STA 235H - RCTs and Observational Studies

Fall 2023

McCombs School of Business, UT Austin

Housekeeping

- Let's talk about ChatGPT.
  - Should be use as a *complement* of learning, not a *substitute*.
  - ChatGPT is mainly useful when you are able to check the *accuracy* of its answers.
  - You need to do your <u>own work</u>.
- No Office Hours this Thursday.
  - I will hold OH (for this week) on Tues (4pm - 5:30pm) and Wed (10:30am - 11:30am)

Last week

- We talked about the Ignorability Assumption
- Started discussing randomized controlled trials.
  - Why they are the gold standard.
  - How to analyze them.

Today

- Discuss about limitations of RCTs:
  - Generalizability
  - Spillover/General equilibrium effects.
- What is selection on observables?:
  - Omitted Variable Bias
  - Regression Adjustment
  - Matching

Limitations of RCTs

Recap

- RCTs make the **ignorability assumption hold by design**

How?

Examples of RCTs

LinkedIn Ran Social Experiments on 20 Million Users Over Five Years

A study that looked back at those tests found that relatively weak social connections were more helpful in finding jobs than stronger social ties.

Give this article 163

Researchers examined changes that LinkedIn had made to its "People You May Know" algorithm to test what sociologists call the "strength of weak ties." Sundry Photography/Alamy

The NYT Open Team Follow

Mar 4, 2021 · 8 min read · Listen

Save

How We Rearchitected Mobile A/B Testing at The New York Times

We use A/B tests to make decisions about the products and features we release, but our mobile test allocation wasn't separating users properly and we had to figure out how to fix it.

Steps to analyze a RCT?

1) Check for balance

(Remember to transform categorical variables into binary ones!)

2) Estimate diff. in means

(Simple regression between Y and Z)

2)* Estimate diff. in means with covariates

(Multiple regression between Y and Z, adding other baseline covariates X)

Potential issues to have in mind

Generalizability of our estimated effects (External Validity)

- Where did we get our sample for our study from? Is it representative of a larger population?

