POLS0012 Causal Analysis: Tutorial Exercise 2

Question 1

In this question we will demonstrate that randomised experiments work by creating an imaginary experiment, using the dataset "a" contained in the file "experiment.Rda." It includes the potential outcome under control ($y0$) and the potential outcome under treatment ($y1$) for 100 units that form the sample for our experiment. Note that this is a purely hypothetical scenario. In reality, we never observe potential outcomes under both treatment and control for the same units: we only observe one of them (the fundamental problem of causal inference). By creating a randomised experiment with this dataset, we'll demonstrate that experiments overcome the fundamental problem.

a) Find the true Average Treatment Effect for all units, using $y0$ and $y1$

   Code:

   ```
   mean(a$y1-a$y0)
   ```

   The true ATE is 20.9952

b) Now, we'll randomly assign half of the units to treatment and half to control by creating a new variable indicating treatment status. As described in the lecture, we can do this using the following steps:[1]

   i) Assign each unit a random number from 1 to 100: create a new column in the dataset named *rand*, using the `sample()` command and a vector of the numbers 1 to 100
   ii) Re-order the dataset from lowest to highest value of *rand* using the code `a <- a[order(a$rand),]`
   iii) Create a treatment variable named *tr* that equals 1 for the first 50 units and 0 for the second 50 using the code `c(rep(1,50),rep(0,50))`

   Code:

   ```
   set.seed(1)
   a$rand <- sample(c(1:100))
   a <- a[order(a$rand),]
   a$tr <- c(rep(1,50),rep(0,50))
   ```

c) Conduct a test to assess whether the treatment and control groups have the same average potential outcomes under control ($y0$). Has randomisation succeeded in creating treatment and control groups with equivalent potential outcomes under control?

   Code:

[1] First, input the command `set.seed(1)` so that your results are the same as the solutions

```
t.test(a$y0[a$tr==1], a$y0[a$tr==0])
```

There is a small, but not statistically significant, difference in potential outcomes under control between the two groups. Remember that randomisation isn't guaranteed to equalise potential outcomes in any one instance. Instead, it does so in expectation over many repeated randomisations, as we show below.

d) Find the Average Treatment Effect from the experiment. How similar is it to the true Average Treatment Effect?

Code:

```
mean(a$y1[a$tr==1]) - mean(a$y0[a$tr==0])
```

The estimated ATE is 19.43, which is close to the true ATE but not exactly the same.

e) Now, let's see how our experimental procedure performs over repeated randomisations, using a simulation:

   i) First, create a function that takes in the dataset a, carries out a randomised experiment, and reports the ATE estimated in (d). You can do this by combining code from (b) and (d), without using `set.seed`
   Code Hint: Remember that a function in R has the following format:
   `function.name <- function(input) { function commands }`

   ii) Second, find the results of 10,000 randomised experiments by running the function 10,000 times and storing the results in a variable.
   Code Hint: You can do this using the `replicate()` command, with two arguments: the first is the number of simulations and the second is the function being evaluated

What is the mean ATE from your 10,000 experiments? Does this suggest that the experimental procedure is unbiased?

Code:

```
  experiment.sim <- function(a){
  a$rand <- sample(c(1:100))
  a <- a[order(a$rand),]
  a$tr <- c(rep(1,50),rep(0,50))

  mean(a$y1[a$tr==1]) - mean(a$y0[a$tr==0])
}

sims <- replicate(10000,experiment.sim(a))
mean(sims)
```

The estimated ATEs will be slightly different for everyone due to randomness, but you should find that the mean ATE is now extremely close to the true ATE. This shows that randomised experiments work! On average across repeated randomisations, we obtain the true ATE. In any one instance, the estimate will not be exactly the same, but the estimator is unbiased because it recovers the true ATE in expectation.

f) Finally, repeat (e), calculating the mean difference in potential outcomes under control (y0) between the treatment and control groups instead of the ATE. What is the mean difference from your 10,000 experiments?

Code:

```
experiment.sim2 <- function(a){
  a$rand <- sample(c(1:100))
  a <- a[order(a$rand),]
  a$tr <- c(rep(1,50),rep(0,50))

  mean(a$y0[a$tr==1] - a$y0[a$tr==0])
}

sims2 <- replicate(10000,experiment.sim2(a))
mean(sims2)
```

Again, the estimated differences will be slightly different for everyone due to randomness, but you should find that the mean difference is now extremely close to zero. This demonstrates that on average, experiments remove selection bias. Again, in any one instance the treatment and control groups will not be exactly alike, but across repeated randomisations, their potential outcomes are identical.

Question 2

Why do people bother to vote? One hypothesis is adherence to social norms. Voting is widely regarded as a civic duty and people worry that others will think badly of them if they fail to participate. According to this theory, voters may receive two different types of utility from voting; (a) the intrinsic rewards from performing this duty and (b) the extrinsic rewards received when others observe them doing so. To gauge the effects of priming intrinsic motives and applying varying degrees of extrinsic pressure on voting behaviour, Gerber, Green, and Larimer conducted a famous field experiment in Michigan prior to the August 2006 primary election.[2] The sample for

[2] For the original paper, see: `//isps.yale.edu/sites/default/files/publication/2012/12/ISPS08-001.pdf`

the experiment was 344,084 voters. They were randomly assigned to either the control group or one of four treatment groups.

