicator of race: black vs. white. Then run the following specification: race =

α + γabove\_HighSchool + ε. Comment about the point estimate of γˆ.

Text

Description automatically generated

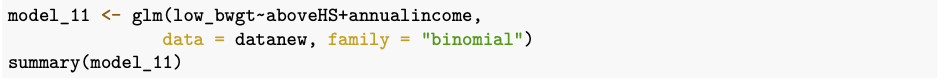
Since we have the aboveHS1 as a dummy variable, being aboveHS1 reduces the log odds of being under category race black by 0.2062407.

1. Using your realizations from 9, comment on the case of omitted variable bias when estimating

low\_birthweight = α + βabove\_HighSchool + ε in 6.

The proportion of having education past highschool can be used to influence the race type of the mother, but the education level of the mother can influence the low birth weight category in a similar direction for the log odds. As such there may be some sort of correlation between the race categorization and above high school relative to the low birth weight.

1. Run the regression specified as low\_birthweight = α + βabove\_HighSchool + κannualincome + ε. Compare the estimate of β with the specification low\_birthweight = α + βabove\_HighSchool + ε. Comment on which specification you’d prefer and why.



Text, letter

Description automatically generated

**Section 2: Lectures**

1. The demand for doctor visits can be written as:

doc visits = g(Insurance, Education, Income, Demographic, ε),

where ε contains unobserved factors that a researcher cannot perceive. One example is risk preference – as a researcher you cannot observe the risk preference of an individual. Now, assuming a linear functional form, you can express the demand for doctor visits as:

doc visits = α + βInsurance + γEducation + δincome + κDemographic + ε.

1. After estimating the regression, say you get β > 0. Does this mean that having an insurance increases

doctor visits compared to those without insurance? Explain.

The α estimate term is represents the difference between means of those with and those without insurance, as such the β is the means of those with insurance added onto mean of those without insurance which is the intercept, whenever other variables are kept constant. As such, we can make the assumption that having insurance can increase the number of doctor visits.

1. Based on part a. can you say that insurance has a causal effect on increasing doc;visits? Why or why not?

Not at all. The association between the number of doctor visits and insurance simply shows that the two variable values vary together. However, it may not necessarily imply that changes in having an insurance leads to changes in the the doctor visits.

1. Based on the evidence that insured individuals face lower price compared to uninsured and insured individuals are more likely to go to the doctor (β > 0), can you state that the demand for doctor visits is downward sloping? Why or why not?

Not at all. The association between the prices and the insurance uptake does not necessarily extend the correlation to the model that does not include the prices offered to the insured or not insured. Therefore, the inclusion may not influence the slopping, but with the β > 0, the demand for the doctors may be based purely on the value of the intercept term that acts as a reference.

This problem pertains to the randomized control trial. Say, you want to find out whether the demand for doctor visits is downward sloping among young adults. To do so, you do a lottery and randomly assign health insurance for 25 people, who are technically termed as the “treated group." The other 25 people who did not receive insurance are control group.

a. Consider the specification: visits = α + βT reat + ε.

Where visits represent the total number of doctor visits, Treat is the group that was assigned insurance through the lottery draw, and ε is the error term. Say, after estimating this specification, you find that β > 0. What does this suggest?

The β > 0 suggests that the exp(β) > 1, and as such the expected number of visits is exp(β) times higher as compared to those in the control group.

* 1. Does the estimate of β > 0 (in this case of randomization) suggest that there is a causal relationship between getting insured and increased doctor visits? Why or why not?

Not, it does not suggest causation. The association between the number of visits and being in the treated group of insurance simply shows that the two values vary together, but it may not necessarily depict that changes in having an insurance leads to changes in the the number of visits.

* 1. Now, in a different analysis based on the survey data, you find that those insured tend to have higher doctor visits. From this finding can you say that insurance leads to more doctor visits? Explain in resonance to your answer from part b.

The insured number of visits are based on the number of visits in the control group that act as the reference group. Away from causation, there may be an association, but it may not necessarily mean that there will be causation to that effect given that there may be instances where it may vary.

D. Briefly describe the Oregon Health Insurance experiment.

The Oregon health insurance experiment is one in which the effects of the coverage of Medicaid on health care and health care outcomes were assessed. The experiment offered enrolled was based on lottery selection for the treatment group, while the remaining control group was selected through the normal process. The findings with regard to attributes from the studies, of the participants such as their ages and their ethnicity were found to be vary much statistically similar. As such, factors that influence the criteria for selection such as the health status of the the participants were isolated and pointed out.

