Section 6

What are we doing in section today?

Compliance with R and an Attrition Example

Packages needed

```
library(tidyverse)
library(estimatr)
```

Example

"Guan and Green report the results of a canvassing experiment conducted in Beijing on the eve of a local election. Students on the campus of Peking University were randomly assigned to treatment or control groups. Canvassers attempted to contact students in their dorm rooms and encourage them to vote. No contact with the control group was attempted. Of the 2,688 students assigned to the treatment group, 2,380 were contacted. A total of 2,152 students in the treatment group voted; of the 1,334 students assigned to the control group, 892 voted. One aspect of this experiment threatens to violate the exclusion restriction. At every dorm room they visited, even those where no one answered, canvassers left a leaflet encouraging students to vote"

Dataset

```
data <- read_csv(here("~/Desktop/section6.csv"))%>%
  mutate(cluster = as.character(cluster))
head(data)
```

```
## # A tibble: 6 × 4
##       Y     D cluster     Z
##   <dbl> <dbl> <chr>   <dbl>
## 1     0     0 1010101     0
## 2     0     0 1010101     0
## 3     0     0 1010101     0
## 4     0     0 1010102     0
## 5     0     0 1010102     0
## 6     0     1 1010103     1
```

Estimate the ITT

What is the ITT?

What variables do we need for the ITT in this dataset?

Estimate the ITT

```
itt <- data %>%
  group_by(Z)%>%
  summarise(across(Y, mean))%>%
  mutate(Z = if_else(Z == 1, "T", "C"))%>%
  pivot_wider(names_from = Z, values_from = Y)%>%
  mutate(ITT = T-C)%>%
  pull(ITT)
itt
```

```
## [1] 0.1319296
```

CACE Estimation

Assume the leaflet has no effect on turnout. What is the CACE?

```
# Use a ratio estimate
ittd <- data %>%
  group_by(Z)%>%
  summarise(across(D, mean))%>%
  mutate(Z = if_else(Z == 1, "T", "C"))%>%
  pivot_wider(names_from = Z, values_from = D)%>%
  mutate(ITT_D = T-C)%>%
  pull()
ittd
```

```
## [1] 0.8857887
```

CACE Ratio Estimator

```
cace <- itt/ittd
cace
```

```
## [1] 0.1489402
```

CACE Estimation

Assume the leaflet has no effect on turnout. What is the CACE?

```
## Use an IV estimator
cace <- iv_robust(Y~D|Z, data = data, cluster = cluster)%>%
  tidy()%>%
  filter(term == "D")%>%
  select(estimate)%>%
  pull()
cace
```

```
## [1] 0.1489402
```

Attrition and IPW

```
data_a <- read_csv(here("~/Desktop/attritonEx.csv"))
head(data_a)
```

```
## # A tibble: 6 × 15
##       D   sex match_i match_ic match_i7   age phone checkid  math  read
##   <dbl> <dbl>   <dbl>    <dbl>    <dbl> <dbl> <dbl>   <dbl> <dbl> <dbl>
## 1     0     0       0        0        0    10     1       1     0     0
## 2     0     0       1        1        1    10     0       1    39    54
## 3     0     0       1        1        1    10     0       1    42    41
## 4     0     0       1        1        1    10     1       1    45    52
## 5     0     0       0        0        0    10     1       1     0     0
## 6     0     0       1        1        1    10     1       1    50    54
## # … with 5 more variables: match_ic7 <dbl>, readcens1 <dbl>, readcens10 <dbl>,
## #   mathcens1 <dbl>, mathcens10 <dbl>
```

Step 1: Model of missingness

```
modelData <- data_a %>%
  mutate(observed = 1 - (read == 0),
         probobs = glm(observed~(D*sex)+(D*phone)+(D*age),
                       family = binomial(link = "logit"))$fitted,
         weights = 1/probobs)%>%
  select(D, read, sex, observed, phone, age, probobs, weights)
```

Step 1: Model of missingness

D	read	sex	observed	phone	age	probobs	weights
0	0	0	0	1	10	0.8332338	1.200143
0	54	0	1	0	10	0.8875910	1.126645
0	41	0	1	0	10	0.8875910	1.126645
0	52	0	1	1	10	0.8332338	1.200143
0	0	0	0	1	10	0.8332338	1.200143
0	54	0	1	1	10	0.8332338	1.200143

Step 2: Run a weighted regression

```
weightedModel <- lm_robust(read ~ D,
                           data = modelData%>%
                             filter(observed==1),
                           weights = weights)%>%
  tidy()%>%
  mutate(model = "w")
```

Step 3: Compare to unweighted Regression

```
unweightedModel <- lm_robust(read~D,
                             data = modelData %>%
                               filter(observed ==1))%>%
  tidy()%>%
  mutate(model = "u")
```

Step 4: Coefficient Plots

```
cpData <- bind_rows(weightedModel, unweightedModel)%>%
  filter(term == "D")%>%
  ggplot(., aes(x = estimate, y = model))+
  geom_point()+
  geom_errorbarh(aes(xmin = conf.low,
                     xmax = conf.high,
                     height = 0.1))+
  theme_xaringan()+
  geom_vline(xintercept = 0, linetype = 2)
```

Step 4: Coefficient Plots