POLS0012 Causal Analysis: Tutorial Exercise 3

Question 1

In this question, we study a field experiment on voting carried out in Benin in 2001 by Leonard Wantchekon. He investigated what types of campaign messages are most effective in increasing voter turnout in this developing African nation.[1] He persuaded three presidential campaigns to randomly use different types of campaign messages in different villages throughout the campaign. There were 16 villages in this part of the experiment, and randomisation was carried out by blocks. Villages were divided into blocks of 2 villages based on their geographic locations. Within each block, the two villages were assigned to one of two conditions:

- *Public Policy*: Wantchekon describes this treatment condition as: "It was decided that any public policy platform would raise issues pertaining to national unity and peace, eradicating corruption, alleviating poverty, developing agriculture and industry, protecting the rights of women and children, developing rural credit, providing access to the judicial system, protecting the environment, and/or fostering educational reforms."
- *Clientelist*: Wantchekon describes this treatment as: "A clientelist message, by contrast, would take the form of a specific promise to the village, for example, for government patronage jobs or local public goods, such as establishing a new local university or providing financial support for local fishermen or cotton producers."

The data is contained in the file "benin.Rda" and consists of:

- *block*: a variable indicating the block number
- *reg.voters*: a background covariate, the number of registered voters in the village
- *vote.pop*: the outcome variable, the proportion of village that voted
- *treatment*: =1 if in the "clientelist" condition, 0 otherwise

a) Using a suitable regression, perform a balance test with the *reg.voters* variable between villages in the two treatment conditions (ignoring the blocking). What do you conclude?

Code:

```
#regress treatment on the covariate
summary(lm(treatment ~ reg.voters,data=b))
```

[1] For the original paper, see http://www.nyu.edu/gsas/dept/politics/faculty/wantchekon/research/WP_0331.pdf

Answer:
There is no evidence of imbalance between villages in the two treatment conditions, based on this covariate. The coefficient is extremely close to zero and not statistically significant, as we would expect.

b) Estimate the average treatment effect and its standard error, using the difference in means estimator (ignoring the blocking). What does the ATE tell you? Is it significant at the 5% level?

Code:

```
summary(lm(vote.pop~treatment,data=b))

# equivalently (but doesn't give standard error):
t.test(b$vote.pop[b$treatment==1],b$vote.pop[b$treatment==0])
```

Answer:
Note that a regression is the easiest way to obtain a standard error. The t-test gives the same answer but doesn't report the standard error. The ATE suggests that using clientelist rather than public policy messages boosts turnout by 15.75 percentage points. However, this effect is not significant at the 5% level (p=0.1).

c) Using a suitable regression, estimate the same average treatment effect while controlling for block membership. Are there any differences in the results compared to part (b)? Why or why not?

Code Hints:
- Linear regression can be used to calculate Average Treatment Effects
- `factor()` can be used to turn a categorical variable into a full set of dummy variables , 1 for each category

Code:

```
summary(lm(vote.pop ~ treatment + factor(block), data=b))
```

Answer:
The ATE is unchanged, as we would expect. However, the standard error is substantially reduced, and the effect is now significant at the 5% level. This illustrates how useful blocking is in small samples like this one. It is like a 'free lunch.' We still get the correct coefficient, now estimated with greater precision.

NB: Here, you needed to control for block membership using a dummy variable for each block. The block numbers (1-8) are not meaningful in their own right

Question 2

For this question, we will examine the original data from the Tennessee STAR experiment. Recall that in the original study, within schools students were randomly assigned to "small", "regular" or "regular plus aide" classes for four years. To keep things simple, the dataset for this problem only contains students in the "small classes" or "regular classes" conditions for kindergarten and 1st grade only (the first two years of the study). Remember that a key problem with the experiment was that some children left the study early, and others did not comply with their assignments to treatment and control. In this problem we are going to assess both attrition and non-compliance.

The data is contained in the file "star.Rda" and includes the following variables for each student:

- *gkclasstype* - Class type enrolled in kindergarten, the beginning of the study
- *g1classtype* - Class type enrolled in first grade
- *gender* - gender (1 if female, 0 if male)
- *race* - race (1 if black, 0 if not black)
- *gkfreelunch* - gender (1 if child qualifies for free lunches, 0 if not). This is often used a proxy for family poverty
- *gktreadss* - SAT test score in reading at end of kindergarten
- *gktmathss* - SAT test score in maths at end of kindergarten
- *g1treadss* - SAT test score in reading at end of first grade
- *g1tmathss* - SAT test score in maths at end of first grade
- *gkschid* - Unique code identifying each school in the study

a) Create a variable called "treat" that equals 1 if a child was assigned to a small class in kindergarten and 0 otherwise

   Code:

```
s$treat <- ifelse(s$gkclasstype=="SMALL CLASS",1,0)
```

b) Using t tests, obtain p-values to assess the null hypotheses of no imbalance between the "small class" and "regular class" groups in terms of gender, race or free lunches. What do you conclude about balance between the two groups?

