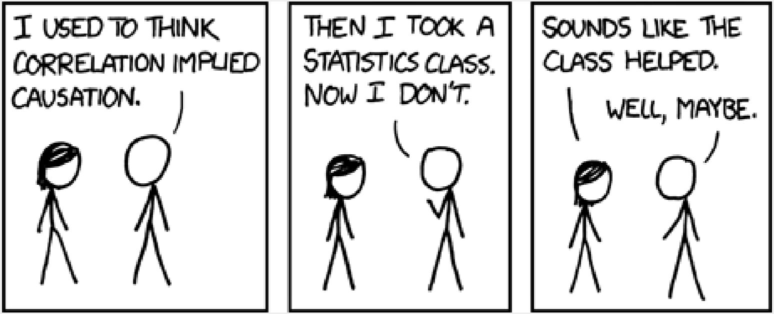
Causality


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Causal Questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up aŏter the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it
  - Would Jamal have gotten a job if he were white?

Logic of Causality

- Our intuition about causation relies too heavily on simple comparisons of pre-post change in outcomes before and after something happens
  - No change: no causation
  - Increase in outcome: positive effect
  - Decrease in outcome: negative effect
- Comparison between factual and counterfactual
- **Fundamental problem of causal inference**: We must infer counterfactual outcomes
- No causation without manipulation: **immutable characteristics**

Formalizing the logic

- The causal effect of the treatment T is the difference in Y with and without T

$$Y(T = 1) - Y(T = 0)$$

- We can never observe Y where T = 1 and T = 0 at the same time!

How do we draw causal inference?

- Anecdotes
- Correlations
- Regressions

These methods are all severely prone to error. Causal inference is a hard problem and invalid causal reasoning is one of the most common errors in human judgment, news reporting, and scientific studies!

Anecdotes

- "My grandpa Boris ate two cloves of garlic everyday and lived until he was 95 years old."
- For every anecdote you know, there might be many that you do not know that show the opposite pattern
- We often only raise those anecdotes that we like to see to justify actions or behaviors
- All that the anecdotes suggests is that Boris was prone to have a long life
- The key question for causal inference is about the unobserved counterfactual: how long would Boris have lived had she never ate garlic?

HEART BREAKING One glass of wine a day is worse for the heart than binge drinking, study finds

Shaun Wooller

18 Oct 2019, 10:10 | Updated: 18 Oct 2019, 10:13

- The problem with correlations for causal inference is that they often arise for reasons that have nothing to with the causal process under investigation (spurious correlation)
- Correlations are often driven by *selection effects*:
  - People who drink one glass of wine may already have exisiting health issues (Hence they're not binge drinking...)
  - Basketball players are tall, but does playing basketball make you taller?

Correlations

- Correlations are often driven by *confounding factors*:
  - ice cream sales are correlated with murder rates throughout a typical year.
  - Does not mean ice cream causes murders. Confounding factor:weather.
- Correlations are neither a necessary nor sufficient condition for causality.

Regressions

- Regressions are simply refined correlations that try to control for other confounding factors.
- Problems:
  - The list of all potential confounding factors is a bottomless pit.
  - How to properly control for confounders is often up for debate / unknown

Study Says Obesity Can Be Contagious

By Gina Kolata

July 25, 2007

- People whose friends tend to be obese might differ in many ways from those whose friends are not obese:
  - They might be poorer economically, live in areas with easier access to healthier food, have less access to sports, different hobbies, eating habits, etc.
- For causal inference we need to ask: among people who are identical in all respects, does making friends with obese persons really make them more likely to become obese?

Changing hearts and minds

- Question: can we effectively persuade people to change their minds?
- Hugely important question for media companies, 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.
- 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)

- What does T_i causes Y_i mean ~ **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.

- We want to estimate the average causal effects over all units: **Sample Average Treament Effect (SATE)**

$$\frac{1}{n}\sum_{i=1}^{n} Y_i(1) - Y_i(0)$$

- What we can estimate instead: **Difference in Means**

$$\bar{Y}_{treated} - \bar{Y}_{control}$$

- $\bar{Y}_{treated}$: observed average outcome for treated group
- $\bar{Y}_{control}$: 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

- 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
    - Outcome in control group ~ what would have happened to treatment group if they had control.