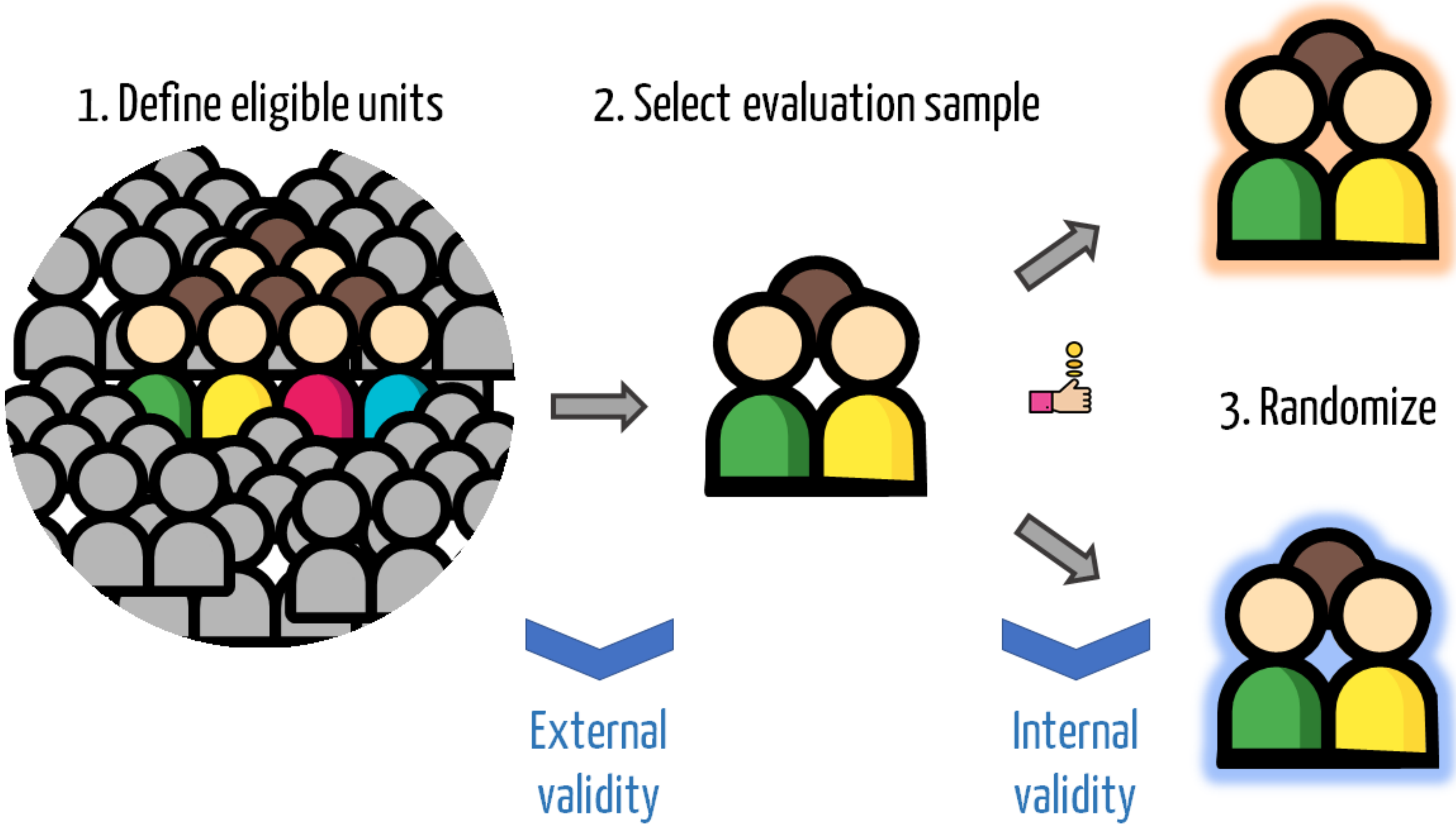
Spillover effects

- Can an individual in the control group be affected by the treatment?

General equilibrium effects

- What happens if we scale up an intervention? Will the effect be the same?

External vs Internal Validity



- Many times, RCTs use convenience samples

SUTVA: No interference

- Aside from ignorability, RCTs rely on the Stable Unit Treatment Value Assumption (SUTVA)

"The treatment applied to one unit does not affect the outcome for other units"

- No spillovers
- No general equilibrium effects

Network effects (spillover) example

- RCT where students where randomized into two groups:
  - Treatment: Parents receive a text message when student misses school.
  - Control: Parents receive a general text message.
- Estimate the effect of the intervention on attendance.
  - Difference in average attendance between treated students and control students.
- Potential problem: Students usually skip school with a friend.

Why could this be a problem for causal inference?

Network effects

Can we do something about this?

1. Randomize at a higher level (e.g. neighborhood, school, etc. instead of at the individual level)
2. Model the network!

General Equilibrium Effects

- Usually arise when you **scale up** a program or intervention.
- Imagine you want to test the effect of providing information about employment and expected income to students to see whether it affect their choice of university and/or major.

What could happen if you offer it to everyone?

Let's see another example

Get Out The Vote

- "Get out the Vote" Large-Scale Mobilization experiment (Arceneaux, Gerber, and Green, 2006)
  - "Households containing one or two registered voters where randomly assigned to treatment or control groups"
  - Treatment: GOTV phone calls
  - Stratified RCT: Two states divided into competitive and noncompetitive (randomized within state-competitiveness)

Checking for balance

Balance Table by Stratum

	Non-competitive		Competitive		Non-competitive		Competitive	
	Treat	Control	Treat	Control	Treat	Control	Treat	Control
female2	0.552	0.546	0.541	0.535	0.549	0.545	0.543	0.541
fem_miss	0	0	0	0	0.026	0.025	0.022	0.021
age	52.157	51.977	50.81	50.862	55.795	55.782	53.481	53.464
newreg	0.117	0.116	0.133	0.134	0.048	0.049	0.048	0.046
persons	1.496	1.497	1.513	1.518	1.539	1.538	1.529	1.533
vote98	0.231	0.227	0.258	0.259	0.572	0.574	0.599	0.594
vote00	0.564	0.567	0.595	0.593	0.734	0.732	0.781	0.78

Let's go to R

Estimating the effect

- One important thing to note in the previous analysis is that **assignment to treatment $\neq$ contact**

```
d_s1 %>% count(treat_real, contact)
```

```
##   treat_real contact     n
## 1          0       0 17186
## 2          1       0  1626
## 3          1       1  1374
```

Does this break the ignorability assumption?

