1. Noncompliance in Recycling Experiment

Suppose that you want to conduct a study of recycling behavior. A number of undergraduate students are hired to walk door to door and provide information about the benefits of recycling to people in the treatment group. Here are some facts about how the experiment was actually carried out.

- 1,500 households are assigned to the treatment group.
- The undergrads tell you that they successfully managed to contact 700 households.
- The control group had 3,000 households (not contacted by any undergraduate students).
- The subsequent recycling rates (i.e. the outcome variable) are computed and you find that 500 households in the treatment group recycled. In the control group, 600 households recycled.

```
library(data.table)
d <- data.table(id = 1:4500)
d[, assigned := ifelse(id <=1500, 1, 0)]
d[, treated := ifelse(id <=700, 1, 0)]
d[, recycled := ifelse(id <=500, 1, ifelse(id>=3900, 1, 0))]

head(d)
```

```
##    id assigned treated recycled
## 1:  1        1       1        1
## 2:  2        1       1        1
## 3:  3        1       1        1
## 4:  4        1       1        1
## 5:  5        1       1        1
## 6:  6        1       1        1
```

1. What is the ITT? Do the work to compute it, and store it into the object `recycling_itt`.

```
# 1st way to calculate ITT (via regression)
recycling_itt <- coef(d[, lm(recycled ~ assigned)])[2]
recycling_itt
```

```
## assigned
##    0.133
```

```
# 2nd way to calculate ITT (via mean of treatment/control)
recycling_itt = d[assigned ==1, mean(recycled)] - d[assigned ==0, mean(recycled)]
recycling_itt
```

```
## [1] 0.133
```

We show two ways to compute ITT which show the results are identical to each other at: 0.133

2. What is the CACE? Do the work to compute it, and store it into the object `recycling_cace`.

```
# calculating ITT_d below
recycling_itt_d <- coef(d[, lm(treated ~ assigned)])[2]
recycling_itt_d
```

```
##  assigned
## 0.4666667
```

```
# calculating CACE below as a fraction of ITT and ITT_d
recycling_cace <- recycling_itt / recycling_itt_d
recycling_cace
```

```
## assigned
##    0.285
```

CACE is computed as: 0.285

There appear to be some inconsistencies regarding how the undergraduates actually carried out the instructions they were given.

- One of the students, Mike, tells you that they actually lied about the the number of contacted treatment households and that the true number was 500.
- Another student, Andy, tells you that the true number was actually 600.

3. What is the CACE if Mike is correct?

```
#We first calculate the ITT_d for the number of households which Mike contacted
d[, treated_mike := ifelse(id <=500, 1, 0)]
recycling_itt_d_mike <- coef(d[, lm(treated_mike ~ assigned)])[2]
recycling_itt_d_mike
```

```
##  assigned
## 0.3333333
```

```
cace_mike <- recycling_itt / recycling_itt_d_mike
cace_mike
```

```
## assigned
##    0.399
```

We see that the CACE computation for Mike's contacted households comes to: 0.399

4. What is the CACE if Andy is correct?

```
d[, treated_andy := ifelse(id <=600, 1, 0)]
recycling_itt_d_andy <- coef(d[, lm(treated_andy ~ assigned)])[2]
recycling_itt_d_andy
```

```
## assigned
##      0.4
```

```
cace_andy <- recycling_itt / recycling_itt_d_andy
cace_andy
```

```
## assigned
##   0.3325
```

We see that the CACE computation for Andy's dataset is: 0.3325

For the rest of this question, suppose that **in fact** Mike was telling the truth.

5. What was the impact of the undergraduates's false reporting on our estimates of the treatment's effectiveness?

We see that as per dataset given by Mike, CACE calculated from that (Mike's) dataset was larger (0.399 vs. 0.285). Hence if Mike was correct, we would have underestimated the treatment effect in our earlier estimates.

6. Does your answer change depending on whether you choose to focus on the ITT or the CACE?

Yes our answer changes depending on what metric (ITT or CACE) we focus on. We see that ITT is not impacted by the take-out rate. However, CACE is a fraction of ITT and ITT_d (a.k.a take-out rates). And since ITT_d is impacted by the number of compliers in experimental studies, so does CACE. Therefore, we do see that CACE gets impacted by the number of people complying in the experimentation. > yes, ITT is not affected by take out rate, but CACE does get impacted by number of people.