   Code:

```
t.test(s$gender[s$treat==1],s$gender[s$treat==0])
t.test(s$race[s$treat==1],s$race[s$treat==0])
t.test(s$gkfreelunch[s$treat==1],s$gkfreelunch[s$treat==0])
```

Answer:

Table 1: **P-Values for Balance Tests between the Two Class Sizes**

Variable	p-value
Gender	0.85
Race	0.01
Free Lunch	0.12

For gender and free lunch, we cannot reject the null hypothesis of no difference between the two groups of students. There is no evidence for failures of randomisation in either case. For race, children in small classes were slightly less likely to be black (31.6% versus 35.8%), a difference that is statistically significant. This could be a source of concern if race is also correlated with educational achievement (e.g. because black families in the US are likely to have lower incomes), because the children in small classes may be higher-achieving on average, even before the experiment began. This may have occurred of white parents were 'pushier' and worked harder to get their children's assignment changed to a small class.

c) Calculate Average Treatment Effects for both maths and reading scores in kindergarten, for children in small classes compared to regular classes. Estimate results with and without controlling for the school attended. Interpret the results, including their statistical significance

Code:

```
ate.reading1 <- lm(gktreadss ~ treat,data=s)
summary(ate.reading1)$coef[2,]
ate.reading2 <- lm(gktreadss ~ treat + factor(gkschid),data=s)
summary(ate.reading2)$coef[2,]

ate.maths1 <- lm(gktmathss ~ treat,data=s)
summary(ate.maths1)$coef[2,]
ate.maths2 <- lm(gktmathss ~ treat + factor(gkschid),data=s)
summary(ate.maths2)$coef[2,]
```

Answer:

The results show that being in a small class compared to a regular class leads to an 8.44-unit increase in reading score and an 11.77-unit increase in maths score, when controlling for

Table 2: **ATEs and Standard Errors for being in small class compared to regular class in kindergarten**

	Uncontrolled		Controlling for School	
Outcome	ATE	Standard Error	ATE	Standard Error
Reading Score	8.04	1.15	8.44	1.07
Maths Score	11.4	1.76	11.77	1.62

school. The results are slightly larger when controlling for school. All results are significant at all conventional significance levels.

d) Calculate the proportion of children who left the experiment (i.e., attrition) between kindergarten and first grade, accounting for both children who are missing from the sample in first grade altogether, and children who did not report outcome data in first grade.

Code Hints:

- Start by creating an indicator variable equalling 1 if a child was missing and 0 otherwise
- `is.na()` selects missing observations of a variable

Code:

```
s$missing <- ifelse( is.na(s$g1classtype) | is.na(s$g1treadss) |
is.na(s$g1tmathss) ,1,0)
sum(s$missing) / length(s$missing)

s$missing <- ifelse( (is.na(s$g1classtype)) | (is.na(s$g1treadss) &
is.na(s$g1tmathss)) ,1,0)
sum(s$missing) / length(s$missing)
```

Answer:
Attrition was very high between kindergarten and first grade, although the answer will differ slightly depending on exactly how you defined attrition here. Defining attrition as anyone who did not take part in first grade or didn't report one of maths or reading scores (the top piece of code), it was 37.1%. Or, if we define it as as anyone who did not take part in first grade or didn't report both maths *and* reading scores (the bottom piece of code), it was 35.4%

e) Assess the extent of non-compliance amongst children in first grade by calculating:

i) The proportion of children who were assigned to regular classes in kindergarten that were enrolled in small classes in first grade

ii) The proportion of children who were assigned to small classes in kindergarten that were enrolled in regular classes in first grade

Hint: You need to exclude missing responses in first grade

What type of non-compliance do we have here? Did non-compliance differ by class assignment in kindergarten?

```
length(s$g1classtype[s$g1classtype=="SMALL CLASS" & s$gkclasstype=="REGULAR CLASS" &
                       !is.na(s$g1classtype)]) /
  length(s$gkclasstype[s$gkclasstype=="REGULAR CLASS"])

length(s$g1classtype[s$g1classtype=="REGULAR CLASS" & s$gkclasstype=="SMALL CLASS" &
                       !is.na(s$g1classtype)]) /
  length(s$gkclasstype[s$gkclasstype=="SMALL CLASS"])
```

Answer:

This is two-sided non-compliance: 8.2% of children initially assigned to regular classes were in small classes in first grade, while 3.2% of children initially assigned to small classes were in regular classes in first grade. The pattern is not surprising: since small classes were more desirable, parents were more likely to try to push their children into smaller classes than the other way around.