- **Non-compliance**: When the treatment assignment (e.g. calling the household) is not the same as the treatment (e.g. actually receiving a call/ making contact with the household)

- What was **randomly assigned** was calling the household.

- Usually, <u>the effect of calling should be lower than the effect of actually receiving the call</u>.

Can we do something if we can't randomize??

Controlling by Confounders

Controlling by Confounders

- We can control by a confounder by **including it in our regression**:
  - After we control for it, we are doing a fair comparison (e.g. *"holding X constant"*)

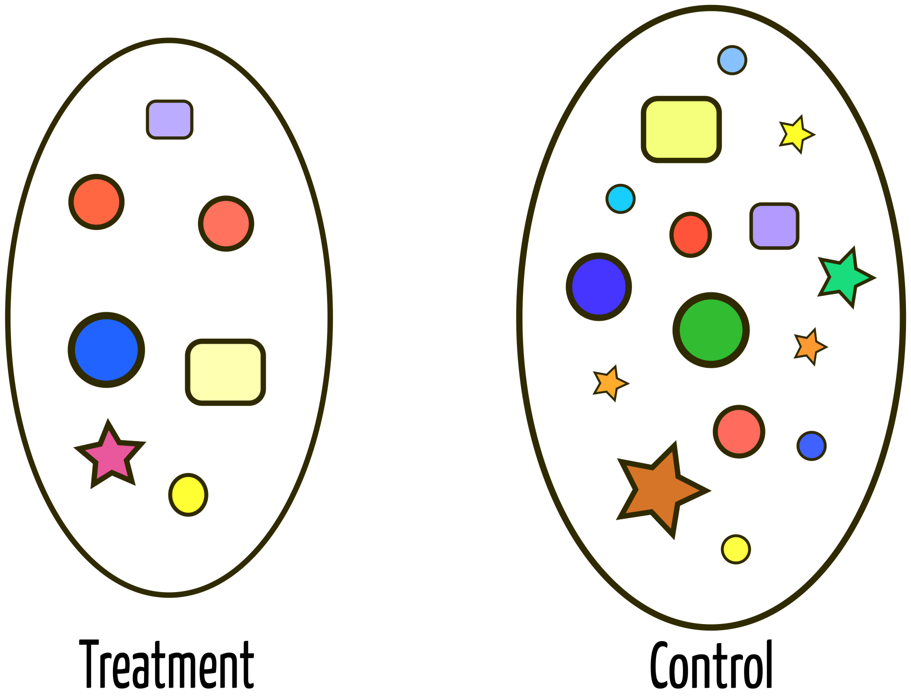
Conditional Independence Assumption (CIA)

- *"<u>Conditional on X</u>, the ignorability assumption holds."*
- But is there another way to control for confounders?

