Experiments and Randomization

EC420 MSU

Justin Kirkpatrick

Last updated March 10, 2021

From before...

When we think of the "choice or occurrance" we want to study as the treatment, then *selection bias* is the distortion of the statistic of interest (the Average Treatment Effect) due to the fact that a person's outcomes ($Y_{0,i}$ and $Y_{1,i}$) depend on selection of treatment.

That is, the expectation of the treated value and the expectation of the untreated value change with treatment.

$$E[Y_0|D=0] \neq E[Y_0|D=1]$$

We had examples from our previous lecture

1. Effect of smoking on Life Expectancy

2. Effect of energy consumption information on energy use

3. Anyone remember any others?

Selection bias

The energy consumption example:

Our previous example

Name	Treated	Use Without Tmt	Use With Tmt
Allison	TRUE	4	3
Bertrand	TRUE	5	4
Carmine	FALSE	3	2
Dalia	FALSE	4	3

Our **Mastering Metrics** reading is about *randomization* to eliminate selection bias, so let's try it.

Get into a group by row of seating

Randomization

In groups, "draw from the data" by randomly assigning treatment

- First, randomly decide to which group each member belongs.
- Second, build a new observed dataset with the randomly realized observed values
  - The "with" value if the person is randomly placed in the with-treatment
  - The "without" value if the person is randomly placed in the without-treatment
- Third, calculate the group means and the difference between the group means.
- Repeat 15 times.

Name	Use Without Tmt	Use With Tmt
Allison	4	3
Bertrand	5	4
Carmine	3	2
Dalia	4	3

Randomization

Report the 15 trial results: the mean with, the mean without, and the difference between means

Calculate the average over all the randomized trials. What is the average with- and without-treatment value? What is the average difference?

What was our previous result comparing the means?

We know that the Average Treatment Effect is 1. Which one is closer to the true value of 1?

Randomization (of treatment) eliminates selection bias

Let's see if we can show this *formally*....

Recall that in a previous lecture, we used some algebra on conditional expectations to *decompose* the naive difference in means into two parts:

$$
\begin{aligned}
& E[Y|D=1] - E[Y|D=0] \\
= {} & E[Y_1|D=1] - E[Y_0|D=0] \\
= {} & E[Y_1|D=1] - E[Y_0|D=0] - E[Y_0|D=1] + E[Y_0|D=1] \\
= {} & E[Y_1|D=1] - E[Y_0|D=1] + E[Y_0|D=1] - E[Y_0|D=0] \\
= {} & \underbrace{E[(Y_1 - Y_0)|D=1]}_{\text{SATE}} + \underbrace{E[Y_0|D=1] - E[Y_0|D=0]}_{\text{selection bias}}
\end{aligned}
$$

MICHIGAN STATE UNIVERSITY

$$
\begin{aligned}
& E[Y|D=1] - E[Y|D=0] \\
&= E[Y_1|D=1] - E[Y_0|D=0] \\
&= E[Y_1|D=1] - E[Y_0|D=0] - E[Y_0|D=1] + E[Y_0|D=1] \\
&= E[Y_1|D=1] - E[Y_0|D=1] + E[Y_0|D=1] - E[Y_0|D=0] \\
&= \underbrace{E[(Y_1 - Y_0)|D=1]}_{\text{SATE}} + \underbrace{E[Y_0|D=1] - E[Y_0|D=0]}_{\text{selection bias}}
\end{aligned}
$$

Now, we saw that randomization got us closer to the correct answer. It might even be unbiased! But what does randomization do?

- It randomly assigns treatment
- So what would $E[Y_0|D=1]$ be if treatment, D, is random? It would be:

$$E[Y_0|D=1] = E[Y_0|D=0] = E[Y_0] \quad \textbf{under randomization}$$

So what happens to *selection bias*?

Under randomization, where treatment is randomly assigned, selection bias is equal to zero in expectation

Remember the definition of mean independence: $E[Y|X] = E[Y]$ for RV's Y, X

Applying the same thing to **potential outcomes** of Y and D:

- Randomization gives us $(Y_1, Y_0) \perp D$
- $(Y_1, Y_0) \perp D \Rightarrow E[Y_0|D] = E[Y_0]$

A comparison of means is then an *unbiased estimate* of the $SATE$

How would we randomize to answer the following with a simple cross-sectional comparison of means

- Effect of energy consumption information on future energy consumption
- Effect of health insurance on health outcomes
- Effect of college education on future earnings

Randomization, in it's simplest form, is quite straightforward. If the econometrician* can reliably and with certainty assign treatment to individuals, then great!

But even in closely-managed operation, this won't always work.

- People are good at finding ways around assignment
- Even drug trials, which are about as controlled as you can get, are subject to shenanigans.
- *Endogenous treatment* is the term frequently used for this
    - It means that treatment is determined "within the system" (by an individual affected by the treatment).
    - You will be sick of the term "endogenous" by the end of this semester.

