Lectures on causal inference and experimental methods

Macartan Humphreys

Roadmap

Road Map

- Lecture 1: ▸ What is a cause? ▸ Potential Outcomes ▸ Estimands ▸ Endogenous subgroups
- Lecture 2: ▸ What's an experiment? ▸ What is a Design? ▸ Declare Your Design
- Lecture 3: ▸ Assignment Schemes ▸ Balance ▸ Power
- Lecture 3b: ▸ Assignments with `DeclareDesign` ▸ Simple ▸ Complete ▸ Blocked
- Lecture 4: ▸ Inference ▸ No more regressions ▸ Estimating variance ▸ Randomization Inference
- Lecture 5: ▸ Complications ▸ LATE ▸ Spillovers ▸ Mediation
- Lecture 6: ▸ Opportunities ▸ Ethics ▸ Transparency ▸ Constraints

▸ credits

Take home ideas

- Random assignment to treatment is random sampling from alternative universes.
- Be able to demonstrate that your design is complete for some purpose.
- You have to have an estimand.
- Analyze as you randomize.
- Regression requires many more assumptions in order to lay claim to estimating average causal effect.
- Use the data to understand a model and not a model to understand the data.
- Start worrying after the first stage.
- Think about spillovers at the design stage.
- Seek generality.

Lecture 1: What's a cause?

Motivation

The *intervention* based motivation for understanding causal effects:

- We want to know if a particular intervention (like aid) caused a particular outcome (like reduced corruption).
- We need to know:
  1. What happened?
  2. What would the outcome have been if there were no intervention?
- The problem
  1. . . . this is hard
  2. . . . this is impossible

  The problem in 2 is that you need to know what would have happened if things were different. You need information on a **counterfactual**

The Potential Outcomes Framework

Potential Outcomes

- For each unit we assume that there are two **post-treatment** outcomes: $Y_i(1)$ and $Y_i(0)$.
- eg $Y(1)$ is the outcome that **would** obtain *if* the unit received the treatment.
- The **causal effect** of Treatment (relative to Control) is:

$$\tau_i = Y_i(1) - Y_i(0)$$

- Note:
  - the causal effect is defined at the *individual level*.
  - there is no "data generating process" or functional form
  - the causal effect is defined relative to something else and so a counterfactual must be conceivable (did Germany cause the second world war?)
  - are there any substantive assumptions made here so far?

Potential Outcomes

Now that we have a concept of causal effects available, let's answer two **questions**:

- If for a given unit A causes B and B causes C, does that mean that A causes C?

- Say A causes B — does that mean that there is a spatiotemporally continuous sequence of causal intermediates?

Potential Outcomes

Now that we have a concept of causal effects available, let's answer two **questions**:

- If for a given unit A causes B and B causes C, does that mean that A causes C?
  A boulder is flying down a mountain. You duck. This saves your life. So the boulder caused the ducking and the ducking caused you to survive. So: *did the boulder cause you to survive?*
- Say A causes B — does that mean that there is a spatiotemporally continuous sequence of causal intermediates?
  Person A is planning some action Y; Person B sets out to stop them; person X intervenes and prevents person B from stopping person A. In this case Person A may complete their action, producing Y, without any knowledge that B and X even exist; in particular B and X need not be anywhere close to the action. So: *did X cause Y?*

Causal claims: Contribution or attribution?

The counterfactual model is all about contribution, not attribution, except in a very conditional sense.

- Focus is on non-rival contributions
- Not: what caused Y but what is the effect of X?
- At most it provides a conditional account

Consider at outcome Y that might depend on two causes X_1 and X_2:

$$Y(0,0) = 0$$

$$Y(1,0) = 0$$

$$Y(0,1) = 0$$

$$Y(1,1) = 1$$

What caused Y? Which cause was most important?

Causal claims: Contribution or attribution?

The counterfactual model is all about contribution, not attribution, except in a very conditional sense.

- Focus is on non-rival contributions
- Not: what caused Y but what is the effect of X?
- At most it provides a conditional account
- This is problem for research programs that define "explanation" in terms of figuring out the things that cause Y
- Real difficulties conceptualizing what it means to say one cause is more important than another cause. What does that mean?

Causal claims: Contribution or attribution?

The counterfactual model is all about contribution, not attribution, except in a very conditional sense.

- Focus is on non-rival contributions
- Not: what caused Y but what is the effect of X?
- At most it provides a conditional account
- *Erdogan's increasing authoritarianism was the most important reason for the attempted coup*
  - More important than Turkey's history of coups?
  - What does that mean?

Causal claims: No causation without manipulation

- Some seemingly causal claims not admissible.
- To get the definition off the ground, manipulation must be imaginable (whether practical or not)
- This renders thinking about effects of race and gender difficult
- What does it mean to say that Aunt Pat voted for Brexit because she is old?

Causal claims: No causation without manipulation

- Some seemingly causal claims not admissible.
- To get the definition of hte ground, manipulation must be imaginable (whether practical or not)
- This renders thinking about effects of race and gender difficult
- **Compare**: What does it mean to say that Southern counties voted for Brexit because they have many old people?

Causal claims: Causal claims are everywhere

- Jack exploited Jill
- It's Jill's fault that bucket fell
- Jack is the most obstructionist member of Congress
- Melania Trump stole from Michelle Obama's speech
- Activists need causal claims

Causal claims: What is actually seen?

- We have talked about what's potential, now what do we *observe*?
- Say Z_i indicates whether the unit i is assigned to treatment ($Z_i = 1$) or not ($Z_i = 0$). It describes the treatment process. Then what we observe is:

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

- Say Z is a random variable, then this is a sort of data generating process. BUT the key things to note is
  - Y_i is random but the randomness comes from Z_i — the potential outcomes, $Y_i(1)$, $Y_i(0)$ are fixed
  - Compare this to a regression approach in which Y is random but the X's are fixed. eg:

$$Y \sim N(\beta X, \sigma^2) \text{ or } Y = \alpha + \beta X + \epsilon, \epsilon \sim N(0, \sigma^2)$$

Causal claims: The estimand and the rub

- The causal effect of Treatment (relative to Control) is:

$$\tau_i = Y_i(1) - Y_i(0)$$

- This is what we want to estimate
- BUT: We never can observe both $Y_i(1)$ and $Y_i(0)$!
- This is the **fundamental problem** (Holland)

Causal claims: The rub and the solution

- Now for some magic. We really want to estimate:

$$\tau_i = Y_i(1) - Y_i(0)$$

- BUT: We never can observe both $Y_i(1)$ and $Y_i(0)$
- Say we lower our sights and try to estimate an **average** treatment effect:

$$\tau = E(Y(1) - Y(0))$$

- Now make use of the fact that

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

- In words: *The average of differences is equal to the difference of averages*; here, the average treatment effect is equal to the difference in average outcomes in treatment and control units.
- The magic is that *while we can't hope to measure the differences; we are good at measuring averages*.

Causal claims: The rub and the solution

- So we want to estimate $E(Y(1))$ and $E(Y(0))$.
- We know that we can estimate averages of a quantity by taking the average value from a random sample of units
- To do this here we need to select a random sample of the $Y(1)$ values and a random sample of the $Y(0)$ values, in other words, we **randomly assign** subjects to treatment and control conditions.
- When we do that we can in fact estimate:

$$E_N(Y_i(1)|Z_i = 1) - E_N(Y_i(0)|Z_i = 0)$$

which in expectation equals:

$$E(Y_i(1)|Z_i = 1 \text{ or } Z_i = 0) - E(Y_i(0)|Z_i = 1 \text{ or } Z_i = 0)$$

- This highlights a deep connection between **random assignment** and **random sampling**: when we do random assignment *we are in fact randomly sampling from different possible worlds*.

Causal claims: The rub and the solution

This provides a **positive argument** for causal inference from randomization, rather than simply saying with randomization "everything else is controlled for"

Let's discuss:

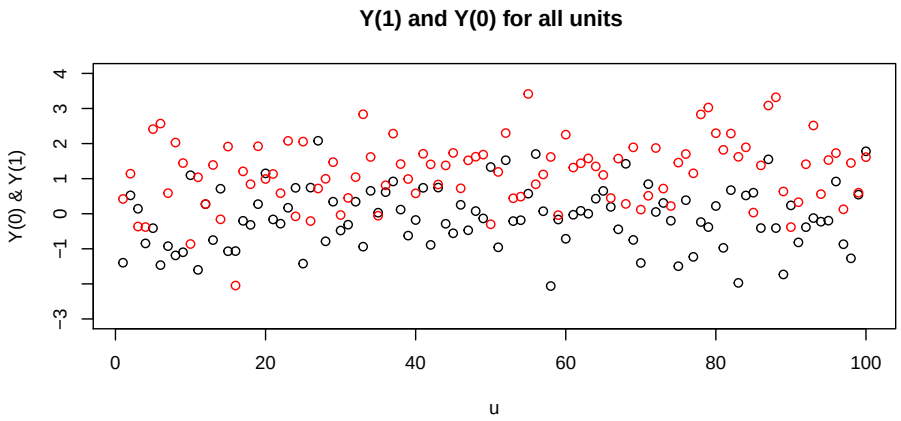
- Does the fact that an estimate is unbiased mean that it is right?
- Can a randomization "fail"?
- Where are the covariates?

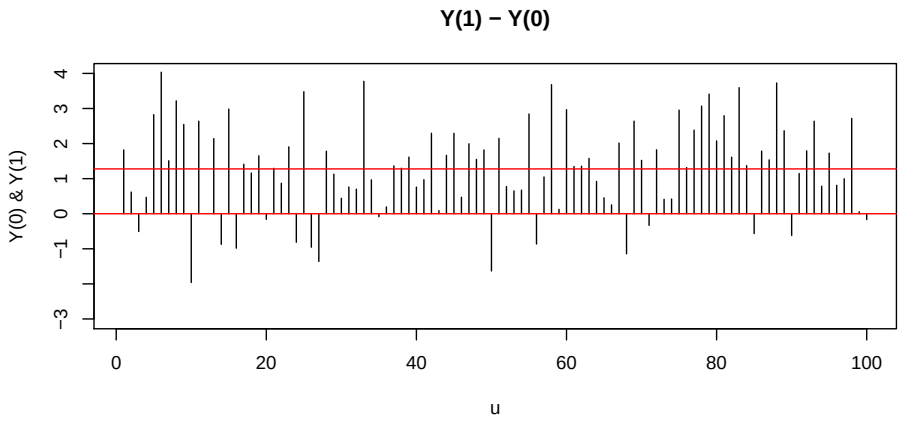
Idea: random assignment is random sampling from potential worlds: to understand anything you find, you need to know the sampling weights

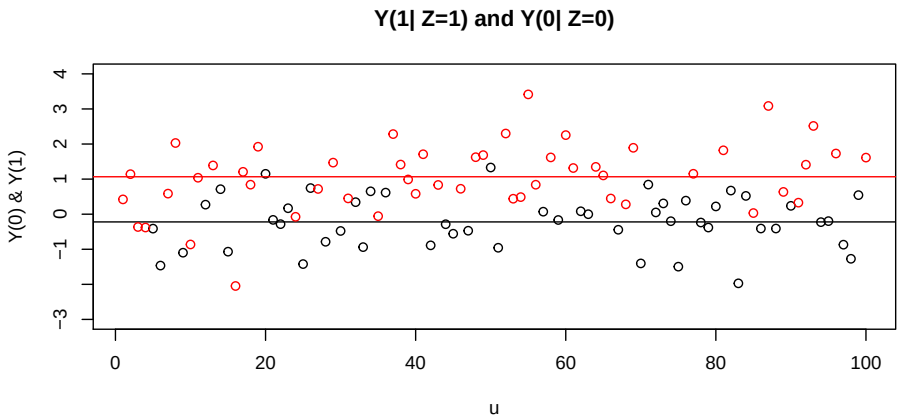
Potential outcomes: why randomization works

