Gov 50: 4. Randomized Experiments

Matthew Blackwell

Harvard University

Fall 2018

1/ Today's agenda

Where are we?

- What you've been doing:
    - Reading QSS, 2.1–2.4
    - Tried playing with RStudio and looked at R Markdown Playground project
    - Decided which section to attend this week.
    - DataCamp Assignment 2 due tonight at 11:59pm
- HW:
    - On Canvas and rstudio.cloud now.
    - Due 9/20 at 11:59 ET
    - Get started early!

Where are we going?

1. Introduction to randomized experiments
   - Causal effects
   - Role of randomization
2. Applied example: changing minds about gay marriage
   - Conditional statements, subsetting, factor variables

2/ Introduction to randomized experiments

Changing minds on gay marriage

- Question: can we effectively persuade people to change their minds?
- Hugely important question for political campaigns, companies, NGOs, etc.
- Psychological studies show it isn't easy.
- **Contact Hypothesis**: outgroup hostility diminished when people from different groups interact with one another.
- Today we'll explore this question the context of support for gay marriage and contact with a member of the LGBT community.
  - Y_i = support for gay marriage (1) or not (0)
  - T_i = contact with member of LGBT community (1) or not (0)

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would citizen i have supported gay marriage if they had been exposed to the LGBT community?
- Two **potential outcomes**:
  - $Y_i(1)$: would i have supported gay marriage if they **had** contact with a member of the LGBT community?
  - $Y_i(0)$: would i have supported gay marriage if they **didn't have** contact with a member of the LGBT community?
- **Causal effect**: $Y_i(1) - Y_i(0)$
- **Fundamental problem of causal inference**: only one of the two potential outcomes is observable.

Sigma notation

- We will often refer to the **sample size** (number of units) as n.
- Therefore, we often have n measurements of some variable, $(Y_1, Y_2, \dots, Y_n)$
- For a lot of reasons, we'll often want to refer to the sum of these variables:

$$Y_1 + Y_2 + Y_3 + \cdots + Y_n$$

- But this is cumbersome, so we often use the **Sigma notation**:

$$\sum_{i=1}^{n} Y_i = Y_1 + Y_2 + Y_3 + \cdots + Y_n$$

- $\sum_{i=1}^{n}$ says:
    1. Initialize the running sum to the case when $i = 1$.
    2. Increment i by 1 and add the new expression to the running sum.
    3. Repeat step 2 until $i = n$.

Averages

- The **sample average** or **sample mean** is simply the sum of all values divided by the number of values.
- Sigma notation allows us to write this in a compact way:

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- Suppose we surveyed 6 people and 3 supported gay marriage:

$$\overline{Y} = \frac{1}{6}(1+1+1+0+0+0) = 0.5$$

Quantity of interest

- We want to estimate the average causal effects over all units:

$$\textbf{Sample Average Treatment Effect (SATE)} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$$

- What we can estimate instead:

$$\textbf{Difference in means} = \overline{Y}_{\text{treated}} - \overline{Y}_{\text{control}}$$

- $\overline{Y}_{\text{treated}}$: observed average outcome for treated group
- $\overline{Y}_{\text{control}}$: observed average outcome for control group
- How do we ensure that the difference-in-means is a good estimate of the SATE?

Randomized control trials (RCT)

- Randomize!
- Key idea: **Randomization** of the treatment makes the treatment and control groups "identical" on average.
- The two groups are similar in terms of *all* characteristics (both observed and unobserved).
  - Control group is similar to treatment group
  - $\rightsquigarrow$ outcome in control group $\approx$ what would have happened to treatment group if they had control.

Some potential problems with RCTs

- **Placebo effects**:
  - Respondents will be affected by any intervention, even if they shouldn't have any effect.
- **Hawthorne effects**:
  - Respondents act differently just knowing that they are under study.

Balance checking

- Can we determine if randomization "worked"?
- If it did, we shouldn't see large differences between treatment and control group on **pretreatment variable**.
  - Pretreatment variable are those that are unaffected by treatment.
- We can check in the actual data for some pretreatment variable X
  - $\overline{X}_{\text{treated}}$: average value of variable for treated group.
  - $\overline{X}_{\text{control}}$: average value of variable for control group.
  - Under randomization, $\overline{X}_{\text{treated}} - \overline{X}_{\text{control}} \approx 0$

Multiple treatments

- Instead of 1 treatment, we might have multiple **treatment arms**:
  - Control condition
  - Treatment A
  - Treatment B
  - Treatment C, etc
- In this case, we will look at multiple comparisons:
  - $\overline{Y}_{\text{treated, A}} - \overline{Y}_{\text{control}}$
  - $\overline{Y}_{\text{treated, B}} - \overline{Y}_{\text{control}}$
  - $\overline{Y}_{\text{treated, A}} - \overline{Y}_{\text{treated, B}}$

3/ Gay marriage example

Changing minds on gay marriage

- Question: can we effectively persuade people to change their minds?
- Two randomized control trials in Los Angeles (2013)
- Timed around the Supreme Court decision to legalize gay marriage in CA
- LaCour & Green (2015). “When contact changes minds: An experiment of transmission of support for gay equality.” *Science*.

