POLS0012 Causal Analysis: Tutorial Exercise 6

Question 1

In China, limited democracy exists in some local elections for low-level party officials, but people are often coerced into participation by their state-run employers. However, in some places voters have been permitted a degree of independence and autonomy. One such place is Peking University in Beijing, where students are generally free to choose whether or not to vote. In a paper from 2006, Mei Guan and Donald Green investigate whether the sorts of 'nudges' that work in a Western democratic context, such as canvassing from candidates and their supporters, will cause higher turnout in this more authoritarian setting.[1] They tested this by randomly assigning some of a sample of Beijing University students to receive a canvassing visit at their dorm room, with the others randomly assigned to receive no visit. They then measured whether each student subsequently voted. Here we'll investigate compliance and treatment effects. The data is contained in the file "guangreen.Rda" and includes the following variables for each student:

- *turnout* - =1 if the student voted, 0 otherwise
- *treat2* - =1 if the student was assigned to receive a canvassing visit, 0 otherwise
- *contact* - =1 if the student was contacted by a canvasser, 0 otherwise

a) Calculate (i) the number of students who were assigned to control but received the treatment and (ii) the number of students who were assigned to treatment but received the control. Using these answers, what type of non-compliance occurred in this experiment?
Code Hint: Try using the function `length()` to count the length of `contact` for particular subsets of people defined by square brackets

Code:

```
  # assigned to control, contacted
length(g$contact[g$contact==1&g$treat2==0])
  # assigned to treatment, not contacted
length(g$contact[g$contact==0&g$treat2==1])
```

Answer:
No students who were assigned to the control were contacted (treated), but 307 students who were assigned to the treatment were not contacted (i.e., received the control). Therefore this is a case of one-sided non-compliance.

[1] Guan and Green (2006). "Non-Coercive Mobilization in State-Controlled Elections: An Experimental Study in Beijing." *Comparative Political Studies* 39 (10): 1175-1193

b) In this experiment, what type of student is a complier?

Answer:
A complier is a student who always gets contacted when assigned to the treatment group and always does not get contacted when assigned to the control group

c) Calculate the proportion of compliers
Code Hint: Try using the function `sum()`

Code:

```
prop.c <- sum(g$contact)/sum(g$treat2)
```

Answer:
The proportion of compliers is 0.886

d) Calculate the intent-to-treat effect, and interpret what it means here

Code:

```
itt <- mean(g$turnout[g$treat2==1]) - mean(g$turnout[g$treat2==0])
```

Answer:
The intent-to-treat effect is 0.132. This means that on average, attempting to contact a student led to a 13.2 percentage point increase in their probability of voting.

e) Calculate the complier average treatment effect, and interpret what it means here

Code:

```
cace <- itt/prop.c
```

Answer:
The complier average causal effect is 0.149. This means that on average amongst the compliers, actually being contacted led to a 14.9 percentage point increase in the probability of voting

Question 2

In the example discussed at the end of today's class, a paper by Clingingsmith, Khwaja and Kremer (2009) examines the impact of the Hajj pilgrimage on the subsequent beliefs and values of Pakistani Muslims. They use a natural experiment – the random assignment of eligibility for a visa by lottery. You will need to load the `AER` package for this problem. The data is contained in the file "hajjdata.Rda" and includes the following key variables:

- *moderacy* - an index ranging from 0 to 4 constructed from opinion questions, where higher values indicate more moderate views on Islamic practices, Islamist terrorism, and the status of women[2]
- *success* - =1 if the respondent won the lottery for a Hajj visa, 0 otherwise
- *hajj2006* - =1 if the respondent went on the Hajj, 0 otherwise
- *age* - in years
- *literate* - =1 if respondent is literate, 0 otherwise
- *urban* - =1 if respondent lives in an urban area, 0 otherwise

a) Calculate (i) the *proportion* of people who won the lottery and did not go on the Hajj and (ii) the proportion of people who lost the lottery and went on the Hajj. Using these answers, what type of non-compliance occurred in this natural experiment?

Code:

```
# assigned to control, went on hajj
length(h$hajj2006[h$success==0&h$hajj2006==1])/length(h$hajj2006[h$success==0])
# assigned to treatment, didn't go on hajj
length(h$hajj2006[h$success==1&h$hajj2006==0])/length(h$hajj2006[h$success==1])
```

Answer:
13.7% of lottery losers went on the Hajj, while just 0.82% of lottery winners did not go. Therefore this is (just) two-sided non-compliance, but non-compliance was very rare amongst those assigned to treatment.

b) In this study, who are the compliers and who are the always-takers?

The compliers are people who always go on the Hajj when they win the visa lottery and always don't go on the Hajj when they lose it. The always-takers are people who always go on the Hajj regardless of their lottery outcome

c) Using a similar approach to Problem 1, calculate the intent-to-treat effect, proportion of compliers and complier average treatment effect, using moderacy as the outcome variable. Interpret the meaning of all three quantities in this natural experiment.
Hint: You will now need to include two terms when calculating the proportion of compliers, one for the encouraged group and one for the unencouraged group

Code:

[2]Specifically, it includes questions asking whether governments should force people to undertake Islamic religious practices, whether Osama Bin Laden was fighting for a just cause, and whether women should go to school and be allowed to work.

```
itt <- mean(h$moderacy[h$success==1]) - mean(h$moderacy[h$success==0])

prop.c <- sum(h$hajj2006[h$success==1])/length(h$hajj2006[h$success==1]) -
          sum(h$hajj2006[h$success==0])/length(h$hajj2006[h$success==0])

cace <- itt/prop.c
```

Answer:
The intent-to-treat effect is 0.107, meaning that winning the visa lottery caused a 0.107-point increase in moderacy along the four-point scale. The proportion of compliers is 0.854, meaning that 85.4% of people in this study are compliers. The CACE is 0.125, meaning that amongst the compliers, going on the Hajj causes an increase in moderacy of 0.125 points.

d) Now calculate the proportion of compliers using a bivariate regression of hajj2006 on success (the first stage of two-stage least squares). Verify that it is identical to the proportion you estimated in part (c).

Code:

```
summary(lm(h$hajj2006 ~ h$success))
```

e) Calculate the complier average treatment effect (CACE) using two-stage least squares and verify that your answer is identical to part (c). Report its standard error. Is the CACE statistically significant?
Code Hint: The code for two-stage least squares takes the form:

```
ivreg(outcome ~ treatment | encouragement)
```

Code:

```
library(AER)
summary(ivreg(moderacy ~ hajj2006 | success,data=h))
```

Answer:
As expected, the result is identical to (c). The standard error is 0.04, meaning that the CACE is statistically significant at all conventional significance levels

f) Add the covariates age, literate and urban to your model from (d). Does the estimated CACE change much? Why or why not?
Code Hint: Covariates need to be added to both parts of the `ivreg()` formula

Code:

```
summary(ivreg(moderacy ~ hajj2006+age+literate+urban |
success+age+literate+urban,data=h))
```

Answer:
The estimated CACE barely changes. This is not surprising, given that the instrument is randomly assigned.