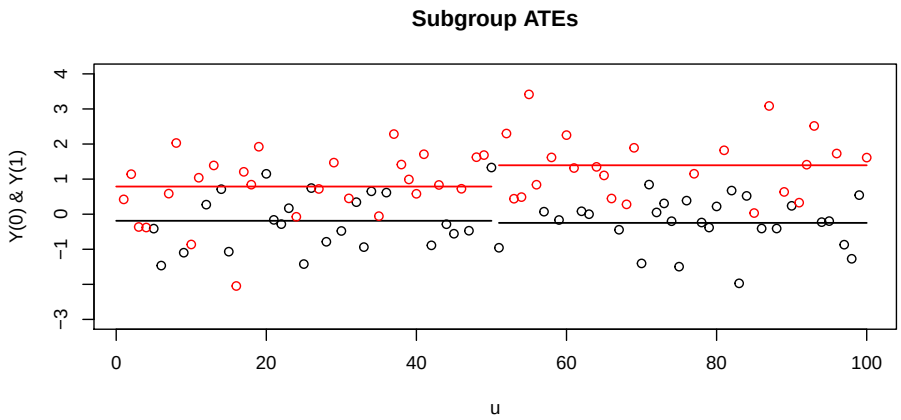
The average of the differences $\approx$ difference of averages

```
po.graph(N, Y0, Y1, u, Z)
```





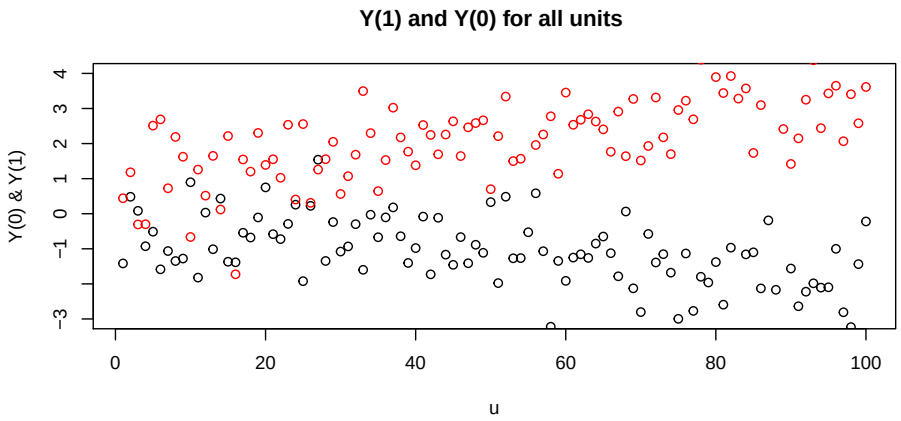


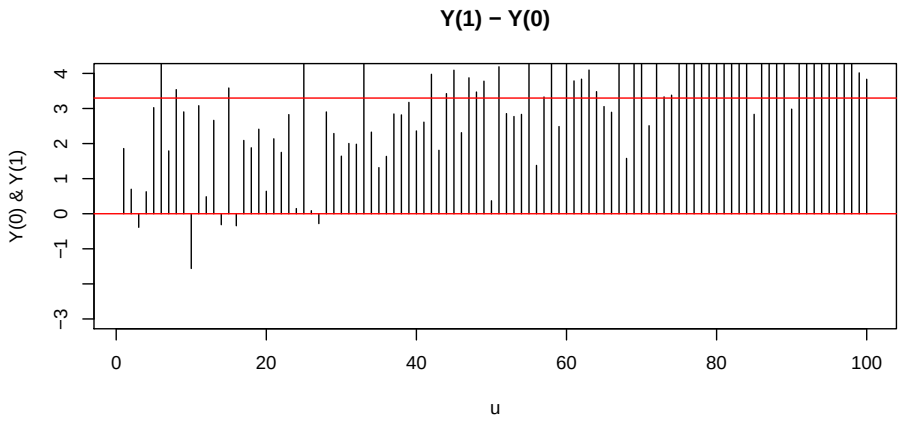


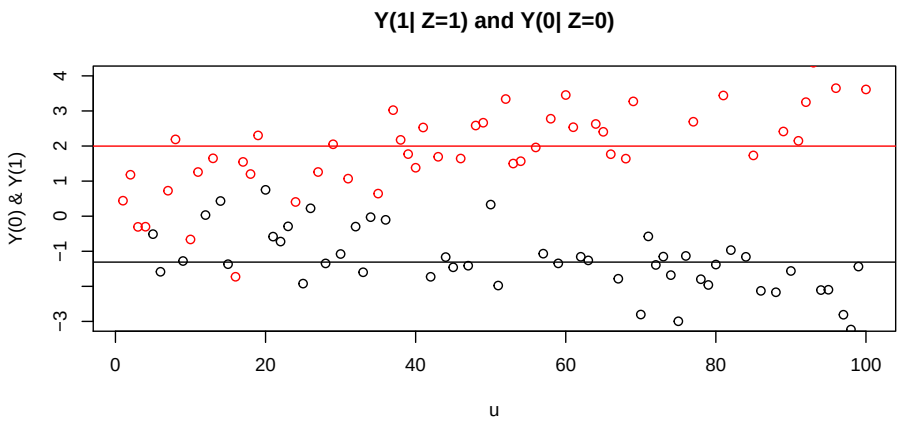
Potential outcomes: heterogeneous effects

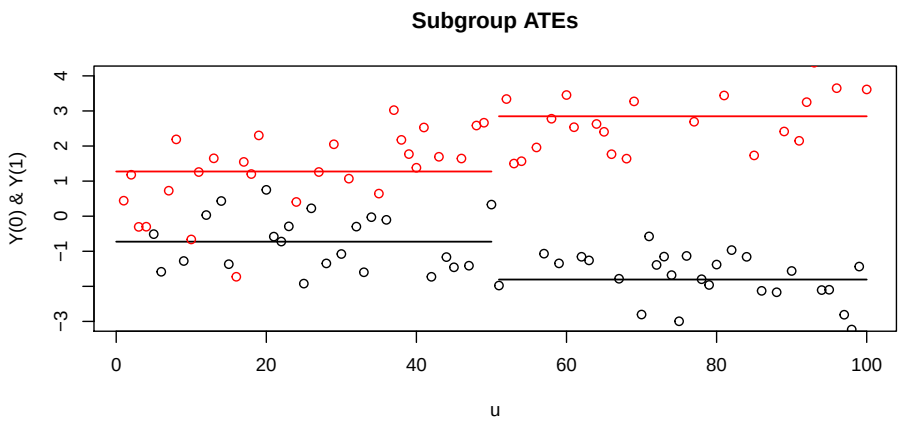
The average of the differences $\approx$ difference of averages

```
po.graph(N, Y0 - u/50, Y1+u/50, u,Z)
```









Potential outcomes: heterogeneous effects

Question: $\approx$ or $=$?

Estimands and Estimators

Estimands

- The estimand is the thing you want to estimate
- If you are estimating something you should be able to say what your estimand is
- You are responsible for your estimand. Your estimator will not tell you what your estimand is
- Just because you can calculate something does not mean that you have an estimand
- You can test a hypothesis without having an estimand

Estimands: ATE, ATT, ATC, S-, P-, C-, ITT, LATE

Say that units are randomly assigned to treatment in different strata (maybe just one); with fixed, though possibly different, shares assigned in each stratum. Then the key estimands and estimators are:

$$
\begin{array}{lllll}
\tau_{ATE} \equiv & E(\tau_i) & = & \sum_x \frac{w_x}{\sum_j w_j}\tau_x & \widehat{\tau}_{ATE} = \sum_x \frac{w_x}{\sum_j w_j}\widehat{\tau}_x \\
\tau_{ATT} \equiv & E(\tau_i | Z_i = 1) & = & \sum_x \frac{p_x w_x}{\sum_j p_j w_j}\tau_x & \widehat{\tau}_{ATT} = \sum_x \frac{p_x w_x}{\sum_j p_j w_j}\widehat{\tau}_x \\
\tau_{ATC} \equiv & E(\tau_i | Z_i = 0) & = & \sum_x \frac{(1-p_x) w_x}{\sum_j (1-p_j) w_j}\tau_x & \widehat{\tau}_{ATC} = \sum_x \frac{(1-p_x) w_x}{\sum_j (1-p_j) w_j}\widehat{\tau}_x
\end{array}
$$

where x indexes strata, p_x is the share of units in each stratum that is treated, and w_x is the size of a stratum.

Here:

- ATE is Average Treatment Effect (all units)
- ATT is Average Treatment Effect on the Treated
- ATC is Average Treatment Effect on the Controls

Estimands: ATE, ATT, ATC, S-, P-, C-, ITT, LATE}

In addition, each of these can be targets of interest:

- for the **population**, in which case we refer to PATE, PATT, PATC and $\widehat{PATE}, \widehat{PATT}, \widehat{PATC}$
- for a **sample**, in which case we refer to SATE, SATT, SATC, and $\widehat{SATE}, \widehat{SATT}, \widehat{SATC}$

And for different subgroups,

- given some value on a covariate, in which case we refer to CATE (conditional average treatment effect)
- for unobservable subgroups, we estimate LATE (Local Average Treatment Effect (see below).

With non-compliance we might estimate ITT —the "intention to treat" effect

Skip to ▸ Fixer or ▸ Inference 1 or ◂ Big Ideas

Exercise your potential outcomes 1

Consider the following potential outcomes table:

Unit	Y(0)	Y(1)	τ_i
1	4	3	
2	2	3	
3	1	3	
4	1	3	
5	2	3	

Questions for us: What are the unit level treatment effects? What is the average treatment effect?

Exercise your potential outcomes 2

Consider the following potential outcomes table:

In treatment?	Y(0)	Y(1)
Yes		2
No	3	
No	1	
Yes		3
Yes		3
No	2	

Questions for us: Fill in the blanks.

- Assuming a constant treatment effect of $+1$
- Assuming a constant treatment effect of -1
- Assuming an *average* treatment effect of 0

What is the actual treatment effect?

Endogeneous subgroups

Endogeneous Subgroups

Experiments often give rise to endogenous subgroups. The potential outcomes framework can make it clear why this can cause problems.

Heterogeneous Effects with Endogeneous Categories

- Problems arise in analyses of subgroups when the categories themselves are affected by treatment
- Example from our work:
  - You want to know if an intervention affects reporting on violence against women
  - You measure the share of all subjects that experienced violence that file reports
  - The problem is that which subjects experienced violence is itself a function of treatment

Heterogeneous Effects with Endogeneous Categories

It is possible that in truth no one's reporting behavior has changed, what has changed is the propensity of people with different propensities to report to experience violence:

	Violence(Treatment)		Reporting(Treatment, Violence)			
	V(0)	V(1)	R(0,1)	R(1,1)	R(0,0)	R(1,0)
Type 1 (reporter)	1	1	1	1	0	0
Type 2 (non reporter)	1	0	0	0	0	0

Expected reporting given violence in control $=$ Pr(Type 1)

Expected reporting given violence in treatment $= 100\%$

Question: What is the actual effect of treatment on the propensity to report violence?

Heterogeneous Effects with Endogeneous Categories

It is possible that in truth no one's reporting behavior has changed, what has changed is the propensity of people with different propensities to report to experience violence:

	Reporters		Non reporters		
	Experience Violence		Experience Violence		
	No	Yes	No	Yes	% Report
Control	25	25	25	25	$\frac{25}{25+25} = 50\%$
Treatment	25	25	50	0	$\frac{25}{25+0} = 100\%$

Heterogeneous Effects with Endogeneous Categories

This problem can arise as easily in seemingly simple field experiments. Example:

- In one study we provided constituents with information about performance of politicians
- we told politicians in advance so that they could take action
- we wanted to see whether voters punished poorly performing politicians
- what's the problem?