Study design

- Randomized treatment:
  - gay ($n = 22$) vs. straight ($n = 19$) canvassers with similar characteristics
  - same-sex marriage vs. recycling scripts (20 min conversation)
  - a total of 4 treatments: 2×2 factorial design
  - control group: no canvassing.

- Persuasion scripts are the same except one important difference:
  - gay canvassers: they would like to get married but the law prohibits it.
  - straight canvassers: their gay child, friend, or relative would like to get married but the law prohibits it.
- What is the recycling script for? $\rightsquigarrow$ **Placebo effect**
- Outcome measured via unrelated **panel survey**: self-reported support for same-sex marriage.
- Why use an "unrelated" survey? $\rightsquigarrow$ **Hawthorne effect**

The Data

- Data file: `gay.csv`

Name	Description
`study`	Source of the data (1 = Study1, 2 = Study2)
`treatment`	Five possible treatment assignment options
`wave`	Survey wave (a total of 7 waves)
`ssm`	5 point scale on same-sex marriage, higher scores indicate support.

- Load the data and create a cross-tabulation by `study` and `wave`:

```
gay <- read.csv("data/gay.csv")
table(gay$study, gay$wave)
```

```
##
##         1    2    3    4    5    6    7
##   1 9507 8465 8651 8672 8339 9013 6560
##   2 2441 2132 2113 2171    0    0 1528
```

Subsetting

- Let's focus on the baseline survey in Study 1:

```
study1.wave1 <- subset(gay, (study == 1) & (wave == 1))
```

- Examine the distribution of treatments:

```
prop.table(table(study1.wave1$treatment))
```

```
##
##                                      No Contact
##                                           0.551
##                Recycling Script by Gay Canvasser
##                                           0.110
##           Recycling Script by Straight Canvasser
##                                           0.109
##        Same-Sex Marriage Script by Gay Canvasser
##                                           0.121
## Same-Sex Marriage Script by Straight Canvasser
##                                           0.109
```

What do we expect if randomization is done correctly?

```r
tapply(study1.wave1$ssm, study1.wave1$treatment, mean)
```

```
##                                No Contact
##                                      3.04
##           Recycling Script by Gay Canvasser
##                                      3.13
##      Recycling Script by Straight Canvasser
##                                      3.01
##       Same-Sex Marriage Script by Gay Canvasser
##                                      3.03
## Same-Sex Marriage Script by Straight Canvasser
##                                      3.10
```

Estimating SATEs 3 days later (Wave 2)

- What is the effect of gay vs no canvasser?

```
study1.wave2 <- subset(gay, (study == 1) & (wave == 2))
none.ssm.w2 <- subset(study1.wave2,
                     treatment == "No Contact")
gay.ssm.w2 <- subset(study1.wave2,
   treatment == "Same-Sex Marriage Script by Gay Canvasser")
## estimated SATEs
mean(gay.ssm.w2$ssm) - mean(none.ssm.w2$ssm)
```

```
## [1] 0.0999
```

- What is the effect of straight vs no canvasser?

```
straight.ssm.w2 <- subset(study1.wave2,
    treatment == "Same-Sex Marriage Script by Straight Canvasser")
mean(straight.ssm.w2$ssm) - mean(none.ssm.w2$ssm)
```

```
## [1] 0.122
```

Script effect?

- Any effects of scripts for gay canvassers?

```
gay.rec.w2 <- subset(study1.wave2,
    treatment == "Recycling Script by Gay Canvasser")
mean(gay.ssm.w2$ssm) - mean(gay.rec.w2$ssm)
```

```
## [1] 0.032
```

- Any effects of scripts for straight canvassers?

```
straight.rec.w2 <- subset(study1.wave2,
    treatment == "Recycling Script by Straight Canvasser")
mean(straight.ssm.w2$ssm) - mean(straight.rec.w2$ssm)
```

```
## [1] 0.158
```

After the SCOTUS Decision (Wave 5)

```r
study1.wave5 <- subset(gay, (study == 1) & (wave == 5))
none.ssm.w5 <- subset(study1.wave5,
                     treatment == "No Contact")
gay.ssm.w5 <- subset(study1.wave5,
   treatment == "Same-Sex Marriage Script by Gay Canvasser")
## estimated SATEs
mean(gay.ssm.w5$ssm) - mean(none.ssm.w5$ssm)
```

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

```r
straight.ssm.w5 <- subset(study1.wave5,
    treatment == "Same-Sex Marriage Script by Straight Canvasser")
mean(straight.ssm.w5$ssm) - mean(none.ssm.w5$ssm)
```

```
## [1] 0.0986
```