Randomisation

- Randomisation ensures the treatment is not correlated with any other variable
- Randomisation generates balance between treatment and control groups
- Treatment and control groups identical on average
- Allows us to estimate the average treatment effect simply as a difference in means

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

Internal versus external validity

- RCTs have very strong internal validity, that is, there is very little chance the result is derived from causes other than the treatment
- However, they may not generalise well. Why?
  - Samples may not reflect the whole population of interest
  - Treatment may be unrealistic ## 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*
  - X … average value of variable for treated group.
  - … e of variable for control group.

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:
    - $\bar{Y}_{treated,A} - \bar{Y}_{control}$
    - $\bar{Y}_{treated,B} - \bar{Y}_{control}$
    - $\bar{Y}_{treated,A} - \bar{Y}_{treated,B}$

Gay Marriage Example

- 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, M. J., and D. P. Green. 2014. “When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality.” *Science* 346(6215): 1366–69.

- Canvassers were given a script leading to conversations that averaged about twenty minutes. A distinctive feature of this study is that *gay and straight canvassers were randomly assigned to households and canvassers revealed whether they were straight or gay in the course of the conversation*. The experiment aims to test the **contact hypothesis**, which contends that out-group hostility (towards gays in this case) diminishes when people from different groups interact with one another.

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.

- `study` Study (1 or 2)
- `treatment` Treatment assignment: `No contact`, `Same-Sex Marriage Script by Gay Canvasser`, `Same-Sex Marriage Script by Straight Canvasser`, `Recycling Script by Gay Canvasser`, and `Recycling Script by Straight Canvasser`
- `wave` Survey wave (1-7). Note that Study 2 lacks wave 5 and 6.
- `ssm` Support for gay marriage (1 to 5).
  - Higher scores indicate more support.

- 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? ~ **Placebo effect**
- Outcome measured via unrelated **panel** survey: self-reported support for same-sex marriage.
- Why use an "unrelated" survey? ~ **Hawthorne effect**

Changing Minds on Gay Marriage

- Using the baseline interview wave before the treatment is administered (`wave == 1`), examine whether randomisation was properly conducted. Base your analysis on the three groups of Study 1:
    - ‘Same-Sex Marriage Script by Gay Canvasser’
    - ‘Same-Sex Marriage Script by Straight Canvasser’
    - ‘No Contact.’

Load and Filter Data

```
library(tidyverse)
gay = read.csv("C:/Users/afisher/Documents/R Code/qss/CAUSALITY/gay.csv")
str(gay)
```

```
## 'data.frame':    69592 obs. of  4 variables:
##  $ study    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ treatment: Factor w/ 5 levels "No Contact","Recycling Script by Gay Canvasser",..: 1 1 1
##  $ wave     : int  3 4 1 6 2 7 7 6 1 2 ...
##  $ ssm      : int  5 5 5 5 5 5 5 4 4 4 ...
```

```
# Subset to Study 1
gay = gay %>%
  filter(study == 1)

wave1 = gay %>%
  filter(wave == 1)

# Subset by treatment, wave, and study
w1_gay = gay %>%
  filter(wave==1 & treatment == "Same-Sex Marriage Script by Gay Canvasser")

w1_straight = gay %>%
  filter(wave==1  & treatment == "Same-Sex Marriage Script by Straight Canvasser")

w1_control = gay %>%
  filter(wave==1 & treatment == "No Contact")
```

Get Mean support by group

```r
# Mean support in gay canvasser group before treatment
mean(w1_gay$ssm)
```

```
## [1] 3.025195
```

```r
# Mean support in straight canvasser group before treatment
mean(w1_straight$ssm)
```

```
## [1] 3.09971
```

```r
# Mean support in control group before treatment
mean(w1_control$ssm)
```

```
## [1] 3.042764
```

