Live Tidy Meeting Week 5: Effect of Newspaper Endorsements

Matthew Blackwell

Gov 51 (Harvard)

Introduction

- Can the media persuade people to vote differently?
- Hard problem from a causal POV because people **choose** their media.
    - Liberals choose to watch MSNBC and conservatives choose to watch Fox News
- Could do a lab experiment, but concerns about external validity.

Data setup

- Two political scientists tried to get around these issues in a particular moment in UK politics.
- Between 1992 and 1997, 4 UK newspapers switched their endorsements from Conservatives to Labour
    - Big surprise to the public!
- Our question: did readers of the switching newspapers vote differently in 1997?

Data

```r
library(tidyverse)
news <- read.csv("data/newspapers.csv")
```

Name	Description
to_labour	Read a newspaper that switched endorsement to Labour between 1992 and 1997 (1=Yes, 0=No)?
vote_lab_92	Did respondent vote for Labour in 1992 election (1=Yes, 0=No)?
vote_lab_97	Did respondent vote for Labour in 1997 election (1=Yes, 0=No)?
age	Age of respondent
male	Does the respondent identify as Male (1=Yes, 0=No)?
parent_labour	Did the respondent's parents vote for Labour (1=Yes, 0=No)?
work_class	Does the respondent identify as working class (1=Yes, 0=No)?

Question 1

Calculate the average treatment effect of reading a switching to Labour paper on voting for Labour in 1997 under a cross-sectional design.

Why might we not believe this estimate of the average treatment effect?

Answer 1

```r
## couple of ways to do this.
## first way is "pure tidy"
ate_pure <- news %>%
  group_by(to_labour) %>%
  summarize(across(vote_lab_97, mean), .groups = "drop") %>%
  summarize(
    ate = vote_lab_97[to_labour == "1"] -
      vote_lab_97[to_labour == "0"]
  )

ate_pure
```

```
## # A tibble: 1 x 1
##     ate
##   <dbl>
## 1 0.148
```

Answer 1 (cont'd)

```r
## alternative is to use filter() to create subsets and then use mean()
switched <- news %>%
  filter(to_labour == 1)

no_change <- news %>%
  filter(to_labour == 0)

ate <- mean(switched$vote_lab_97) - mean(no_change$vote_lab_97)
ate
```

```
## [1] 0.148
```

Question 2

Let's estimate the treatment effect using a before-and-after design. Calculate the difference between average vote for Labour in 1997 and average vote for Labour in 1992 **in the treated group**.

How does this estimate of the treatment effect compare to the cross-sectional design?

Answer 2

```r
ate_ba <- mean(switched$vote_lab_97) - mean(switched$vote_lab_92)
ate_ba
```

```
## [1] 0.172
```

Question 3

With the cross-sectional design, we might be worried about confounding here, so let's use **statistical control** and calculate the difference in means of `vote_lab_97` across the treatment and control groups within levels of one possible confounder: did the respondent's parents vote for Labour?

Answer 3

```r
news %>%
  group_by(parent_labour, to_labour) %>%
  summarize(across(vote_lab_97, mean)) %>%
  pivot_wider(
    names_from = to_labour,
    values_from = vote_lab_97
  ) %>%
  mutate(diff_by_parent = `1` - `0`)
```

```
## # A tibble: 2 x 4
## # Groups:   parent_labour [2]
##   parent_labour   `0`   `1` diff_by_parent
##           <int> <dbl> <dbl>          <dbl>
## 1             0 0.270 0.47          0.200
## 2             1 0.599 0.679         0.0804
```

