Lecture 14

Alex Stephenson

10-13-2021

What did we cover last time?

We went over in abbreviated form the entirety of the semester so far

What are we doing today?

Remind everyone that there is both a checkpoint and a weekly practice due this Friday

Reminder to turn in your make up for PS2 by Wednesday

Today we are going to talk about attrition

Attrition

Attrition: Outcome data is missing for units in the study

When attrition is systematically related to potential outcomes removing observations from the dataset leads to selection bias.

Why does attrition occur?

Units may refuse to cooperate with researchers

Researchers lose track of units in an experiment

There is interference with researchers finding out an outcome from a unit

Outcomes are intrinsically unavailable

Researchers drop units

Attrition Notation

Consider potential outcomes defined as $Y_i(z)$ where $z \in \{0, 1\}$

Under no attrition:

$$Y_i = Y_i(0)(1 - z_i) + Y_i(1)z_i$$

Under attrition

$$r_i = r_i(0)(1 - z_i) + r_i(1)Z_i$$

Which implies when $r_i = 1$

$$Y_i = Y_i(0) + [Y_i(1) - Y_i(0]z_i$$

and missing otherwise

When does attrition lead to bias

Assume we are interested in the ATE. We can rewrite this to include missing data as

$$E[Y_i(1)|R_i(1) = 1] - E[Y_i(0)|R_i(0) = 1]$$

Bias of this equation is based on the relationship between missingness R_i and outcomes Y_i.

If missing outcomes are completely at random (MCAR), then we do not have bias.

If missing outcomes are not completely at random, we begin to worry about possible bias.

Forms of attrition

MCAR: Missing completely at random

MAR: Missing at random

MNAR: Missing not at random

Estimands under attrition

ATE among "Always Reporters"

Suppose that all units are either never missing or always missing outcomes.

In this case, an unweighted comparison of average treatment and control group outcomes provides an unbiased estimate of the ATE among always reporters.

Such a situation might occur when there is a delay between intervention and measurement of the outcome.

Bounding ATE under attrition

If attrition is non random we may still be able to place bounds on the value of the ATE.

We choose to be purposely conservative with these bounds to make minimal assumptions (Manski 1989)

Extreme Value Bounds gauge the potential consequences of attrition by examining how the estimated ATE varies depending on how one fills in missing potential outcomes.

Reducing the threat of attrition

Attrition should be considered at the design stage

To mitigate the threat of attrition attempts should be made BEFORE data is collected to plan for follow up data.

The most promising approach is to plan for an intensive effort to reduce attrition under a random sample of subjects with missing data.

What if covariate data is missing?

Missing data for covariates is far less series. In a purely randomized experiment, missing covariate data does not affect the consistency of our estimates. Why?

If a covariate is missing, use the following strategy:

1. Assign an arbitrary value to units with missing values
2. Create a dummy variable for missing values
3. Regress outcomes on the treatment, covariates, and the new dummy

Should we only analyze subgroups without attrition

Comparing subgroups that do not suffer from attrition will produce biased treatment effects of the ATE

Example:

Suppose we run an experiment to help units run faster, but units have differential responses. Some units will love the program and keep going. Others will hate it and quit.

Example

Unit	Pair	Y(0)	Y(1)	r(0)	r(1)	Y(0) Given r(0)	Y(1) Given r(1)
1	A	8	12	1	1	8	12
2	A	8	4	1	0	9	Missing
3	B	16	16	1	1	16	16
4	B	18	18	1	1	18	18

Example

Imagine we drop the pair where there is attrition. Using the remaining data can we recover the true ATE of 0?

No

Units assigned to T	Units assigned to C	Estimated ATE	Drops?
{1,3}	{2,4}	1	No
{1,4}	{2,2}	3	No
{2,3}	{1,4}	-2	Yes
{2,4}	{1,3}	2	Yes

Our estimator produces $\frac{1+3+2-2}{4} = 1 \neq 0$