Mean by treatment

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

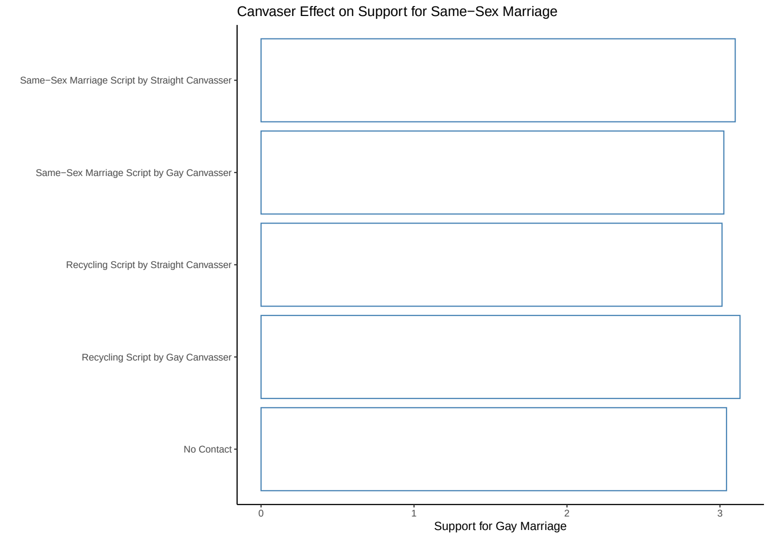
```
##                                No Contact
##                                  3.042764
##          Recycling Script by Gay Canvasser
##                                  3.130975
##     Recycling Script by Straight Canvasser
##                                  3.013474
##     Same-Sex Marriage Script by Gay Canvasser
##                                  3.025195
## Same-Sex Marriage Script by Straight Canvasser
##                                  3.099710
```

Plotting support for gay marriage across treatments

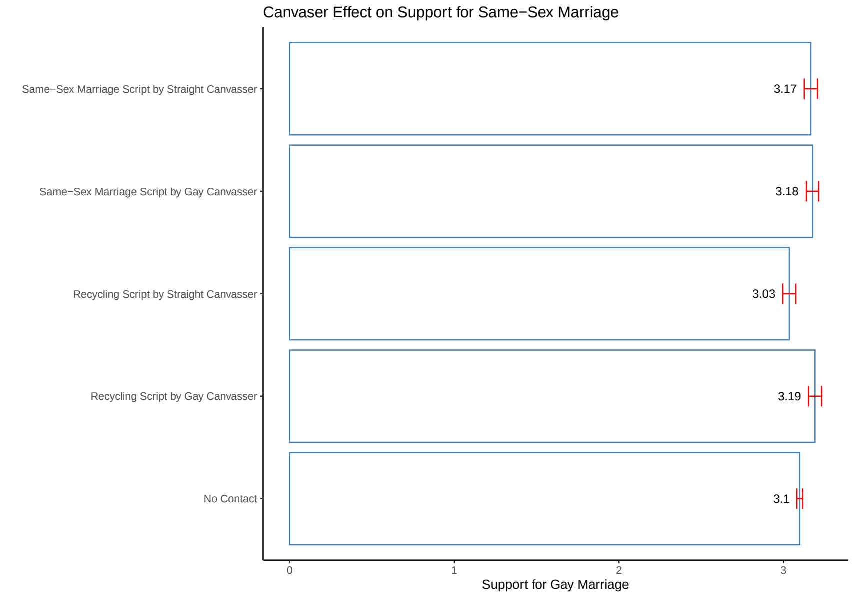
```
library(ggplot2)
# calculate the means of each combination of 'grade'
means <- wave1 %>% group_by(treatment) %>%
   summarise(mean = mean(ssm))

# making the plot
ggplot(means, aes(x = treatment, y = mean)) +
    geom_bar(stat = "identity") +
    coord_flip()
```

Plotting support for gay marriage across treatments



Adding Confidence Intervals



- The second wave of survey was implemented two months after the canvassing. Using Study 1, estimate the average treatment effects of gay and straight canvassers on support for same-sex marriage, separately (`wave == 2`). Give a brief interpretation of the results.