Heterogeneous Effects with Endogeneous Categories

Question for us:

Setting:

- Quotas for women are randomly placed in a set of constituencies in year 1. All winners in these areas are women; in other areas only some are.
- In year 2 these quotas are then lifted.

Questions Which problems face an endogenous subgroup issue?:

1. You want to estimate the likelihood that a woman will stand for reelection in treatment versus control areas in year 2.
2. You want to estimate how much incumbents are more likely to be reelected in treatment versus control areas in year 2.
3. You want to estimate how much treatment areas have more relected incumbents in elections in year 2 compared to control.

Heterogeneous Effects with Endogeneous Categories

In such cases you can:

- Examine the joint distribution of multiple outcomes
- Condition on pretreatment features only
- Engage in mediation analysis

Missing data can create an endogeneous subgroup problem

- It is well known that missing data can undo the magic of random assignment.
- One seemingly promising approach is to match into pairs *ex ante* and drop pairs together *ex post*.
- Say potential outcomes looked like this (four units divided into two pairs):

Table 1: Full profile of potential outcomes

Pair	I	I	II	II	
Unit	1	2	3	4	Average
Y(0)	0	0	0	0	
Y(1)	-3	1	1	1	
τ	-3	1	1	1	

Missing data

- Say though that cases are likely to drop out of the sample if things go badly (eg they get a negative score or die)
- Then you might see no attrition in cases in which people that are likely to drop out if treated do not get treated.
- You might assume you have no problem (after all, no attrition).

Table 2: No missing data when the normal cases happens to be selected

Pair	I	I	II	II	
Unit	1	2	3	4	Average
Y(0)	0		0		0
Y(1)		1		1	1
$\hat{\tau}$					1

Missing data

- But in cases in which you have attrition, dropping the pair doesn't necessarily help.
- The problem is potential missingness still depends on potential outcomes
- The kicker is that the method can produce bias even if (*in fact*) there is no attrition!

Table 3: Missing data when the vulnerable cases happens to be selected

Pair	I	I	II	II	
Unit	1	2	3	4	Average
Y(0)		[0]	0		0
Y(1)	[-3]			1	1
$\hat{\tau}$					1

Missing data

[Footnote: The right way to think about this is that bias is a property of the strategy over possible realizations of data and not normally a property of the estimator conditional on the data.]

Multistage games

Multistage games can also present an endogenous group problem since collections of late stage players facing a given choice have been created by early stage players.

Multistage games

Question: Does **visibility** alter the extent to which subjects follow norms to punish antisocial behavior (and reward prosocial behavior)? Consider a trust game in which we are interested in how information on receivers affects their actions

Table 4: Return rates given investments under different conditions

		% invested	Average % returned	
		(average)	...when 10% invested	...when 50% invested
Treatment	Masked information on respondents	30% (avg)	20%	40%
	Full information on respondents	30% (avg)	0%	60%

What do we think? Does visibility make people react more to investments?

Multistage games

Imagine you could see all the potential outcomes, and they looked like \ this:

Table 5: Potential outcomes with (and without) identity protection

		Responder's return decision (given type)						Avg.
		Nice 1	Nice 2	Nice 3	Mean 4	Mean 4	Mean 6	
Offerer behavior	Invest 10%:	60%	60%	60%	0%	0%	0%	30%
	Invest 50%:	60%	60%	60%	0%	0%	0%	30%

Conclusion: Both the offer and the information condition are **completely irrelevant** for all subjects. . .

Multistage games

Unfortunately you only see a sample of the potential outcomes, and that looks like this:

Table 6: Outcomes when respondent **is visible**

		Responder's return decision (given type)						Avg.
		Nice 1	Nice 2	Nice 3	Mean 4	Mean 4	Mean 6	
Offerer behavior	Invest 10%:				0%	0%	0%	0%
	Invest 50%:	60%	60%	60%				60%

False Conclusion: When not protected, responders condition behavior *strongly* on offers (because offerers can select on type accurately)

Multistage games

Unfortunately you only see a sample of the potential outcomes, and that looks like this:

Table 7: Outcomes when respondent **is not visible**

		Responder's return decision (given type)						Avg.
		Nice 1	Nice 2	Nice 3	Mean 4	Mean 4	Mean 6	
Offerer behavior	Invest 10%:			60%		0%	0%	20%
	Invest 50%:	60%	60%		0%			40%

False Conclusion: When protected, responders condition behavior less strongly on offers (because offerers can select on type less accurately)

Multistage games

What to do? **Solutions?**

1. Analysis *could* focus on the effect of treatment on respondent behavior, directly.
   - This would get the correct answer but to a different question [Does information affect the share of contributions returned by subjects on average? No]
2. **Strategy method** can sometimes help address the problem, **but** that is also (a) changing the question and (b) putting demands on respondent imagination and honesty
3. First mover action could be **directly manipulated**, but unless deception is used that is also changing the question
4. First movers could be **selected** because they act in predictible ways (bordering on deception?)

Idea: Proceed with extreme caution when estimating effects beyond the first stage.

Skip to ▸ Mediation or ◂ Big Ideas

Recap: Ten things you need to know about causal inference

1. A causal claim is a statement about what didn't happen.
2. There is a fundamental problem of causal inference.
3. You can estimate average causal effects even if you cannot observe any individual causal effects.
4. If you know that A causes B and that B causes C, this does not mean that you know that A causes C.
5. The counterfactual model is all about contribution, not attribution.
6. X can cause Y even if there is no "causal path" connecting X and Y.
7. Correlation is not causation
8. X can cause Y even if X is not a necessary condition or a sufficient condition for Y.
9. Estimating average causal effects does not require that treatment and control groups are identical.
10. There is no causation without manipulation

`http://egap.org/resources/guides/causality/`

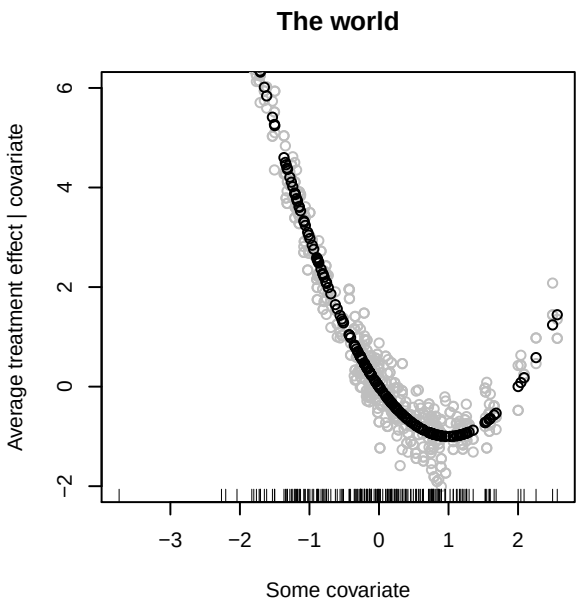
Lecture 2: What's an experiment?

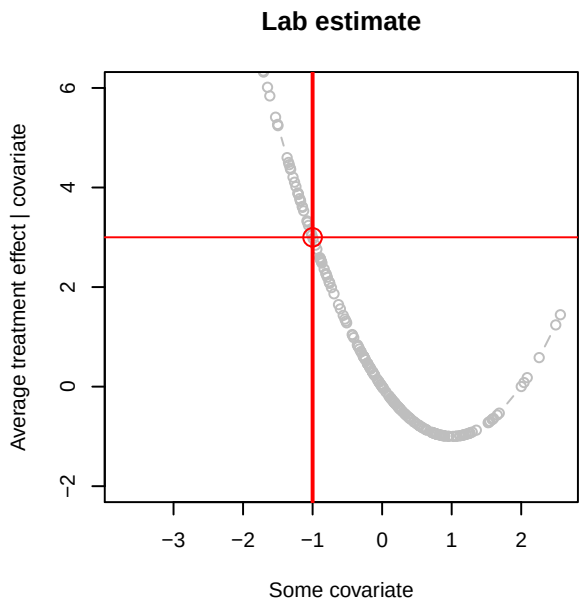
Investigations

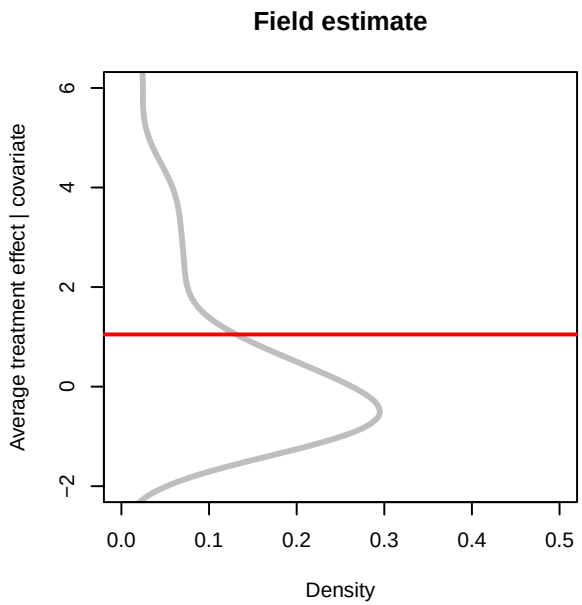
Experiments

- Experiments are investigations in which an intervention, in all its essential elements, is under the control of the investigator. (Cox & Reid)
- Two major types of control:
  1. control over assignment to treatment – this is at the heart of many field experiments
  2. control over the treatment itself – this is at the heart of many lab experiments
- Main focus today is on 1 and on the question: *how does control over assignment to treatment allow you to make reasonable statements about causal effects?*

Experiments







Research designs

Formalizing A Research Design

1. **Causal Variables**. Let D denote a collection of *causal variables*.
   - $\mathcal{V}(D^j)$ denotes the space of possible values for causal variable D^j in D.
   - $a : D \rightarrow 2^D$ denotes the ancestors of variable D^j.

- Three classes of variables:
  - Population (X) with associated probability distribution p_X
  - Manipulands (Z) with associated probability distribution p_Z
  - Outcome Variables (Y): Let $f^j : \times_{D^k \in a(Y_j)} \mathcal{V}(D^k) \rightarrow \mathcal{V}(Y^j)$ denote the potential outcomes function for variable Y^j and f the collection of all such functions.

Formalizing A Research Design

- The **manipulands** are a special class of variables.
  - For every Y^j there is a manipuland Z^j where $\mathcal{V}(Z^j)$ is an augmentation of $\mathcal{V}(Y^j)$ that allows possible values `idle` (or "$\oslash$") for each unit.
  - f^j satisfies:

1. *compliance with controlled assignments*: $y_i^j = z_i^j$ when $z_i^j \neq$ `idle`.
2. *manipuland exclusion*: y^j is independent of z^k for all $k \neq j$ conditional on $a(Y^j) \cap Y$.

So: $f^j(Z^j = z^j, a(Y^j)) = z^j$ whenever $z_i^j \neq \oslash$. When $z_i^j = \oslash$ then y_i^j is unconstrained by z^j and is determined by $a(Y^j)$.