9 months later (Wave 7)

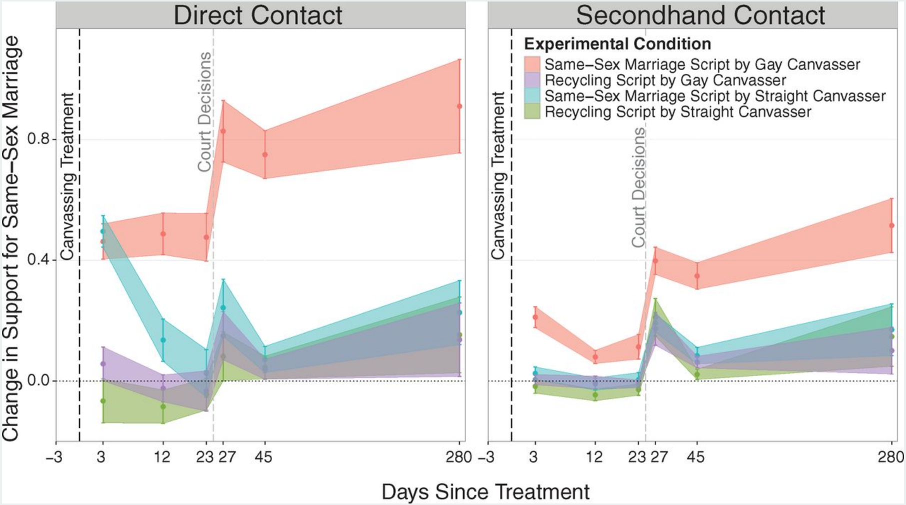
```r
study1.wave7 <- subset(gay, (study == 1) & (wave == 7))
none.ssm.w7 <- subset(study1.wave7,
                      treatment == "No Contact")
gay.ssm.w7 <- subset(study1.wave7,
   treatment == "Same-Sex Marriage Script by Gay Canvasser")
## estimated SATEs
mean(gay.ssm.w7$ssm) - mean(none.ssm.w7$ssm)
```

```
## [1] 0.0594
```

```r
straight.ssm.w7 <- subset(study1.wave7,
    treatment == "Same-Sex Marriage Script by Straight Canvasser")
mean(straight.ssm.w7$ssm) - mean(none.ssm.w7$ssm)
```

```
## [1] -0.0425
```

Big and lasting effects of persuasion



Retraction & media coverage

SECTIONS HOME SEARCH The New York Times SUBSCRIBE NOW LOG IN

Cuba's Environmental Concerns Grow

MATTER The Cambrian Explosion's Strange-Looking

Should Swimmers Worry About Sharks?

Liberia Reports New Cases of Ebola

OBSERVATORY Reaction to Smells May Help Diagnose Autism, Study Suggests

Heaven Scent: Finding May Help Restore Fragrance to Roses

It's the Pi to Have S Says

SCIENCE 359 COMMENTS

Doubts About Study of Gay Canvassers Rattle the Field

By BENEDICT CAREY and PAM BELLUCK MAY 25, 2015

Donald P. Green, left, a co-author of a challenged study by Michael LaCour, right, from Mr. LaCour's Facebook page.

Retraction

APR. 7, 2016 AT 2:00 PM

How Two Grad Students Uncovered An Apparent Fraud — And A Way To Change Opinions On Transgender Rights

By Christie Aschwanden and Maggie Koerth-Baker

Filed under Scientific Method

Hugh Tims, second from left, gives instructions to canvassers before going out in support of same-sex marriage. New research may change the way that political canvassing work is done. ROBERT F. BUKATY / AP

Follow up

POLITICAL SCIENCE

Durably reducing transphobia: A field experiment on door-to-door canvassing

David Broockman[1*] and Joshua Kalla[2]

Existing research depicts intergroup prejudices as deeply ingrained, requiring intense intervention to lastingly reduce. Here, we show that a single approximately 10-minute conversation encouraging actively taking the perspective of others can markedly reduce prejudice for at least 3 months. We illustrate this potential with a door-to-door canvassing intervention in South Florida targeting antitransgender prejudice. Despite declines in homophobia, transphobia remains pervasive. For the intervention, 56 canvassers went door to door encouraging active perspective-taking with 501 voters at voters' doorsteps. A randomized trial found that these conversations substantially reduced transphobia, with decreases greater than Americans' average decrease in homophobia from 1998 to 2012. These effects persisted for 3 months, and both transgender and nontransgender canvassers were effective. The intervention also increased support for a nondiscrimination law, even after exposing voters to counterarguments.

4/ Wrapping up

For next time

- Complete DataCamp Assignment 2
- Work on HW 1 (due next Thursday)
- Go to sections (see website/google calendar for times/locations)
- Read QSS 2.5 on Observational Studies for next time.