Treatment Effect in Wave 2

```r
w2_gay = gay %>%
  filter(wave==2 & treatment == "Same-Sex Marriage Script by Gay Canvasser")

w2_straight = gay %>%
  filter(wave==2  & treatment == "Same-Sex Marriage Script by Straight Canvasser")

w2_control = gay %>%
  filter(wave==2 & treatment == "No Contact")

# ATE for gay canvassers
mean(w2_gay$ssm) - mean(w2_control$ssm)
```

```
## [1] 0.09987463
```

```r
# ATE for straight canvassers
mean(w2_straight$ssm) - mean(w2_control$ssm)
```

```
## [1] 0.122248
```

What did we learn

- Approval of gay marriage increased on average by 0.1 for gay canvassers and by 0.122 for straight canvassers.
- Do results persist over time?

Treatment effects in Wave 7

```r
w7_gay = gay %>%
  filter(wave==7 & treatment == "Same-Sex Marriage Script by Gay Canvasser")

w7_straight = gay %>%
  filter(wave==7  & treatment == "Same-Sex Marriage Script by Straight Canvasser")

w7_control = gay %>%
  filter(wave==7 & treatment == "No Contact")

# ATE for gay canvassers
mean(w7_gay$ssm) - mean(w7_control$ssm)
```

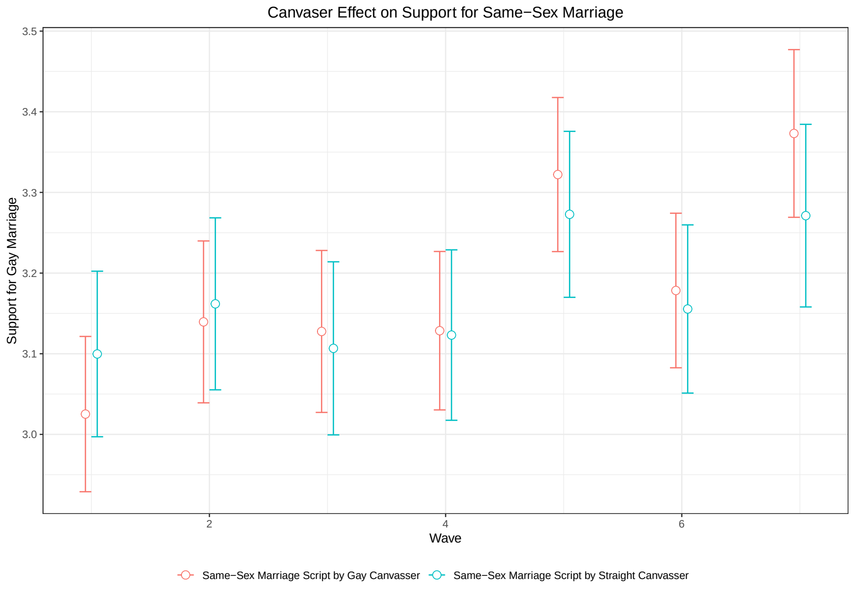
```
## [1] 0.05936835
```

```r
# ATE for straight canvassers
mean(w7_straight$ssm) - mean(w7_control$ssm)
```

```
## [1] -0.04253721
```

- Approval of gay marriage increased on average by 0.059 for gay canvassers and by -0.043 for straight canvassers. The results show that the effect for gay canvassers is persistent over time.

Any Lasting Effects?



Observational Studies

- Example of an **observational study:**
  - We as researchers observe a naturally assigned treatment
  - Very common: often can't randomize for ethical/logistical reasons.
- **Internal validity:** are the causal assumption satisfied? Can we interpret this as a causal effect?
  - RCTs usually have higher internal validity.
  - Observational studies less so, because pre-treatment variable may differ between treatment and control groups
- **External validity:** can the conclusions/estimated effects be generalized beyond this study?
  - RCTs weaker here because often very expensive to conduct on representative samples.
  - Observational studies often have larger/more representative samples that improve external validity

- **Confounder:** a pre-treatment variable affecting treatment and the outcome.
  - Example:
- **Confounding bias** in the estimated SATE due to these differences
  - $Y_{control}$ not a good proxy for $Y_i(0)$ in treated group.
  - one type: **selection bias** from self-selection into treatment