Formalizing A Research Design

Note:

- Our use of `idle` (or $\oslash$) is similar to that in Pearl except that we define variables for collections of units and allow some units to be assigned and others not.
- $Z^j = \vec{\oslash}$ is the case in which Y^j is not directly manipulated in any component and is determined by other ancestors of Y^j
- $Z^j = (0, 1, \oslash)$ is a more complex quantity in which Y^j is directly manipulated for units 1 and 2, but not for unit 3.
- It is still possible however that $f_3^j(Z^j = (0, 1, \oslash)) \neq f_3^j(Z^j = \vec{\oslash})$

Formalizing A Research Design

1. **Causal Variables**
2. **Data**. Let p_D denote a probability distribution over D induced by p_Z and p_X.
   - A realization of the data is denoted d.
   - Let $\mathcal{D}$ denote the set of all possible data (the "**superdata**").
3. **Estimands** Let $\tau(\mathcal{D}, d, p_Z)$ denote an estimand and (τ) a collection of such estimands.
   - An estimand is a summary of potential outcomes, (recorded in $\mathcal{D}$), but it may also depend on which units are sampled or assigned to different conditions (recorded in d), and assignment strategies.
   - A design-independent estimand depends on $\mathcal{D}$ only
4. **Summary statistic** is a function of data and p_Z, these might include test statistics, p values and so on.
5. **Estimator** Let $\hat{\tau}(d, p_Z)$ denote an summary statistic (and $(\hat{\tau})$ a collection of such functions) that is associated with an estimand.

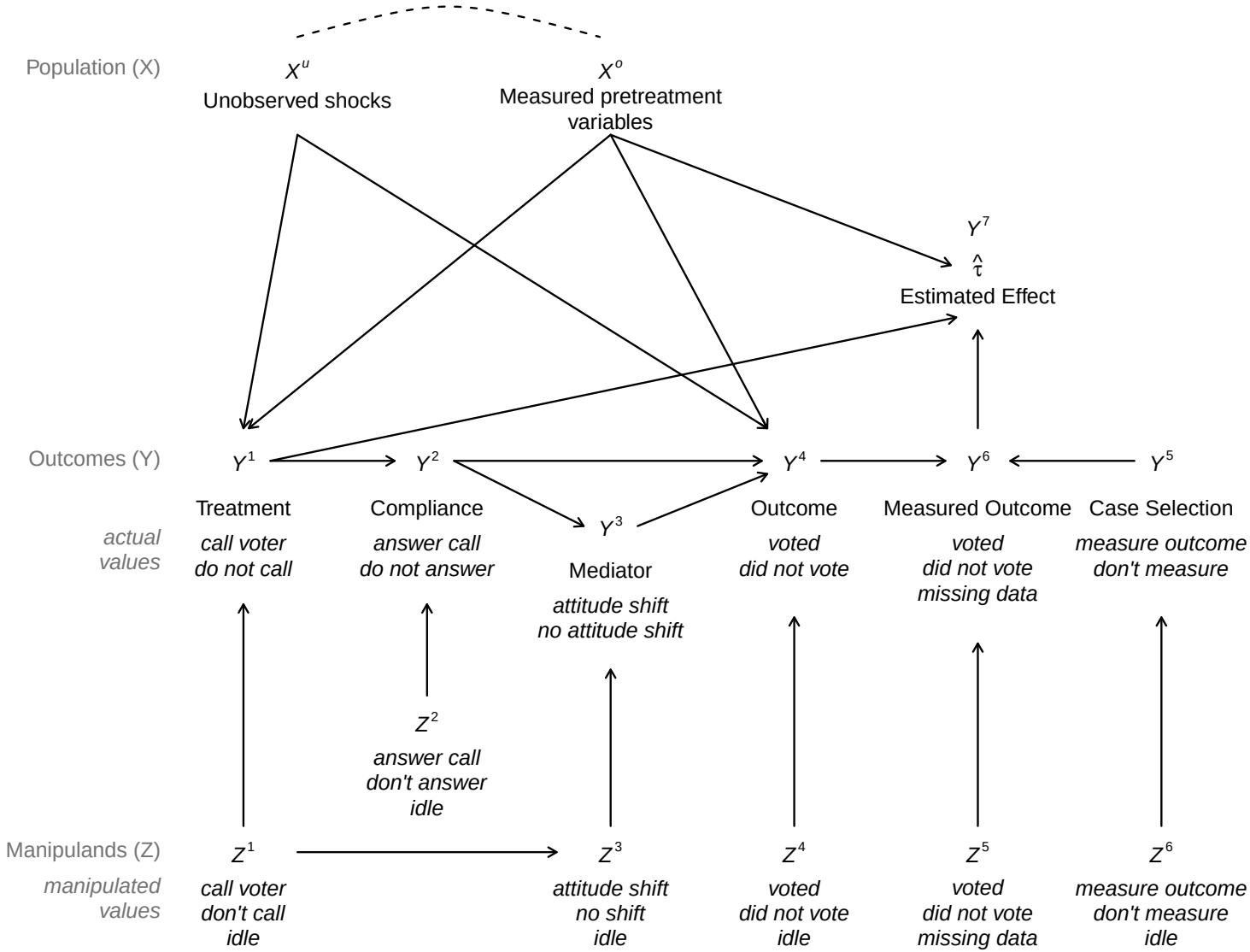
Formalizing A Research Design

5. **Strategy**. Let $\Sigma =< p_Z, (\hat{\tau}) >$ denote a **strategy**
6. **Design**. Let the 4-tuple $\Delta =< p_X, f, \Sigma, (\tau) >$ denote a **design**, it consists of beliefs, conjectures, a strategy, and goals.
7. A **Diagnostic-Statistic**, $t(d, \Delta)$, is a function of Data and the study
   - For example: the difference between $\hat{\tau}$ and τ or whether the p value associated with $\hat{\tau}$ is less than 0.05.
8. A **diagnosand**, θ is a summary of the distribution of a diagnostic statistic. (θ) denotes a collection of diagnosands
   - For example: bias, power.

Formalizing A Complete Research Design

1. We can then say that a design is "θ-complete" if diagnosand θ is calculable from the design.
2. We say that a design is HD-complete when the power, bias, and MSE of the design with respect to estimand τ can be calculated.

Two illustrations: As a DAG



Two illustrations: Simple Case

	p_{Z^1}	Z^1		Y^1		Y^2		$Y^3 = \widehat{\tau}$		t
Assignment Index	Probability of Assignment	Manipuland		Actual Assignment		Potential Outcome		Estimate Diff.-in-means		Diagnostic – Stat. $\widehat{\tau} - \tau$
1.	$\frac{1}{6}$	$\begin{bmatrix}1\\0\\\oslash\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}1\\0\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}4\\4\\4\end{bmatrix}$	$\rightarrow$	0	$\rightarrow$	-4
2.	$\frac{1}{6}$	$\begin{bmatrix}1\\\oslash\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}1\\1\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}4\\4\\0\end{bmatrix}$	$\rightarrow$	4	$\rightarrow$	0
3.	$\frac{1}{6}$	$\begin{bmatrix}\oslash\\1\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}1\\1\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}4\\4\\0\end{bmatrix}$	$\rightarrow$	4	$\rightarrow$	0
4.	$\frac{1}{6}$	$\begin{bmatrix}\oslash\\0\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}1\\0\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}4\\4\\4\end{bmatrix}$	$\rightarrow$	0	$\rightarrow$	-4
5.	$\frac{1}{6}$	$\begin{bmatrix}0\\1\\\oslash\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\1\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\4\\0\end{bmatrix}$	$\rightarrow$	4	$\rightarrow$	0
6.	$\frac{1}{6}$	$\begin{bmatrix}0\\\oslash\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\1\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\4\\4\end{bmatrix}$	$\rightarrow$	4	$\rightarrow$	0
7.	0	$\begin{bmatrix}1\\0\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}1\\0\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}4\\0\\0\end{bmatrix}$	$\rightarrow$	4	$\rightarrow$	0
8.	0	$\begin{bmatrix}0\\1\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\1\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\4\\0\end{bmatrix}$	$\rightarrow$	4	$\rightarrow$	0
9.	0	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\0\\1\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}4\\4\\4\end{bmatrix}$	$\rightarrow$	0	$\rightarrow$	0
10.	0	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\rightarrow$	$\begin{bmatrix}0\\0\\0\end{bmatrix}$	$\rightarrow$	NaN	$\rightarrow$	NaN
								$\mathbb{E}(\widehat{\tau}) = \frac{16}{6}$		$\theta_{\text{bias}} = \mathbb{E}(t) = -\frac{8}{6}$

Table 1: Ten possible assignments of Z^1. Estimand $\tau = 4$

Things to note from this formalization

1. Blocks and clusters are part of a design, p_Z, not part of the world (X)
   - There may be all sorts of dependencies in the generation of potential outcomes but the blocks and clusters that are used for randomization are decided by researchers.

Things to note from this formalization

2. Estimands need to be expressed in terms of full assignments. Otherwise it can, for example, be impossible to assess how biased the estimate of an estimand is in the presence of spillovers.
   - We often write the ATE as $E_i(y_i(1) - y_i(0))$, but this estimand *by definition* excludes spillovers.
   - Instead we define ATE as: $E_i(y_i(Z^{1(i)}) - y_i(Z^0))$ where $Z^{1(i)}$ denotes the assignment in which only unit 1 is treated and all other units are in control and Z^0 is the assignment in which all units are in control, i.e. $(\vec{0})$.
   - More generally: model dependent estimands may make it difficult to assess performance of estimators—avoid model dependency in formal declaration

Things to note from this formalization

3. Estimands should not be determined by estimators, though they may be inspired by them.
   - Sometimes researchers choose an estimator (e.g. regression, IV) and define the estimand in terms of whatever it is that the estimator "shoots" at.
   - But the mapping from estimators to estimands is not onto; it is not even a function. A single estimator may shoot at different estimands and a different estimand can be estimated by different estimators (in fact, by *any* estimator, though generally badly)

Things to note from this formalization

3. Estimands should not be determined by estimators, though they may be inspired by them.
    - How good an estimator is for an estimand depends on other features of a design, such as assignment schemes
    - One can define estimands of the form "the expected $\hat{b}$ that would be returned from regression given f_Z" or even "the average $\hat{b}$ that would be returned from averaging each unit's potential outcomes on potential assignments"

Things to note from this formalization

4. Noncompliance is just a potential outcome
   - as are attrition, spillovers, and even data collection, and estimates, and referee reports

Clarity regarding distinct types of estimator and estimand

Simple ways to formalize important differences between classes of estimand and classes of estimates:

- a post-treatment estimand
- a post-sampling estimand
- a population-estimand
- a design-independent estimand

Designs in Code: DeclareDesign

DeclareDesign: The key idea

A design consists of six objects:

1. **The population**. The set of units about which inferences are sought.
2. **The potential outcomes function**. The outcomes that each unit might exhibit depending on how the causal process being studied changes the world.
3. **The sampling strategy**. The strategy used to select units to include in the study sample.
4. **The estimands**. The specification of the things that we want to learn about the world, described in terms of potential outcomes.
5. **The assignment function**. The manner in which units are assigned to reveal one potential outcome or another.
6. **The estimator function**. The procedure for generating estimates of quantities we want to learn about.

DeclareDesign

Note:

- For experimental research, several components are entirely within the control of researchers (3, 4, 5, 6)
- Some are not (1 and 2) though they still matter for assessing the design.
- For observational work, item 5 (the assignment function) is generally not in the control of researchers, but even in observational research, assumptions about the assignment processes are typically invoked at least implicitly.
- Characteristics of populations (1) as well as potential outcomes (2) may be a matter of speculation.
- Explicit statements of beliefs about these are needed however to *evaluate* a design, in the same way as statements about effect sizes are needed to conduct simple power analyses.

DeclareDesign Workflow}

(1) **Design** $\leftrightarrow$ (2) **Diagnose** $\rightarrow$ (3) **Register** $\rightarrow$ (4) **Sample & Assign** $\rightarrow$ (5) Implement $\rightarrow$ (6) Measure $\rightarrow$ (7) **Report & Reconcile**

How can we exactly characterize each step?

