Potential outcomes model, randomized experiments, and power analysis

Tutorial 3

Stanislav Avdeev

Goal for today's tutorial

1. Enumerate tools used to discuss causal questions
2. Set terminology for causal inference
3. Understand different treatment effects
4. Discuss ways to do power analysis

The fundamental problem of CI

- The main goal in doing **causal inference** (CI) is to make as good a guess as possible about what Y would have been if D had been different
- That "would have been" is called a **counterfactual** - counter to the fact of what actually happened
  - in doing so, we want to think about two people/firms/countries that are basically exactly the same except that one has $D = 0$ and one has $D = 1$
- Suppose there are two variables
  - $Y \in \{0, 1\}$: whether a person is immune to Covid-19
  - $D \in \{0, 1\}$: whether a person gets a vaccine
- Our question: does D cause Y?
- **The fundamental problem of causal inference** (Holland 1986) is that for a given individual, we can only observe one world - either they get the vaccine, or they do not
- What is knowable?
  - first, we need some notation - **potential outcomes model** (Neyman-Rubin causal model)

Potential outcomes model

- The logic we just went through is the basis of the potential outcomes model, which is one way of thinking about causality
  - we can't observe the counterfactual, and must make an estimate of what the **outcome** would **potentially** have been under the counterfactual
  - figuring out what makes a good counterfactual estimate is a key part of causal inference
- What is the key assumption to make causal inference?
  - **SUTVA** - Stable Unit Treatment Variable Assignment, which states that person $i's$ outcome is only affected by their own treatment
  - How can we ensure SUTVA is satisfied? Randomized experiments are one of the available tools

Randomized experiments

- A common way to do causal inference in many fields is an experiment
  - if you can **randomly assign** D, then you know that the people with $D = 0$ are, on average, exactly the same as the people with $D = 1$
  - then we can easily estimate this model

$$Y_i = \alpha + \delta D_i + U_i$$

- However, when we're working with people/firms/countries, running experiments is often infeasible, impossible, or unethical
- So we have to think hard about a **model** of what the world looks like
  - so we can use some **model** to figure out what the counterfactual outcome would be (we will discuss that in the 5^{th} and 6^{th} tutorials)

Randomized experiments: simulation

- Let's simulate a dataset with a randomized treatment
- Let's say that getting a treatment D causes Y to increase by 1
- And let's run a randomized experiment of who actually gets D

```
set.seed(7)
df ← tibble(D = sample(c(0, 1), 1000, replace=T),
            Y0 = rnorm(1000),
            Y1 = Y0 + 1,
            Y_observed = ifelse(D == 1, Y1, Y0))
```

```
## # A tibble: 6 × 4
##       D     Y0     Y1 Y_observed
##   <dbl>  <dbl>  <dbl>      <dbl>
## 1     1 -0.206  0.794      0.794
## 2     0 -0.588  0.412     -0.588
## 3     0 -0.685  0.315     -0.685
## 4     1  1.00   2.00       2.00
## 5     0 -0.773  0.227     -0.773
## 6     1 -1.99  -0.994     -0.994
```

Randomized experiments: simulation

```
# The true effect is 1
df %>% group_by(D) %>% summarize(Y = mean(Y_observed))
```

```
## # A tibble: 2 × 2
##       D       Y
##   <dbl>   <dbl>
## 1     0 -0.0326
## 2     1  0.976
```

```
random ← lm(Y_observed ~ D, df) # we can use lm() to get the difference-in-means
```

	Model 1
(Intercept)	−0.033
	(0.045)
D	1.008***
	(0.063)
Num.Obs.	1000
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Randomized experiments: simulation

- Now this time we can't randomize D

```
set.seed(7)
df ← tibble(Z = runif(1000),
            D = ifelse(Z > 0.7, 1, 0),
            Y0 = rnorm(1000) + Z,
            Y1 = Y0 + 1,
            Y_observed = ifelse(D == 1, Y1, Y0))
```

```
## # A tibble: 6 × 5
##         Z     D     Y0     Y1 Y_observed
##     <dbl> <dbl>  <dbl>  <dbl>      <dbl>
## 1 0.989       1  0.783  1.78        1.78
## 2 0.398       0 -0.190  0.810      -0.190
## 3 0.116       0 -0.570  0.430      -0.570
## 4 0.0697      0  1.07   2.07        1.07
## 5 0.244       0 -0.529  0.471      -0.529
## 6 0.792       1 -1.20  -0.202      -0.202
```

Randomized experiments: simulation

```
# The true effect is 1
df %>% group_by(D) %>% summarize(Y = mean(Y_observed))
```

```
## # A tibble: 2 × 2
##       D     Y
##   <dbl> <dbl>
## 1     0 0.309
## 2     1 1.85
```

```
not_random ← lm(Y_observed ~ D, df)
```

	Model 1
(Intercept)	0.309***
	(0.038)
D	1.540***
	(0.068)
Num.Obs.	1000
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Randomized experiments: simulation

- But if we properly **model** the process and compare apples to apples

```
# The true effect is 1
not_random_modeled ← lm(Y_observed ~ D, df %>%
                          filter(abs(Z - 0.7) < 0.01)) # looks like RDD
```

	Model 1
(Intercept)	0.829+
	(0.429)
D	1.084+
	(0.575)
Num.Obs.	18
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

Identification

- In the first randomized case `lm(Y ~ D, df)` identifies the causal effect of X on Y
  - in other words, when we see the estimate, we can claim that it's the causal effect
- In the second non-randomized case `lm(Y ~ D, df)` does not identify the causal effect
- In the apples-to-apples comparison we could identify the causal effect
  - practically by using RDD (we will discuss RDD in the 6^{th} toturial)
- Causal inference is all about figuring out what **model** we need to identify the effect
  - but what effects are we identifying?