Matching

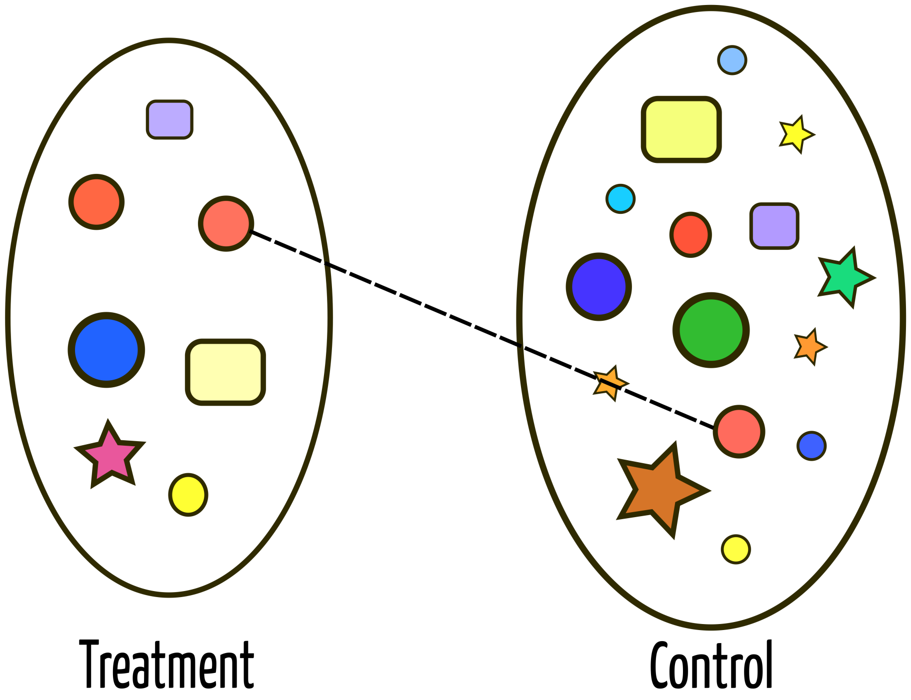
Matching

Start with two groups: A treatment and a control group



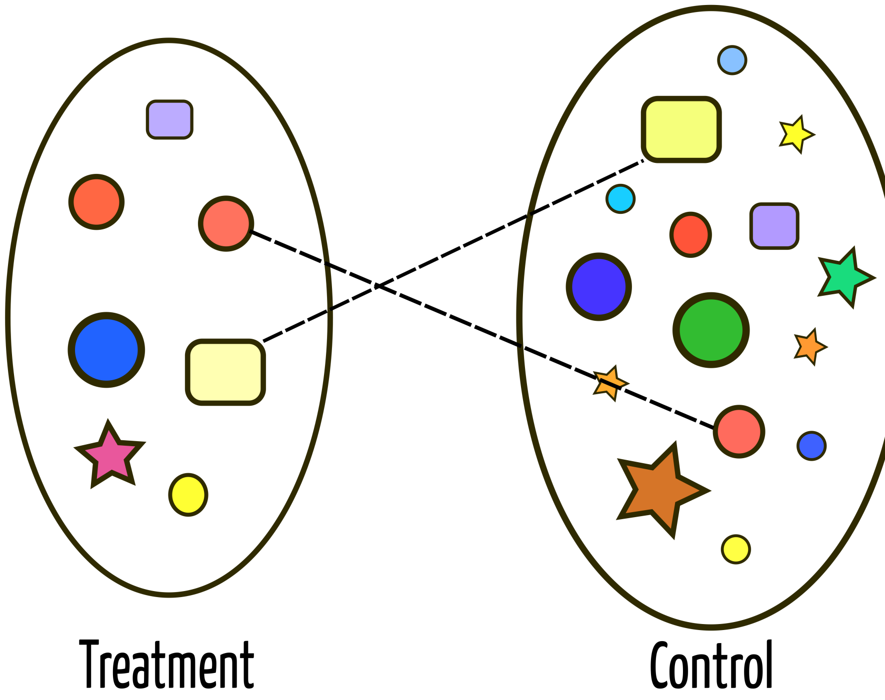
Matching

For each unit in the treatment group, let's find a similar unit in the control group