- Through computer code
- Our implementation: `R` package `DeclareDesign` (with Graeme Blair, Jasper Cooper and Alex Coppock)

Six "declarations" for each design element

```
library(DeclareDesign)

population <- declare_population()
pos        <- declare_potential_outcomes()
sampling   <- declare_sampling()
assignment <- declare_assignment()
estimand   <- declare_estimand()
estimator  <- declare_estimator(estimand = estimand)

my_design  <- declare_design(population = population,
                             sampling = sampling,
                             potential_outcomes = pos,
                             assignment = assignment,
                             estimator = estimator
)
```

Declare Design

```
population    <- declare_population(
  noise = "rnorm(n_)",
  income_bracket = "sample(1:4, n_, replace = TRUE)",
  size = 5000)
```

The declare_population function is extremely flexible and can be used to rapidly generate complex dummy data structures.

Step 1: Design

```
potential_outcomes <- declare_potential_outcomes(
  condition_names = c("control", "treatment"),
  formula =
    Y ~ .1 * (Z == "treatment") +
    .1 * (Z == "treatment") * income_bracket +
    noise)
```

Multiple potential outcomes can be defined, e.g. for compliance or attrition.

Step 1: Design

```
sampling <- declare_sampling(n = 100)
```

Flexible built in options for sampling, though arbitrary user functions can also be employed.

Step 1: Design

```
assignment <- declare_assignment(
  potential_outcomes  = potential_outcomes,
  block_variable_name = "income_bracket",
  block_probabilities = rbind(c(.1, .9),
                              c(.3, .7),
                              c(.7, .3),
                              c(.6, .4)))
```

Flexible built in options for assignment, though arbitrary user functions can also be employed.

Step 1: Design

```
estimand_ATE <- declare_estimand(
  estimand_text = "mean(Y_Z_treatment - Y_Z_control)",
  potential_outcomes = potential_outcomes)
```

Estimands are functions of potential outcomes stored in superdata.

Step 1: Design

```
estimator_lsdv <- declare_estimator(
  formula             = Y ~ Z + factor(income_bracket),
  model               = lm,
  estimates           = get_regression_coefficient,
  coefficient_name    = "Ztreatment",
  estimand            = estimand_ATE,
  labels              = "LSDV")
```

Estimands are associated with estimators when estimators are declared.

Step 1: Design

This set of R objects formally characterizes the entire design.

```
my_design <- declare_design(
  population          = population,
  potential_outcomes = potential_outcomes,
  sampling            = sampling,
  assignment          = assignment,
  estimator           = estimator_lsdv
)
```

Once done my_design is an object that can be posted and shared. It can also be modified, interrogated, and used in many ways.

Step 1: Design

For simple designs this can all be done in one step using a "quick design" template which lets you declare a standard design type in one step.

```
another_design <- quick_design(template = simple_template,
                               N=20, n = 10)
```

Or designs may be modifications of existing designs.

```
another_design <- modify_design(my_design,
                                estimator = estimator_robust)
```

Step 2: Use the design

- Draw dummy data
- Implement sampling and assignment
- Implement analysis
- Report and reconcile

Use the design: Draw Data (Sample and Assign)

Once a design is declared it can be used and interrogated. For example the design has sufficient information to create mock data. These can be used to confirm design features and simulate analysis strategies. `draw_data` functionality can also be used to sample and assign treatments.

```
mock_data <- draw_data(design = my_design)
```

Y_control	Y_treat	Z	Y	noise	income_bracket
-0.642	-0.342	control	-0.642	-0.642	2
1.670	1.970	treatment	1.970	1.670	2
1.816	2.116	treatment	2.116	1.816	2
-0.512	-0.312	treatment	-0.312	-0.512	1
2.661	2.961	treatment	2.961	2.661	2
1.156	1.356	treatment	1.356	1.156	1

Use the design: Implement Analysis

The design also contains the information needed to implement analysis on either real or simulated data.

```
get_estimates(estimator = my_design$estimator,
              data      = mock_data)
```

estimate_label	est	se	p	ci_lower	ci_upper	df	estimator_l
Ztreatment	0.26	0.24	0.27	-0.21	0.74	95	estimator_l

Use the design: Diagnose

Perhaps most importantly, with a design declared it can be diagnosed.

```
diagnose_design(my_design)
```

diagnosand_label	diagnosand	diagnosand_sd_boot
mean(estimand)	0.350	0.000
mean(estimate)	0.370	0.013
sd(estimate)	0.229	0.010
bias	0.020	0.013
RMSE	0.230	0.010
coverage	0.950	0.012
power	0.352	0.025
type S rate	0.058	0.014

Iteration between steps 1 and 2 to improve the design

```
sampling_larger_sample <- declare_sampling(n = 500)
```

```
my_other_design <- modify_design(
  design   = my_design,
  sampling = sampling_larger_sample)
diagnose_design(my_other_design)
```

diagnosand_label	diagnosand	diagnosand_sd_boot
mean(estimand)	0.350	0.000
mean(estimate)	0.386	0.007
sd(estimate)	0.107	0.005
bias	0.036	0.007
RMSE	0.113	0.005
coverage	0.935	0.016
power	0.962	0.012

Group Work 1

Small group work on causal inference

Lecture 3: How to Randomize

Basic randomization

Top

- Basic randomization is very simple. For example, say you want to assign 5 of 10 units to treatment. Here is simple code:

```
1:10 %in% sample(1:10, 5)
```

```
[1] FALSE FALSE TRUE TRUE FALSE TRUE TRUE FALSE TRUE FALSE
```

. . . should be replicable

In general you might want to set things up so that your randomization is **replicable**. You can do this by setting a **seed**:

```
set.seed(20111112)
1:10 %in% sample(1:10, 5)
```

```
[1] TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE
```

```
set.seed(20111112)
1:10 %in% sample(1:10, 5)
```

```
[1] TRUE TRUE FALSE FALSE TRUE FALSE FALSE TRUE FALSE TRUE
```

Basic randomization

Even better is to set it up so that it can reproduce **lots of possible draws** so that you can check the propensities for each unit.

```
set.seed(20111112)
P <- sapply(1:1000, function(i) 1:10 %in% sample(1:10, 5))
apply(P, 1, mean)
```

[1] 0.525 0.486 0.502 0.500 0.511 0.491 0.485 0.484 0.501 0.515

Here the P matrix gives 1000 possible ways of allocating 5 of 10 units to treatment. We can then confirm that the average propensity is 0.5.

- A huge advantage of this approach is that if you make a mess of the random assignment; **you can still generate the P matrix and use that for all analyses**!

Do it in advance

- Unless you need them to keep subjects at ease, leave your spinners and your dice and your cards behind.
- Especially when you have multiple or complex randomizations you are generally much better doing it with a computer in advance

	IDENTIF	VARIATIONS														
Order	ID	V1 SURVEY MP	V1 NAME OF SURVEY MP	V1 PARTY	V2 INTIMIDATIO	V3 FRAUD	V4 ELEMT	V5 CARD	V5 SCORE TYPE	V5 SCORE	V5 ENDORSE	V5 WHICH ENDORSE	V6 SMS	V6 PRICE	V6 SEE	V7 FIND	V7 MP LCV
1	573411	Constituency	Musumba Isaac Isanga	NRM	A	D	Yes	Yes	CONSTIT	E	No		No		.	Yes	LCV
2	573412	Constituency	Musumba Isaac Isanga	NRM	B	D	No	Yes	PLENY	E	Yes	G	Yes	50 Sh	Yes	No	
3	573422	Constituency	Musumba Isaac Isanga	NRM	B	C	Yes	Yes	PLENY	E	No		Yes	Full Price	Yes	No	
4	573421	Constituency	Musumba Isaac Isanga	NRM	A	C	No	Yes	CONSTIT	E	Yes	C	No		.	No	
1	717221	Women	Alitwala Kadaga Rebecca	NRM	B	D	No	Yes	CONSTIT		No		No		.	Yes	WOM
2	717211	Women	Alitwala Kadaga Rebecca	NRM	A	D	Yes	Yes	PLENY		Yes	E	Yes	Free	No	No	
3	717212	Constituency	Balikowa Henry	NRM	B	C	Yes	No	.		.		No		.	No	
4	717222	Constituency	Balikowa Henry	NRM	A	C	No	No	.		.		Yes	Free	No	No	
1	717421	Women	Alitwala Kadaga Rebecca	NRM	A	C	Yes	Yes	CONSTIT		No		Yes	Full Price	Yes	Yes	WOM
2	717412	Women	Alitwala Kadaga Rebecca	NRM	B	C	No	Yes	PLENY		Yes	A	No		.	No	
3	717411	Constituency	Balikowa Henry	NRM	B	D	Yes	Yes	CONSTIT	C	Yes	D	No		.	No	

Figure 1: A survey dictionary with results from a complex randomization presented in a simple way for enumerators

Did the randomization "work"?

- People often wonder: did randomization work?
- Common practice is to implement a set of t-tests to see if there is balance
- This makes no sense.

- If you doubt whether it was **implemented** properly do an F test
- If you worry about **variance** specify controls in advance as a function of relation with outcomes (more on this later)
- If you worry about **conditional bias** then look at substantive differences between groups, not t–tests

Cluster Randomization

Cluster Randomization

- Simply place units into groups (clusters) and then randomly assign the groups to treatment and control.
- All units in a given group get the same treatment

Note: clusters are part of your design, not part of the world.

Cluster Randomization

- Often used if intervention has to function at the cluster level *or* if outcome defined at the cluster level.
- **Disadvantage:** loss of statistical power
- However: perfectly possible to assign *some* treatments at cluster level and then **other** treatments at the individual level

- **Principle:** (unless you are worried about spillovers) generally make clusters as small as possible
- **Principle:** Surprisngly, variability in cluster size makes analysis harder. (See analysis section)
- **Be clear** about whether you believe effects are operating at the cluster level or at the individual level. This matters for power calculations.
- **Be clear** about whether spillover effects operate only within clusters or also across them. If within only you might be able to interpret treatment as the effect of being in a treated cluster...

Cluster Randomization: Block by cluster size

Surprisingly, if clusters are of different sizes the difference in means estimator is *not* unbiased, even if all units are assigned to treatment with the same probability. **Here's the intuition.** Say there are two clusters each with homogeneous

treatment effects:

Cluster	Size	Y0	Y1
1	1000000	0	1
2	1	0	0

Then:

- What is the true average treatment effect?
- What do you expect to estimate from cluster random assignment?

The solution is to block by cluster size. For more see:
`http://gking.harvard.edu/files/cluster.pdf`

Blocked assignments and other restricted randomizations

Blocking

There are more or less **efficient** ways to randomize.