**Section 3: Readings**

1. Card et al. (2008) uses a natural experiment that uses eligibility criteria (age of 65) as an identification strategy to identify the effects of Medicare on mortality outcomes.
2. Explain in detail how the eligibility criteria set for Medicare is used as a quasi-natural experiment. Talk about the approach (method) they use.

A quasi-natural experiment involves treatment that is as a result of social and/or political factors that may include implementation of new or existing laws and programs around healthcare insurance. There is a predetermined criteria around the selection, for which eligibility is determined and the assignment is on purpose in order to seem random. In the study by Carlo et al. (2008), it is stated that eligibility for the Medicare coverage for individuals is set at 65 years or older, or for those less than 65 years being under the Social Security Disability Insurance program. Within the boundary age, changes in the health insurance offered are distinct, sudden and are on the rise for those individuals clocking the threshold age. The medicare hospital insurance is free from onset of 65 years for any individual for part A that alternates with the sharp rise in up-take while part B, covering doctor bills comes at an extra cost. . In their study about the effects of Medicare on mortality outcomes, there are two cases involved; one case is of the people aged sixty-five and older while the other case is of people who are below the cutoff score of age sixty-five. The patients above the cutoff of age sixty-five are assigned a different treatment, while those below age sixty-five are said not to be eligible for Medicare coverage.

1. The analysis of Card et al.’s study is based on hospital level data such that people in the sample are those who are admitted to the hospital. This creates a selection problem if we compare people who are right above the threshold vs. those who are right below the threshold. This is because if a 64 year old plays a waiting game and goes to the doctor when she is 65, she would have waited longer compared to a 64 year old with similar illness but who does not play the waiting game (i.e., goes to the doctor when 64). Hence, comparing the health outcomes of 64 year old individuals (without Medicare) vs. 65 year old individuls (with Medicare) might mean that we are comparing two groups that are systematically different in health stock to begin with. Describe how the researchers get around this selection problem. *(hint: hospital admissions are higher in week days compared to weekends.)*

While using the Regression Discontinuity approach of health outcomes, Card et al. (2008) encountered a challenge in their selection criteria. Comparisons of the health outcomes for the 64 year old individuals without Medicare versus the 65 year old individuals with Medicare are made easier with the inclusion of characteristics of patients who are admitted through the emergency department for severe conditions whose treatment and care required immediate hospitalization of the individuals. They identified diagnosis codes offered during emergency department admission with similar rates of admission during both weekends and weekdays, for which the mortality rates of patients were found to be same based on previous studies. The disregard for the age factor is made through the checks on discontinuities in the count of admissions at the age of 65 along with the patient characteristics for those above the 65 years and below the 65 years threshold. In addition, they used the estimations of the hospital admissions, which recorded high admissions during weekdays and low admissions during weekends. The regression discontinuity method has been glorified as being a crucial tool for researchers. However, this method is not immune to selection problems

* + 1. By using a figure that corresponds to the eligibility criteria (age), briefly describe the findings of their study on: i) insurance coverage, ii) quality of care, and iii) mortality outcomes.

The study findings are categorized into three important parts, including insurance coverage, quality of care, and mortality outcomes (Card et al. ,2009). First, the findings on insurance coverage is that the fraction of patients aged 65 with Medicare as their primary insurer increases by 47%, while the fraction of patients aged 65 or below with Medicare or private insurance covers records a drastic drop. Finally, the findings on mortality outcome estimated a decrease in patients mortality because patients aged 65 with Medicare insurance and had chronic conditions were assigned quality treatment than those without Medicare insurance covers, leading to a reduction of mortality.

After the age of 65, the Medicare is taken up by majority of the patients, while Private drops significantly to second place followed up by Medicaid. The fraction of patients who are uninsured and those with other insurances drops to almost zero over time depicting the higher uptake of Medicare by patients on attaining the age of 65.

d. What are some potential drawbacks of their study?

The study done by Card et al. (2008) is exposed to various drawbacks. First, the approach they used in their study may raise concerns about the study's external validity. The casual inference is another potential drawback of the study. Causal inference in the study may be jeopardized by systematized selection around the eligibility criteria used. An additional potential drawback of the study is that it may need huge databases to extract logical estimates. And also, the changes as a result of the uptake of Medicare is gauged solely on the rate of mortality which may not be a totally effective measure for the overall impact insurance has on the health and lifestyle of the general population

.