---

* that's you

Relating SATE to ATE

- We seem to be able to get to SATE, but what about the ATE?
    - $SATE = E[Y_1 - Y_0 | D = 1]$
    - $ATE = E[Y_1 - Y_0 | D = 1]$
- If we assume that $E[(Y_1 - Y_0)|D = 1] = E[(Y_1 - Y_0)|D = 0]$, then the $SATE$ is the same as the ATE
- That is, the effect of treatment on the treated is the same as the effect of treatment on the untreated

This is the *constant effects assumption*

....OK, back to randomization

Randomized Control Trial

When the econometrician is able to randomly decide treatment, we call the study a *randomized control trial* or *RCT*.

- These are *very* common in development economics as they can be done at low-cost in many places.

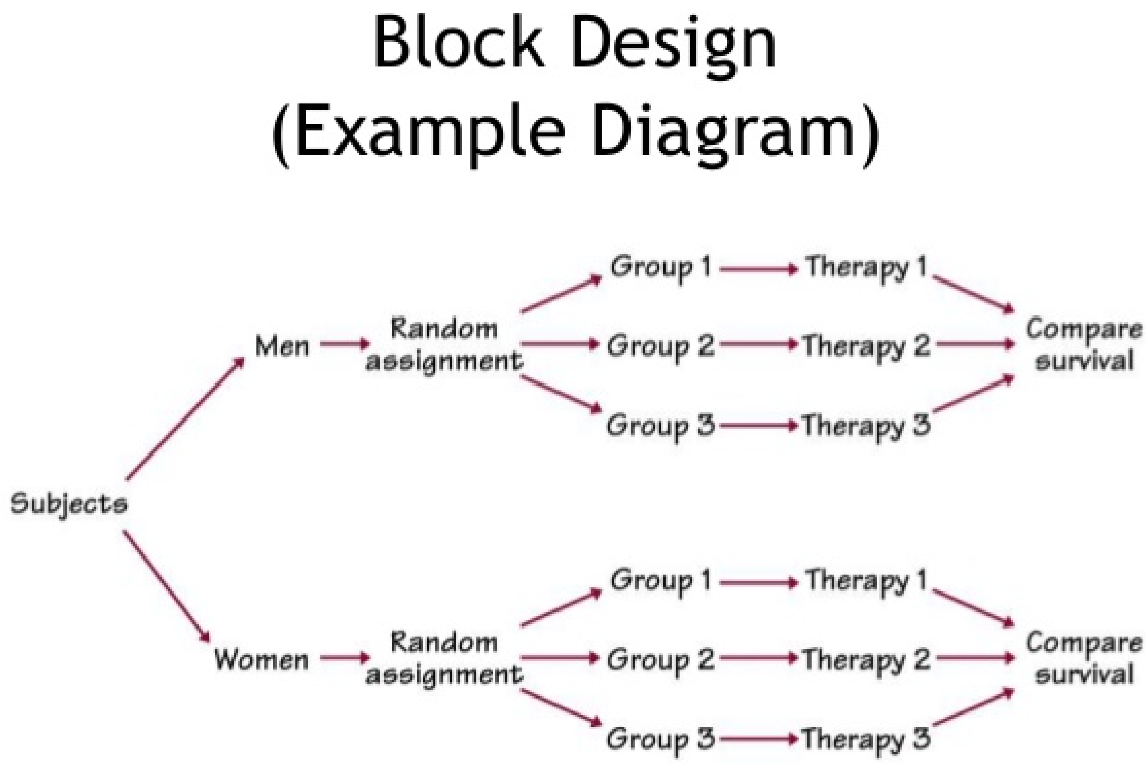
Often, when doing a RCT, we want to make sure we have randomly selected a similar sample into each group (treated, untreated).

- Imagine if you randomly selected people to take part in a smoking campaign, but your "treated" group were mostly older people. If your outcome of interest is "mortality rate", what would happen?

Blocking (stratification)

A *blocked* or *stratified* experiment means we have randomly assigned treatment inside of observable subgroups to ensure balance.

- For example, if our population is 40% college graduates and 60% non-college graduates, and we are assigning treatment to 50% of the participants, we would treat 50% of all college graduates and 50% of all non-college graduates. We would *not* randomly select 50% of all participants.
- "Blocking" because you group observables into "blocks", then select
  - You can block on more than one characteristic: "college graduates over 40", etc..
- This can help ensure balance.
- Of course, this requires being able to assign treatment.



Source: Bravo Students

1. We talked more about selection bias and examles of it.
2. We talked about randomization and how it can alleviate selection bias.
3. We also showed in notation how randomization alleviates selection bias in comparison of means.
   - Because $E[Y_0|D=1] = E[Y_0|D=0]$
4. We defined unbiased and consistent estimators (sample statistics).