- Randomization helps ensure good balance on all covariates (observed and unobserved) *in expectation*.
- But balance may not be so great *in realization*
- Blocking can help ensure balance ex post on observables

Consider a case with four units and two strata. There are 6 possible assignments of 2 units to treatment:

ID	X	Y(0)	Y(1)	R1	R2	R3	R4	R5	R6
1	1	0	1	1	1	1	0	0	0
2	1	0	1	1	0	0	1	1	0
3	2	1	2	0	1	0	1	0	1
4	2	1	2	0	0	1	0	1	1
$\widehat{\tau}$:				0	1	1	1	1	2

Even with a constant treatment effect and everything uniform within blocks, there is variance in the estimation of $\widehat{\tau}$. This can be eliminated by excluding

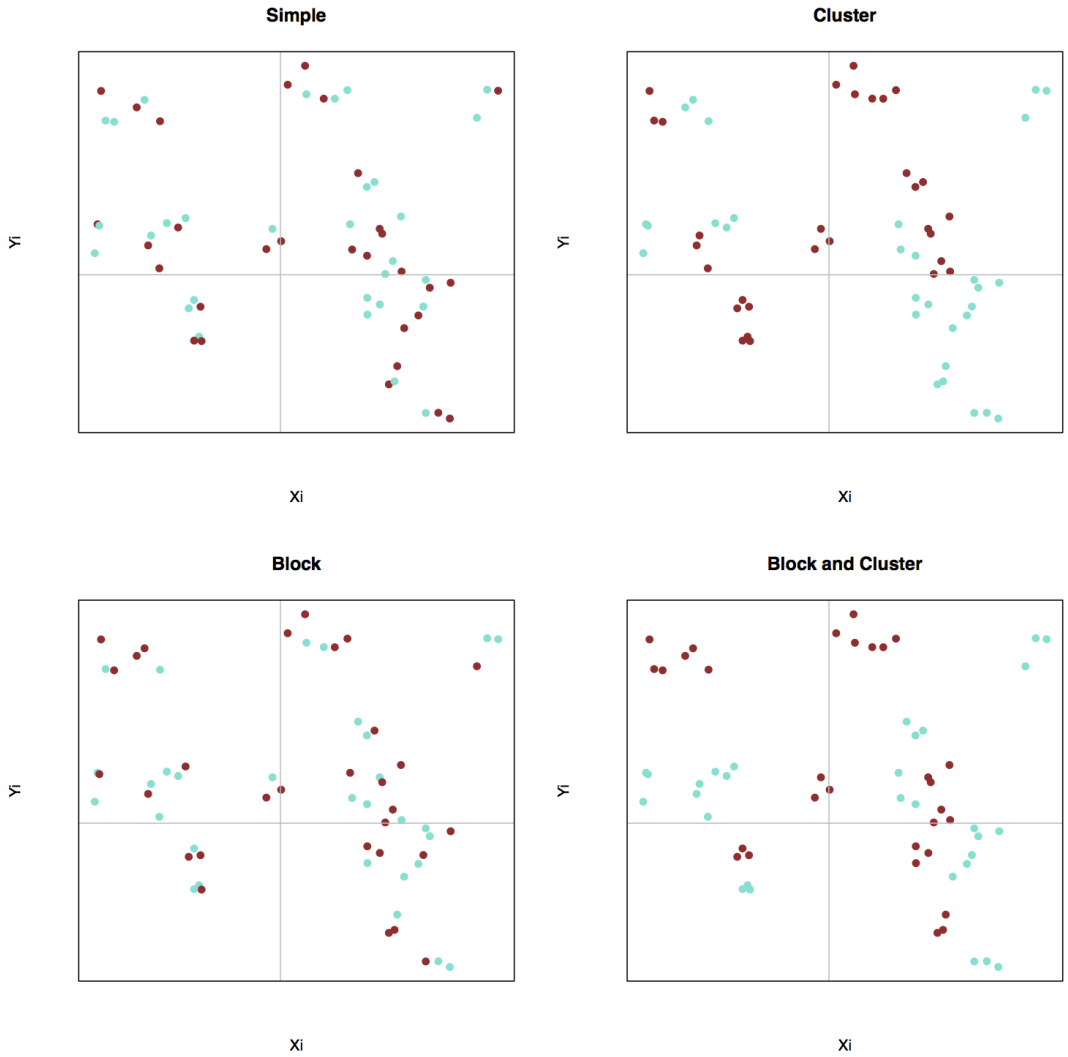
Blocking

Simple blocking in R (5 pairs):

```
sapply(1:5, function(i) rank(runif(2))==1)
```

1	2	3	4	5
FALSE	TRUE	FALSE	TRUE	FALSE
TRUE	FALSE	TRUE	FALSE	TRUE

Of blocks and clusters



Blocking

- Blocking is a case of **restricted randomization**. Although each unit is sampled with equal probability, the *profiles* of possible assignments are not.
- You have to take account of this when doing analysis
- There are many other approaches.
  - "**Matched Pairs**" are a particularly fine approach to blocking
  - You could also randomize and then **replace the randomization** if you do not like the balance. This sounds tricky (and it is) but it is OK as long as you understand the true lottery process you are employing and incorporate that into analysis
  - It is even possible to block on **covariates for which you don't have data** ex ante, by using methods in which you allocate treatment over time as a function of features of your sample (also tricky)

Other types of restricted randomization

- Really you can set whatever criterion you want for your set of treated units to have (eg no treated unit beside another treated unit; at least 5 from the north, 10 from the south, guaranteed balance by some continuous variable etc)
- You just have to be sure that you understand the random process that was used and that you can use it in the analysis stage
- But here be dragons
  - The more complex your design, the more complex your analysis.
  - General injunction (Senn 2004 "as ye randomize so shall ye analyze")
  - In general you should make sure that a given randomization procedure coupled with a given estimation procedure will produce an unbiased estimate. `DeclareDesign` can help with this.

Factorial Designs

Factorial Designs

- Often when you set up an experiment you want to look at more than one treatment.
- Should you do this or not? How should you use your power?

Factorial Designs

- Often when you set up an experiment you want to look at more than one treatment.
- Should you do this or not? How should you use your power?

	$T2 = 0$	$T2 = 1$
$T1 = 0$	50%	0%
$T1 = 1$	50%	0%

	$T2 = 0$	$T2 = 1$
$T1 = 0$	25%	25%
$T1 = 1$	25%	25%

	$T2 = 0$	$T2 = 1$
$T1 = 0$	33.3%	33.3%
$T1 = 1$	33.3%	0%

Factorial Designs

- Surprisingly adding multiple treatments does not eat into your power (unless you are decomposing a complex treatment – then it can. Why?)
- Especially when you use a fully crossed design like the middle one above.
- Fisher: "No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time. The writer is convinced that this view is wholly mistaken."
- However – adding multiple treatments *does* alter the **interpretation** of your treatment effects. If T2 is an unusual treatment for example, then half the T1 effect is measured for unusual situations.

Factorial Designs: In practice

- In practice if you have a lot of treatments it can be hard to do full factorial designs – there may be too many combinations.
- In such cases people use **fractional factorial designs**, like the one below (5 treatments but only 8 units!)

Variation	T1	T2	T3	T4	T5
1	0	0	0	1	1
2	0	0	1	0	0
3	0	1	0	0	1
4	0	1	1	1	0
5	1	0	0	1	0
6	1	0	1	0	1
7	1	1	0	0	0
8	1	1	1	1	1

- Then randomly assign units to rows. Note columns might also be blocking covariates.
- In R, look at `library(survey); hadamard(7)`

Factorial Designs: In practice

- But be careful: you have to be comfortable with possibly not having any simple counterfactual unit for any unit (invoke sparsity-of-effects principle).

Unit	T1	T2	T3	T4	T5
1	0	0	0	1	1
2	0	0	1	0	0
3	0	1	0	0	1
4	0	1	1	1	0
5	1	0	0	1	0
6	1	0	1	0	1
7	1	1	0	0	0
8	1	1	1	1	1

- In R, look at `library(survey); hadamard(7)`

External Validity: Can randomization strategies help?

Principle: Address **external validity** at the design stage

Anything to be done on randomization to address external validity concerns?

- **Note 1**: There is little or nothing about field experiments that makes the external validity problem greater for these than for other "sample based" research
- **Note 2**: Studies that use up the available universe (cross national studies) actually have a distinct external validity problem
- Two ways to think about external validity issues:
  1. Are things likely to operate in other units like they operate in these units? (even with the same intervention)
  2. Are the processes in operation in this treatment likely to operate in other treatments? (even in this population)

Principle: Address **external validity** at the design stage

- Two ways to think about external validity issues:
  1. Are things likely to operate in other units like they operate in these units? (even with the same intervention) 2.Are the processes in operation in this treatment likely to operate in other treatments? (even in this population)
- Two approaches for 1.
  - Try to sample cases and estimate *population average treatment effects*
  - Exploit internal variation: block on features that make the case unusal and assess importance of these (eg is unit poor? assess how effects differ in poor and wealthy components)
- 2 is harder and requires a sharp identification of context free primitives, if there are such things.

Power

Randomization and power

- Power is the probability that you will correctly reject an incorrect null.
- For example if your null is that the average treatment effect is 0 and you reject the null whenever you estimate a t-stat$>$1.96, then your power is the probability that you willindeed estiamte a t-stat greater than 1.96.
- But to know that you have to speculate about some true data generating process.
- **Question:** Say you had a design in which you said: "I will reject the null no matter what the data looks like." What is your power?

Randomization and power

- So power can be silly. But it is normally understood in a framework in which you reject the null on the basis of some p value and the p value has the interpretation that under the null the data pattern you see (or some more extreme pattern) would arise with probability p.
- Many factors influence power. What are
- **Question:** What are the most important factors that influence power?

Randomization and power

1. The variance of outcomes in the treatment and control groups – the lower the variance the greater the power
2. The number of observations you have in the treatment adn control groups
    - These two are related and you can control them—you want the data to be where it will be most effective. If the variance in the treatment and control groups is the same then you should spread the data across these groups (equal assignment to treatment); but otherwise, not
    - You can also control the variance of outcomes in your groups with your assignment strategy (eg blocking)

Also:

3. The true effect size – the more different the true effect from the null the easier it is to reject the null. You can control the true effect size (dosage) and you can choose the null.
4. Your test statistic. This is generated normally by your estimation strategy – some strategies are better than others for reducing variation

Randomization and power

If you have:

- figured out your design—including the sample, estimation strategy, and tests you will focus on
- conjectures about the world

Then you can calculate power.

Illustration of power analysis in DeclareDesign

`DeclareDesign` can act as a big power calculator but one in which you can compare arbitrary aspects of the design.

- Here this is done by declaring a sequence of designs and diagnosing each.

```
ns      <- seq(20, 100, 20)
designs <- lapply(ns,  FUN = function(n)
           quick_design(template = simple_template, n = n))
powers  <- lapply(compare_designs2(design = designs),
           get_stat, stat = "power")
```

Illustration of power analysis in DeclareDesign

DeclareDesign can act as a big power calculator but one in which you can compare arbitrary aspects of the design.

```
## Error in xy.coords(x, y, xlabel, ylabel, log): 'x' and 'y'
```

But please look past power

But done this way you can see that you can query not just arbitrary features of the design but also arbitrary diagnosands. Such as RMSE. frand_dis

```
ns      <- seq(20, 100, 20)
designs <- lapply(ns,  FUN = function(n)
            quick_design(template = simple_template, n = n))
rmses   <- lapply(compare_designs2(design = designs),
            get_stat, stat = "RMSE")
```

But please look past power

```
## Error in xy.coords(x, y, xlabel, ylabel, log): 'x' and 'y'
```

Maintaining balance

Principle: Maintain Balance Everywhere

- Obvious as it may seem, it is critical to ensure that no element of a design is accidentally correlated with treatment
- Principle is violated eg if:
  - Your design is complex: eg you want to see the effect of a political endorsement but the treatment provides exposure both to the endorsement and the more general fact that there is a campaign
  - You take more measurements of subjects in treatment than in control
  - Your delivery of treatment selects subpopulations of subjects (those with phones?) in ways different to control populations
  - You use different survey teams in the treatment and control groups
  - You take measurements for treatment and control at different times

Recap: Ten things you need to know about randomization

1. Some ways are better than others
2. Block randomization: You can fix it so that treatment and control groups are balanced
3. Factorial designs: You can randomize multiple treatments at the same time without using up power
4. You can randomize whole clusters together (but the bigger your clusters, the weaker your power!)
5. You can randomize in a way that makes it easier to see if there are spillovers
6. Different units can be assigned to treatment with different probabilities
7. Restricted randomization: If you don't like what you get you can start over
8. Write randomization code that lets you simulate many possible randomizations
9. You can do randomization as you go along
10. Randomization can sometimes be an ethical way of assigning a treatment, but sometimes it isn't

`http://egap.org/resources/guides/randomization/`

Lecture 3b: Assignments with DeclareDesign

Complete, block, cluster, more

A design: Multilevel data

We will declare a design with hierarchical data and consider a range of assignment strategies.