Observational Studies

- Why is it harder to make causal inference with observational data?
- **Confounders/confounding variables:** variables that are associated both with the treatment and the outcome
- If the treatment is not independent from pre-treatment variables, we cannot be sure about what causes what

Examples of Confounders

- Are incumbents more likely to win elections?
  - Maybe, but they also receive more campaign donations
- Are democratic countries more peaceful than authoritarian ones?
  - Maybe, but they also tend to be richer
- **Selection bias:** groups are not directly comparable

Observational Studies

- Pre-treatment variables have to be controlled for, that is, *held constant* in the analysis
- **Statistical control:** we use statistical methods to create balance between treatment and control and emulate an experiment as best as we can. E.g:
  - Instrumental variables (IV)
  - Regression discontinuity designs (RDD)
  - Differences-in-differences (DD/DiD/Diff-in-diff)

Statistical Control

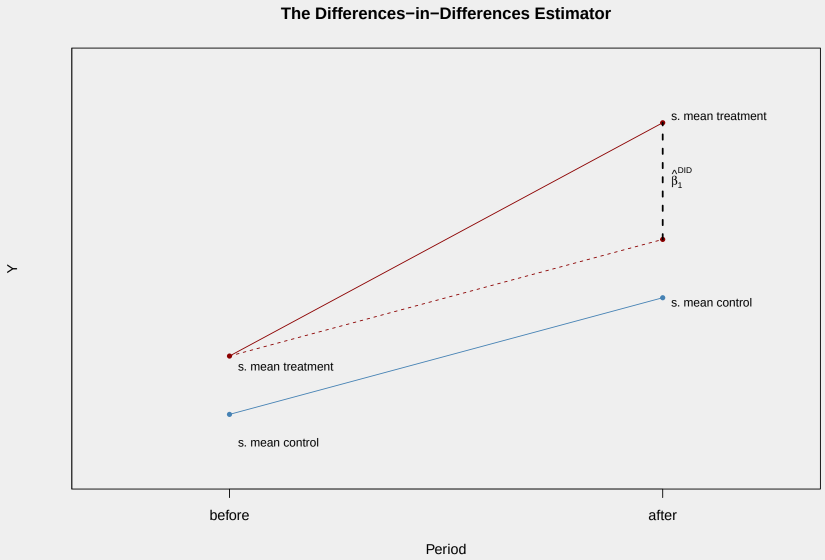
- **Statistical control:** adjust for confounders using statistical procedures.
  - can help to reduce confounding bias.
- One type of statistical control: **subclassification**
  - Compare treated and control groups within levels of a confounding variable.
  - Remaining effect can't be due to the confounder
- **Threat to inference:** we can only control for observed variables ~ threat of unmeasured confounding

- Key idea: use the before-and-after difference of **control group** to infer what would have happend to **treatment group** without treatment.
- Requires data for two time periods (before and after treatment) and at least two units
- **Parallel time trend assumption**
  - Example:
  - Threat to inference: non-parallel trends
- https://www.youtube.com/watch?v=dSlCBJSh96w&feature=youtu.be

For more see...

- https://uclspp.github.io/PUBL0050/
- https://egap.org/methods-guides/10-things-you-need-know-about-causal-inference
- http://www.cookbook-r.com/Graphs/Plotting_means_and_error_bars_(ggplot2)/
- https://cran.r-project.org/web/packages/afex/vignettes/afex_plot_introduction.html
- https://bookdown.org/ccolonescu/RPoE4/indvars.html#the-difference-in-differences-estimator
- https://www.econometrics-with-r.org/13-4-quasi-experiments.html
- https://www.econometrics-with-r.org/13-4-quasi-experiments.html
- https://www.mailman.columbia.edu/research/population-health-methods/difference-difference-estimation

DID Plot


The Differences−in−Differences Estimator
s. mean treatment
$\hat{\beta}_1^{DID}$
s. mean control
s. mean treatment
s. mean control
Y
before
after
Period