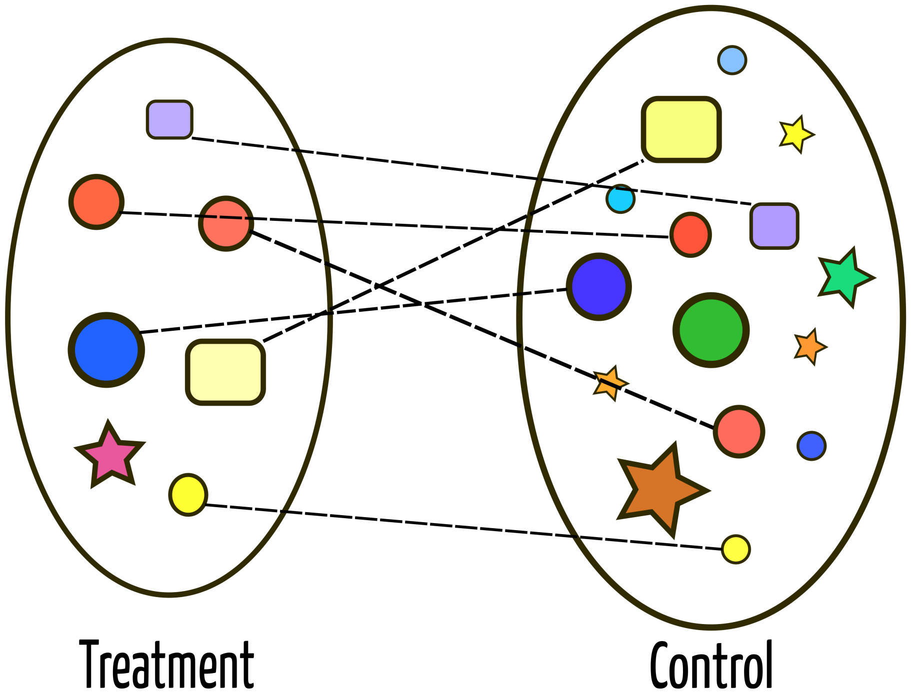
Matching

And we do this for all units



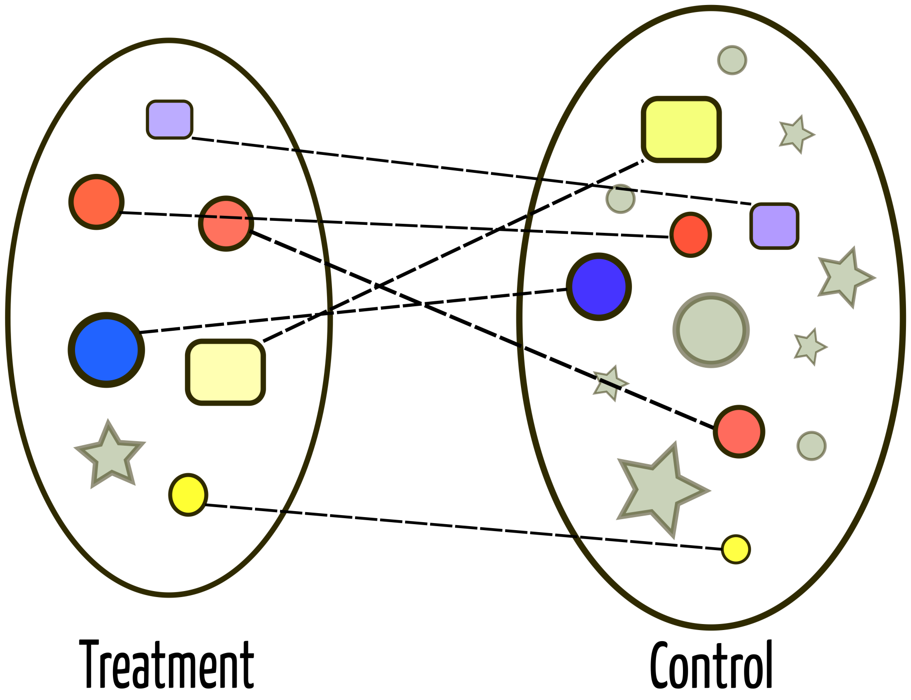
Matching

Note that we might not be able to find similar units for everyone!



Matching

Then we just compare our matched groups



Propensity Score Matching

- It is difficult (impossible) to match on all the variables we want (potential confounders)
    - The curse of dimensionality
- Propensity score: Probability of being in the treatment group given the individuals characteristics.

$$p = Pr(Z = 1) = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k$$

- E.g. Two units have a 50% chance of being treated, but one was actually treated (Z=1) and the other one was not (Z=0).
- Don't need to calculate this by hand; we will use the `MatchIt` package.

Let's go to R

Omitted Variable Bias

- If we are under the presence of **confounders**, then our estimates will be biased (i.e. will not recover the true causal effect) *unless we are able to control by them.*
- **Omitted Variable Bias** represents the bias that stems from not being able to observe a confounding variable.
- If a potential confounder **is in our data**, then it's not a problem!
  - We can **control** for it.
- Our headache will come from **unobserved confounders**.

Wrapping things up

- If the ignorability assumption doesn't hold, I can potentially control by all my confounders.
  - Conditional Independence Assumption.
- Unlikely to hold
- Do we have other alternatives?
  - Let's see next class!