Population

```
N_per_level <- c(students = 500,
                 classrooms = 50, schools = 25)

population <- declare_population(
  student = list(
    noise_student = declare_variable(type = "normal",
                     location_scale = c(mean = 0, sd = 2)
                     )),
  classroom = list(
     noise_classroom = declare_variable(type = "normal",
                     location_scale = c(mean = 0, sd = 2)
                     )),
  school = list(),
  size = N_per_level)
```

Sampling and potential outcomes

```
potential_outcomes <- declare_potential_outcomes(
  condition_names = c("control", "treatment"),
  formula =
    Y ~ 50 + 5*(Z == "treatment") +
        2 * (classroom_ID/10 - 2) +
        2 * (school_ID/10 - 2) +
        noise_classroom +
        noise_student)
```

Samplng and Assignment

```
sampling <- declare_sampling(sampling = FALSE)

my_assignment <- function(data){
                 rbinom(n = nrow(data), size = 1, prob = 0.5)}

assignment <- declare_assignment(
                 custom_assignment_function = my_assignment,
                 potential_outcomes = potential_outcomes)
```

Estimators and Estimands

```
estimand <- declare_estimand(
    estimand_text = "mean(Y_Z_treatment - Y_Z_control)",
    potential_outcomes = potential_outcomes)

estimator <- declare_estimator(Y ~ Z,
        estimates = difference_in_means,
        estimand = estimand,
        labels = "ATE")
```

Design declaration

Now, we are ready to declare our design and draw a sample data set.

```
design <- declare_design(
                population = population,
                sampling = sampling,
                potential_outcomes = potential_outcomes,
                assignment = assignment,
                estimator = estimator)
```

Sample data

Here are the first couple of rows and columns of the resulting data frame.

```
set.seed(1:5)
my_data <- draw_data(design)
kable(head(round(my_data,2))[,1:5])
```

student_ID	classroom_ID	school_ID	noise_classroom	noise_student
1	1	1	0.15	-1.25
2	1	1	0.15	0.37
3	1	1	0.15	-1.67
4	1	1	0.15	3.19
5	1	1	0.15	0.66
6	1	1	0.15	-1.64

Sample data

Here is the distribution between treatment and control:

```
kable(t(as.matrix(table(my_data$Z))),
      col.names = c("control", "treatment"))
```

control	treatment
266	234

Complete Random Assignment using the built in function

```
assignment_complete <- declare_assignment(m = 250,
                                          potential_outcomes = potential_outcome

design_complete <- modify_design(design,
                                 assignment = assignment_complete)
```

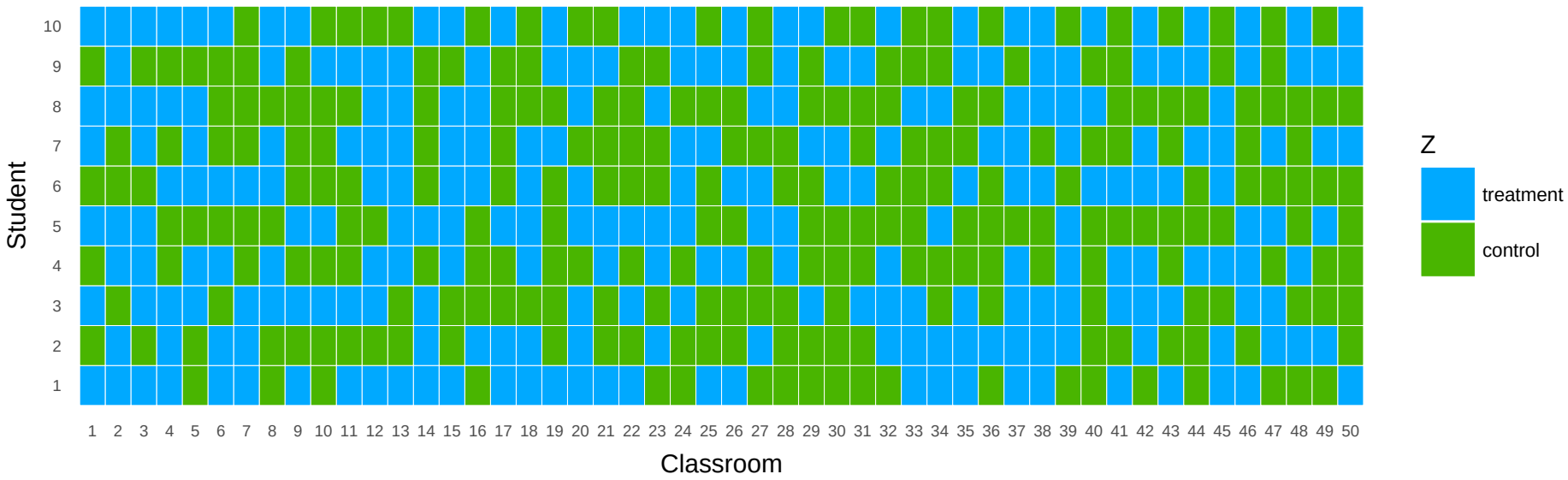
Data from complete assignment

We can draw a new set of data and look at the number of subjects in the treatment and control groups.

```
set.seed(1:5)
data_complete <- draw_data(design_complete)

kable(t(as.matrix(table(data_complete$Z))))
```

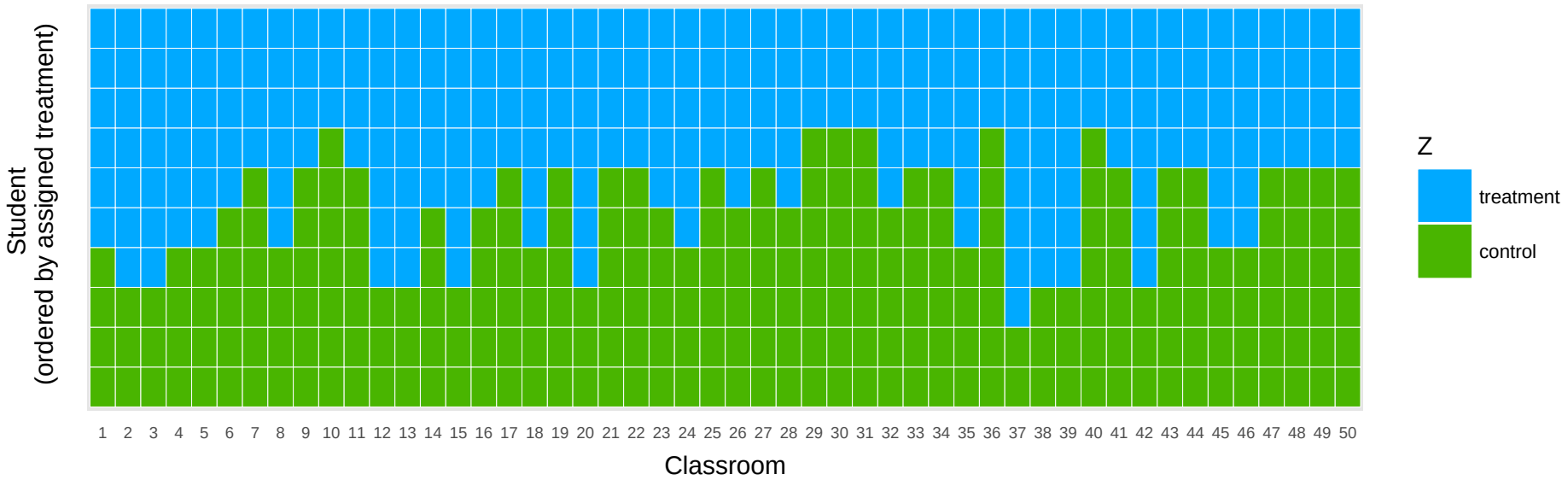
control	treatment
250	250

Plotted


Student
10
9
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50
Classroom
Z
treatment
control

Block Random Assignment

- The treatment and control group will **in expectation** contain the same share of students in different classrooms.
- But as seen in the plot below this does necessarily hold in **realization**
- We make this more obvious by sorting the students by treatment status with schools



Declaring a blocked design

```
assignment_blocked   <- declare_assignment(
  potential_outcomes = potential_outcomes,
  block_variable_name = "classroom_ID")

design_blocked <- modify_design(design,
                          assignment = assignment_blocked)

data_blocked <- draw_data(design_blocked)
```

Illustration of blocked assignment

- Note that subjets are sorted here after the assignment to make it easier to see that in this case blocking ensures that exactly 5 students within each classroom are assigned to treatment.

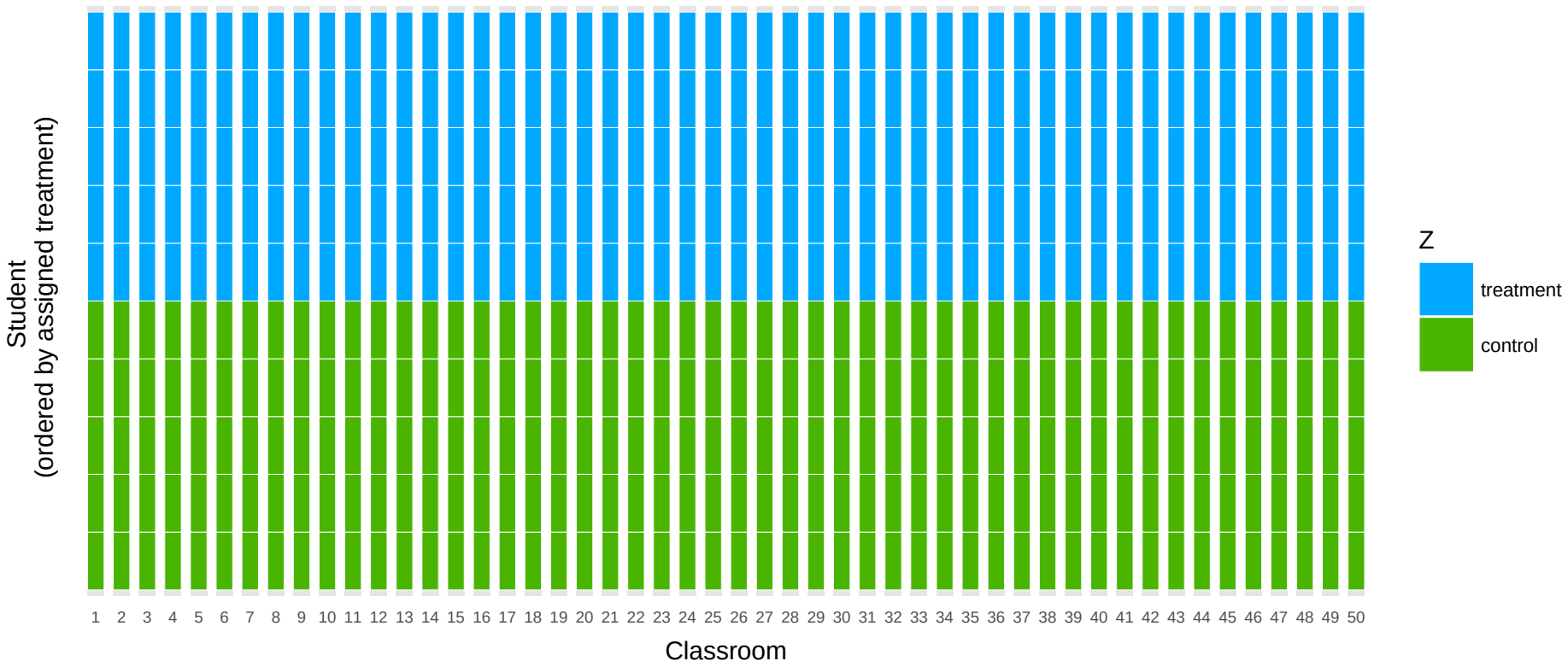
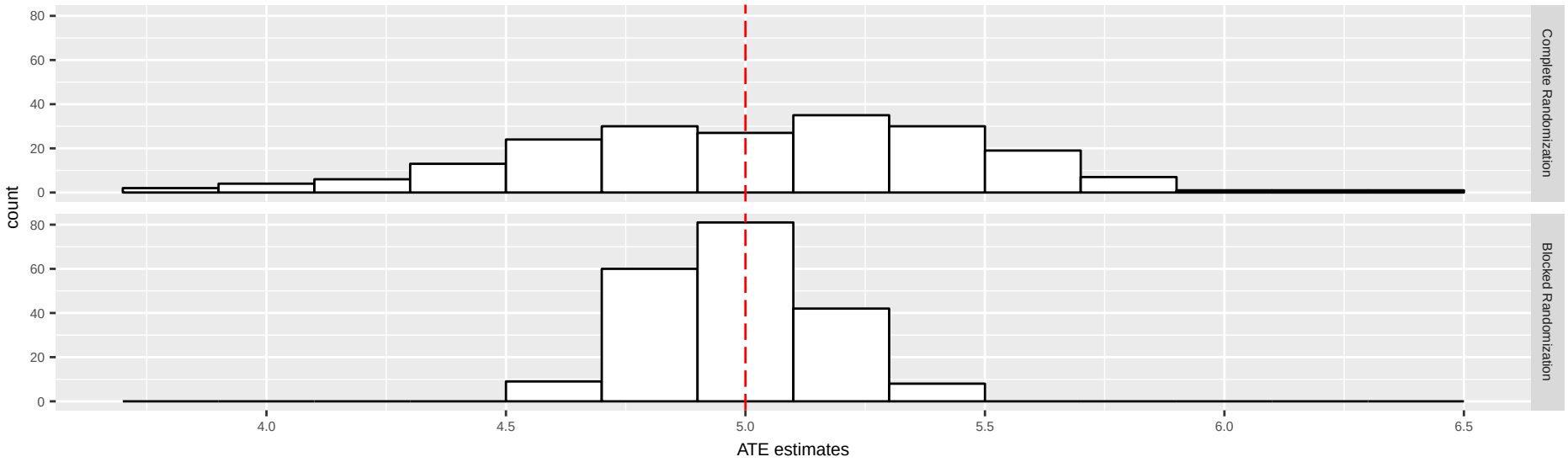