Answer 3 (slightly nicer version)

```
news %>%
  mutate(
    to_labour = ifelse(to_labour == 1, "switched", "no_change"),
    parent_labour = ifelse(
      parent_labour == 1,
      "Parent Voted Labour",
      "Parent Didn't Vote Labour")
  ) %>%
  group_by(parent_labour, to_labour) %>%
  summarize(across(vote_lab_97, mean)) %>%
  pivot_wider(
    names_from = to_labour,
    values_from = vote_lab_97
  ) %>%
  mutate(diff_by_parent = switched - no_change)
```

```
## # A tibble: 2 x 4
## # Groups:   parent_labour [2]
##   parent_labour             no_change switched diff_by_parent
##   <chr>                         <dbl>    <dbl>          <dbl>
## 1 Parent Didn't Vote Labour     0.270    0.47          0.200
## 2 Parent Voted Labour           0.599    0.679         0.0804
```

Question 4

Use the same statistical control strategy to estimate the difference in means between the treated and control groups within levels of whether the respondent identified as working class (`work_class`).

Answer 4

```r
news %>%
  group_by(work_class, to_labour) %>%
  summarize(across(vote_lab_97, mean)) %>%
  pivot_wider(
    names_from = to_labour,
    values_from = vote_lab_97
  ) %>%
  mutate(diff_by_parent = `1` - `0`)
```

```
## # A tibble: 2 x 4
## # Groups:   work_class [2]
##   work_class   `0`   `1` diff_by_parent
##        <int> <dbl> <dbl>          <dbl>
## 1          0 0.257 0.416          0.158
## 2          1 0.551 0.674          0.123
```

Question 5

We might want to check the proportion of switchers there are in more complicated groups of the covariates. Create a new variable that takes on four values for each combination of `male` and `work_class` and calculate the difference in means within these groups.

Answer 5

```
news %>%
  mutate(
    gender_class = case_when(
      male == 1 & work_class == 1 ~ "Working Class Man",
      male == 0 & work_class == 1 ~ "Working Class Woman",
      male == 1 & work_class == 0 ~ "Non-working Class Man",
      male == 0 & work_class == 0 ~ "Non-working Class Woman"
    ),
    to_labour = ifelse(to_labour == 1, "Treated", "Control")
  ) %>%
  group_by(gender_class, to_labour) %>%
  summarize(across(vote_lab_97, mean)) %>%
  pivot_wider(names_from = to_labour, values_from = vote_lab_97) %>%
  mutate(est_ate = Treated - Control)
```

```
## # A tibble: 4 x 4
## # Groups:   gender_class [4]
##   gender_class            Control Treated est_ate
##   <chr>                     <dbl>   <dbl>   <dbl>
## 1 Non-working Class Man     0.242   0.435   0.193
## 2 Non-working Class Woman   0.269   0.387   0.118
## 3 Working Class Man         0.576   0.688   0.112
## 4 Working Class Woman       0.532   0.655   0.122
```

Question 6

Create a variable called `age_group` that groups respondents into the following groups by age:

- age is 25 and under,
- age is between 26-40,
- age is between 41-60, and
- age is 61 and over.

Calculate the difference in means between the treatment and control groups within levels of this variable.

Answer 6

```
news %>%
  mutate(
    age_group = case_when(
      age <= 25 ~ "<= 25",
      age >= 26 & age <= 40 ~ "26-40",
      age >= 41 & age <= 60 ~ "41-60",
      age >= 61 ~ ">= 61",
    ),
    to_labour = ifelse(to_labour == 1, "Treated", "Control")
  ) %>%
  group_by(age_group, to_labour) %>%
  summarize(across(vote_lab_97, mean)) %>%
  pivot_wider(names_from = to_labour, values_from = vote_lab_97) %>%
  mutate(est_ate = Treated - Control)
```

```
## # A tibble: 4 x 4
## # Groups:   age_group [4]
##   age_group Control Treated est_ate
##   <chr>       <dbl>   <dbl>   <dbl>
## 1 <= 25       0.6     0.421  -0.179
## 2 >= 61       0.415   0.614   0.198
## 3 26-40       0.48    0.617   0.137
## 4 41-60       0.375   0.570   0.195
```