Treatment effects

- For any given treatment, there are likely to be **many treatment effects**
- Average Treatment Effect

$$ATE = \mathbb{E}(Y_1^* - Y_0^*) = \mathbb{E}(Y_1^*) - \mathbb{E}(Y_0^*)$$

  - ATE is the effect for the **full** population
- Average Treatment Effect on the Treated

$$ATET = \mathbb{E}(Y_1^* - Y_0^*|D = 1) = \mathbb{E}(Y_1^*|D = 1) - \mathbb{E}(Y_0^*|D = 1)$$

  - ATET is the effect for individuals who actually **received** the treatment
- Heterogeneous Treatment Effect

$$ATE(X) = \mathbb{E}(Y_1^* - Y_0^*|X) = \mathbb{E}(Y_1^*|X) - \mathbb{E}(Y_0^*|X)$$

  - ATE(X) is the effect that is different for individuals with **different** characteristics

Treatment effects

- What we get depends on the research design itself as well as the estimator we use to perform that design
- Which average you want depends on what you want to do with it
  - want to know how effective a treatment would be if applied to **everyone/at random**? ATE
  - want to know how effective a treatment **was** when it was applied? ATET
  - want to know how effective a treatment **was** when it was applied for males? ATE(X)
  - want to know how effective a treatment would be if applied **just a little more broadly?** Local Average Treatment Effect - LATE (next tutorial)
- Different treatment effects are not wrong but we need to pay attention to which one we're getting, or else we may apply the result incorrectly
  - a result could end up representing a different group than you're really interested in

Treatment effects: simulation

- Let's simulate some data and see what different methods give us

```
set.seed(7)
df ← tibble(group = sample(c('A', 'B'), 1000, replace = TRUE),
            b = case_when(group == 'A' ~ rnorm(1000, mean = 5, sd = 2),
                          group == 'B' ~ rnorm(1000, mean = 7, sd = 2)),
            D = rnorm(1000),
            Y = b*D + rnorm(1000))
```

```
## # A tibble: 6 × 4
##   group     b      D      Y
##   <chr> <dbl>  <dbl>  <dbl>
## 1 B     5.49   0.260  0.208
## 2 A     3.82  -1.25  -2.33
## 3 A     3.63  -1.31  -4.94
## 4 B     5.61   1.70  11.2
## 5 A     3.45  -0.511 -2.73
## 6 B     0.611 -1.16  -1.73
```

Treatment effects: simulation

- The true effect for group A is 5, for B is 7

```
m1 ← lm(Y ~ D, data = df)
m2 ← lm(Y ~ D*group, data = df)
m3 ← lm(Y ~ D, data = df[df$group == 'A',])
m4 ← lm(Y ~ D, data = df[df$group == 'B',])
```

	Model 1	Model 2	Model 3	Model 4
D	5.795***	4.880***	4.880***	6.773***
	(0.074)	(0.094)	(0.094)	(0.096)
groupB		−0.065		
		(0.132)		
D × groupB		1.893***		
		(0.135)		
Num.Obs.	1000	1000	484	516
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001				

Power analysis

- In experiments, we have some control over our sample size
  - so **before** collecting any data, we need to do a power analysis - what sample size do we need to be able to identify the effect?
- Power analysis also applies to observational data/non-experimental data
  - we just do not do it as often because we can't control the sample size anyway
  - and it is easier to get huge samples
- You need to have a huge sample to reasonably study small effects
  - so don't pursue effects that are likely to be really tiny, or at least tinier than your sample can handle
  - if you run an underpowered study anyway and **do** get a significant result, it would be more likely to be a false positive than a true positive. That's **low power**

Power analysis

- Power analysis balances five things
  1. size of the effect (coefficient in a regression, a correlation, etc.)
  2. sample size
  3. amount of variation in the treatment (the variance of D, say)
  4. amount of other variation in Y (the R^2, or the variation from the residual after explaining Y with D, etc.)
  5. power (the standard error of the estimate, statistical power, i.e. the true-positive rate)
- In order to do power analysis, you need to be able to fill in the values for four of those five pieces, so that power analysis can tell you the fifth one

Power analysis: implementation

- To calculate the **statistical power**, use standard practices
  - a goal is to achieve $80\% - 90\%$ statistical power
- To calculate the **minimum detectable effect**, use a standard formula

$$\text{MDE} = (t_{1-\alpha/2} - t_{1-q})\sqrt{\frac{1}{p(1-p)}}\sqrt{\frac{\sigma^2}{n}}$$

- To calculate the **smallest sample size**, use a standard formula

$$n = \left(\frac{t_{1-\alpha/2} - t_{1-q}}{MDE}\right)^2 \frac{\sigma^2}{p(1-p)}$$

- Empirically you can do power analysis using
  - `power.t.test()` in `stats`
  - multiple functions in `powerMediation`
  - simulations

References

Books

- Huntington-Klein, N. The Effect: An Introduction to Research Design and Causality, Chapter 10: Treatment Effects
- Cunningham, S. Causal Inference: The Mixtape, Chapter 4: Potential Outcomes Causal Model

Slides

- Huntington-Klein, N. Econometrics Course, Week 8: Experiments
- Huntington-Klein, N. Causality Inference Course, Lecture 3: Causality and Lecture 18: Treatment Effects
- Goldsmith-Pinkham P. Applied Empirical Methods Course, Week 1: Potential Outcomes and Directed Acylic Graphs