f) Assuming missingness-at-random, calculate ATEs and standard errors for maths and reading scores in the first grade, controlling for school attended. Interpret your results
Code Hint: Note that the `lm()` function automatically drops missing values that are labelled as "NA" in R

Code:

```
s$treat2 <- ifelse(s$g1classtype=="SMALL CLASS",1,0)

ate.reading.p2 <- lm(g1treadss ~ treat2 + factor(gkschid),data=s)
summary(ate.reading.p2)$coef[2,]

ate.maths.p2 <- lm(g1tmathss ~ treat2 + factor(gkschid),data=s)
summary(ate.maths.p2)$coef[2,]
```

Answer:

The results show that being in a small class compared to a regular class in first grade leads to a 16.6-unit increase in reading score and a 12.4-unit increase in maths score. Both results are significant at all conventional significance levels.

Table 3: **ATEs and Standard Errors for being in small class compared to regular class in kindergarten, controlling for school**

Outcome	ATE	Standard Error
Reading Score	16.5	2.3
Maths Score	12.4	1.8

g) **[Challenging Question]** Calculate extreme-value bounds for the ATEs in part (e), using the highest and lowest observed test scores. Start by filling in missing values for class types in first grade by assuming that they would have been the same as in kindergarten. Do these bounds suggest that attrition between kindergarten and first grade threatens the validity of the ATEs calculated in (f), or not?

Code:

```
 ## Upper bound ##
   # start by creating a new dataset with new treatment variable, filling in missing valu
 s.upper <- s
 s.upper$treat3 <- ifelse(is.na(s.upper$treat2),s.upper$treat,s.upper$treat2)

   # assign max score to treated units, min score to control units
 for(i in 1:length(s.upper$treat3)){
   if(is.na(s.upper$g1treadss[i]) & s.upper$treat3[i]==0){
     s.upper$g1treadss[i] <- min(na.omit(s$g1treadss))
     s.upper$g1tmathss[i] <- min(na.omit(s$g1tmathss))
   }

   else if(is.na(s.upper$g1treadss[i]) & s.upper$treat3[i]==1){
     s.upper$g1treadss[i] <- max(na.omit(s$g1treadss))
     s.upper$g1tmathss[i] <- max(na.omit(s$g1tmathss))
   }
 }

# new ATEs
 ate.reading.upper <- lm(g1treadss ~ treat3 + factor(gkschid),data=s.upper)
 summary(ate.reading.upper)$coef[2,]

 ate.maths.upper <- lm(g1tmathss ~ treat3 + factor(gkschid),data=s.upper)
 summary(ate.maths.upper)$coef[2,]
```

```
  ## Lower bound ##
    # start by creating a new dataset with new treatment variable, filling in missing valu
  s.lower <- s
  s.lower$treat3 <- ifelse(is.na(s.lower$treat2),s.lower$treat,s.lower$treat2)

    # assign min score to treated units, max score to control units
  for(i in 1:length(s.lower$treat3)){
    if(is.na(s.lower$g1treadss[i]) & s.lower$treat3[i]==0){
      s.lower$g1treadss[i] <- max(na.omit(s$g1treadss))
      s.lower$g1tmathss[i] <- max(na.omit(s$g1tmathss))
    }

    else if(is.na(s.lower$g1treadss[i]) & s.lower$treat3[i]==1){
      s.lower$g1treadss[i] <- min(na.omit(s$g1treadss))
      s.lower$g1tmathss[i] <- min(na.omit(s$g1tmathss))
    }
  }

# new ATEs
  ate.reading.lower <- lm(g1treadss ~ treat3 + factor(gkschid),data=s.lower)
  summary(ate.reading.lower)$coef[2,]

  ate.maths.lower <- lm(g1tmathss ~ treat3 + factor(gkschid),data=s.lower)
  summary(ate.maths.lower)$coef[2,]
```

Answer:
For reading, this gives bounds of $[-82.2, 98.9]$ and for maths $[-91.5, 101.4]$. These intervals contain the ATEs estimated in part (f) and are very wide, encompassing negative treatment effects. This suggests that attrition could threaten the validity of the ATEs calculated in part (f). Nonetheless it is worth remembering that these intervals represent extreme scenarios. In reality, attrition is unlikely to be as extreme as this.

h) **[Challenging Question]** Briefly, explain (using words only) why the two sets of standard errors obtained in part (c) differ depending on whether school is controlled for

Answer:
Randomisation was carried out within each school rather than across schools, so this is akin to a case of block randomisation. As we would expect, schools are very predictive of the outcomes (some schools are better than others), so controlling for school leads to an increase in the precision of the estimated ATEs (i.e., lower standard errors).