We'll practice analysing experiments by focusing on two of their treatments. The first treatment, "civic duty", incolved sending a letter to the voter carrying the message "DO YOUR CIVIC DUTY - VOTE!". The second treatment, "Neighbors" sent the same letter, but also informed the voter that who votes is public information (which is the case by law in the USA). It listed the recent voting record of each registered voter in the household and the voting records of those living nearby, and stated that a follow-up mailing after the election would report back to the household and to their neighbours on who had voted and who had not. The idea was to see whether priming extrinsic motivations would encourage this treatment group to turn out more than the control group. The control group received no letter. For this question we'll use the original data of Gerber et al, contained in the file gerber.Rda. Below is a list of the variable definitions:

- *sex* - gender (1 if female, 0 if male)
- *yob* - year of birth
- *p2004* = 1 if Respondent voted in the 2004 Primary Election, 0 otherwise
- *voting* = 1 if Respondent voted in the 2006 Primary Election, 0 otherwise [the outcome variable]
- *control* = 1 if Respondent is assigned to the control group, 0 otherwise
- *civicduty* = 1 if Respondent is assigned to the "Civic Duty" group, 0 otherwise
- *neighbors* = 1 if Respondent is assigned to the "Neighbors" group, 0 otherwise)

a) For both treatments, calculate the average treatment effect and test whether it is statistically significant. Interpret the results, giving a precise explanation of the magnitude of the treatment effects. What do they suggest about the motivations that people have for voting?

   Code:

```
mean(g$voting[g$civicduty==1]) - mean(g$voting[g$control==1])
t.test(g$voting[g$civicduty==1], g$voting[g$control==1])

mean(g$voting[g$neighbors==1]) - mean(g$voting[g$control==1])
t.test(g$voting[g$neighbors==1], g$voting[g$control==1])
```

   The ATE for the civic duty treatment is 0.018 and the ATE for the neighbors treatment is 0.081. They have t statistics of 6.9 and 30.2 respectively, meaning that they are both statistically significant at all conventional significance levels. Because the outcome variable is binary, the ATEs just tell us the differences in the proportion of people in each group who voted (recall from last year that the mean of a binary variable is the same as the proportion of its observations that equal 1). This means that 1.8 percentage points more people voted in the "civic duty" group compared to the control group and 8.1 percentage points more people voted in the "neighbors" group compared to the control group.

The results support the idea that people have both intrinsic motivations to vote (since "civic duty" led to higher voting) *and* even stronger extrinsic motivations to vote (since public shaming led to even higher voting).

b) For both treatment groups, compare the mean values of *yob*, *sex* and *p2004* to the control group. Do the results suggest that randomisation was successful? Is selection bias likely to be a problem in this experiment?

Code:

```
t.test(g$sex[g$civicduty==1],g$sex[g$control==1])
t.test(g$yob[g$civicduty==1],g$yob[g$control==1])
t.test(g$p2004[g$civicduty==1],g$p2004[g$control==1])

t.test(g$sex[g$neighbors==1],g$sex[g$control==1])
t.test(g$yob[g$neighbors==1],g$yob[g$control==1])
t.test(g$p2004[g$neighbors==1],g$p2004[g$control==1])
```

In all cases, the difference between the treatment group and control group is tiny. This suggests that randomisation was successful. Under true randomisation the control group should provide a valid counterfactual for the treatment group, which means that the average characteristics of the two groups should be virtually identical. This also means that selection bias is unlikely to be a problem in this experiment. If randomisation was successful, there should be zero selection bias.

However, it is worth pointing out three caveats. First, although the differences are very tiny for the *p2004* variable, the difference is statistically significant. As always in statistics, it is important to consider the size of an effect as well as its statistical significance. Here the dataset is huge, making it very easy for even tiny differences to be statistically significant. Second, if there had been large differences between the treatment and control groups, that would not necessarily mean that there was a failure of randomisation. It is perfectly possible for such differences to emerge due to chance alone. If you did find large differences in your own experiment, it would certainly be sensible to check whether an error was made in randomisation. Third, just because we found no difference in terms of observed variables, there could still be selection bias from unobserved variables. Nonetheless, such a case is unlikely.

c) Calculate the ATE for the the *neighbors* treatment using:

i) A regression containing only *neighbors*

ii) A regression containing *neighbors* and the three background characteristics

Note that you will need to subset your data appropriately in order to obtain the correct control group. Are there any big differences in the estimated ATE between the two specifications? Or between these two estimates and the result from part (a)? Why or why not?

Code:

```
# Subset to only neighbors and control observations
g.reg <- g[g$neighbors==1|g$control==1,]

#i)
summary(lm(voting ~ neighbors,data=g.reg))

#ii)
summary(lm(voting ~ neighbors + sex + yob + p2004,data=g.reg))
```

In all cases, the results are virtually identical to the answer obtained in part (a) using a simple difference-in-means, although the standard errors are very slightly smaller. This is not surprising. In part (i) as we saw in week 1, a simple regression with only the treatment variable is numerically exactly identical to a difference-in-means. In part (ii), we would not expect adding in controls for background covariates to make much difference, since they are uncorrelated with the treatment variable (even though they are strongly correlated with the outcome). It is always the case in regression analysis that we can safely omit any variable that is uncorrelated with either the outcome or the existing independent variables.