Illustration of efficiency gains from blocking

```
set.seed(1:3)
diagnosis <- diagnose_design(design_complete,
                             population_draws = 1,
                             sample_draws = 1,
                             assignment_draws = 200)

diagnosis_blocked <- diagnose_design(design_blocked,
                             population_draws = 1,
                             sample_draws = 1,
                             assignment_draws = 200)
```

Illustration of efficiency gains from blocking

```
rand_dist <- diagnosis$simulations$est
rand_dist_blocking <- diagnosis_blocked$simulations$est
```



Lecture 4: Analysis

Basic Analysis

ATE

Unbiased estimates of the (sample) average treatment effect can be estimated (**whether or not there imbalance on covariates**) using:

$$\widehat{ATE} = \frac{1}{n_T}\sum_{T} Y_i - \frac{1}{n_C}\sum_{C} Y_i,$$

Say different strata or blocks $\mathcal{S}$ had different assignment probabilities. Then you could estimate:

$$\widehat{ATE} = \sum_{S \in \mathcal{S}} \frac{n_S}{n}\left(\frac{1}{n_{S1}}\sum_{S \cap T} y_i - \frac{1}{n_{S0}}\sum_{S \cap C} y_i\right) \quad (1)$$

Which also corresponds to the difference in the weighted average of treatment outcomes (with weights given by the inverse of the probability that each unit is assigend to treatment) and control outcomes (with weights given by the inverse of the probability that each unit is assigend to control).

ATE with IPW

- The average difference in means estiamator is the same as what you would get if you weighted inversely by shares of units in different conditions inside blocks.
- But **inverse propensity weighting** is a more general principle, which can be used even if you do not have blocks.
- The intuition for it comes stright from **sampling weights** — you weight up in order to recover an unbiased estiamte of the potential outcomes for all units, whether or not they are assigned to treatment.
- With sampling weights however you can include units even if their weight was 1. *Why can you not include these units when doing inverse propensity weighting?*

Illustration: Estimating treatment effects with terrible treatment assignments: Fixer

Say you made a mess and used a randomization that was correlated with some variable, X. For example:

- The randomization is done in a way that introduces a correlation between Treatment Assignment and Potential Outcomes
- Then possibly, even though there is no true causal effect we naively estimate a large one — enormous bias
- However since we know the assignment procedure we can **fully** correct for the bias
- In the next example we do this using "**inverse propensity score weighting**." This is exactly analogous to standard survey weighting — since we selected different units for treatment with different probabilities, we weight them differently to recover the average outcome among treated units (same for control).

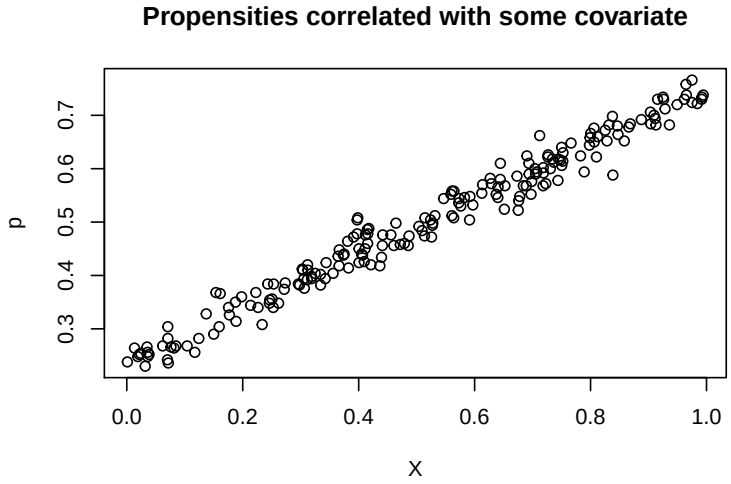
- Then you can still use information on the assignment process to recover the right estimates.
- Skip to ▸ Estimation ◂ Big Ideas

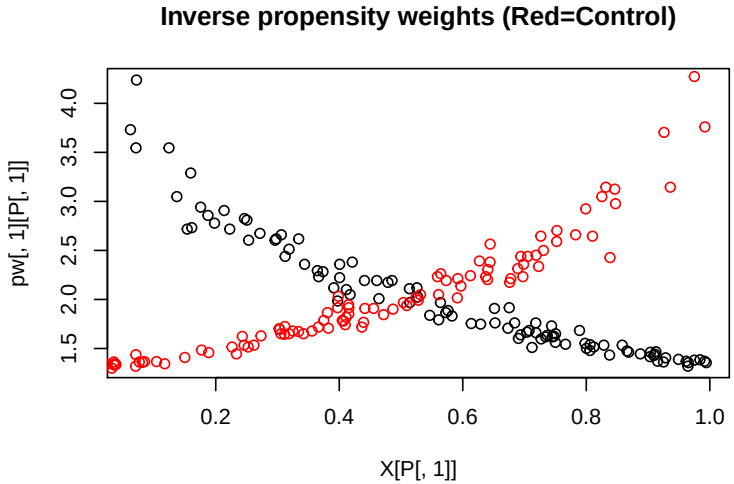
Basic randomization: Fixer

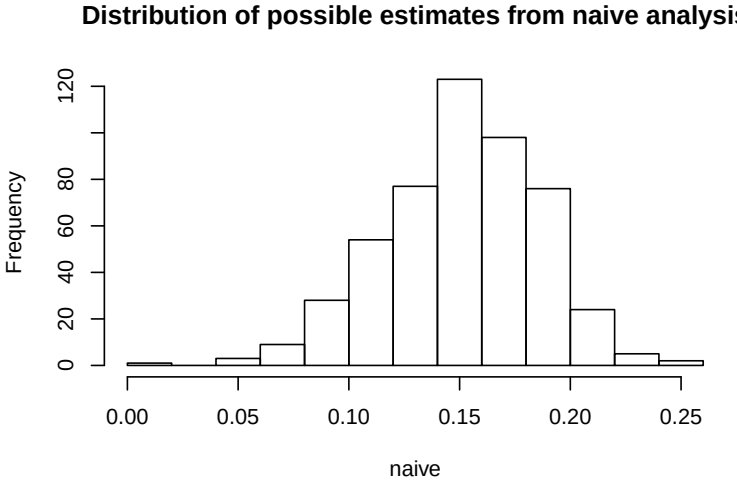
Code to generate bad assignment but proper propensity weights:

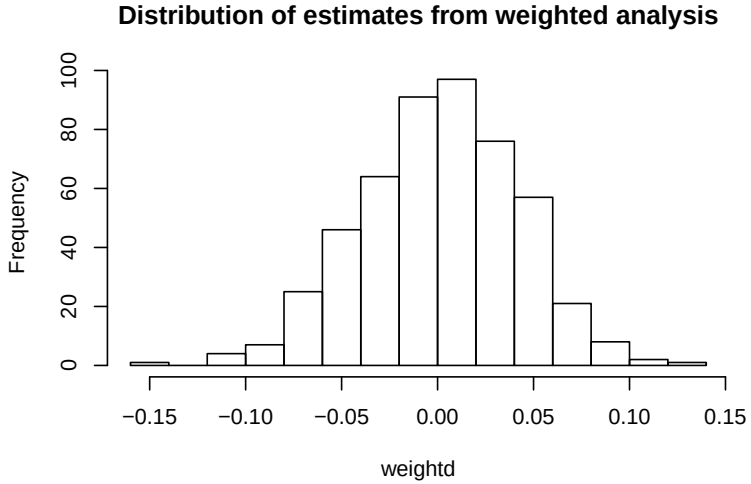
```
n  <- 200; reps <- 500; X  <- runif(n)      # Create a covariat
Y  <- Y1 <- Y0 <- X                         # Say X completely
Z  <- function(i) rank(X+2*runif(n))>(n/2) # Bad randomizatio
P  <- sapply(1:reps, Z)                     # Lots of possible
p  <- apply(P, 1, mean)                     # Recreate propens
pw <- (!P)*(1/(1-p)); pw[P]=(P*(1/p))[P]    # Create inv prop w

naive   <- sapply(1:ncol(P),function(i) {
             mean(Y[P[,i]])-mean(Y[!P[,i]])})
weightd <- sapply(1:ncol(P),function(i) {  # IPW estimates
             weighted.mean(Y[P[,i]],  pw[,i][P[,i]])-
             weighted.mean(Y[!P[,i]], pw[,i][!P[,i]])})
```

Basic randomization: Fixer









Design Based Estimation of Variance

Var(ATE)

- Recall that the treatment effect is gotten by taking a sample of outcomes under treatment and comparing them to a sample of outcomes under control
- Say that there is no "error"
- Why would this procedure produce uncertainty?

Var(ATE)

- Why would this procedure produce uncertainty?
- The uncertainty comes from being uncertain about the average outcome under control from observations of the control units, and from being uncertain about the average outcome under treatment from obervation of the treated units
- In other words, it comes from the variance in the treatment outcomes and variance in the control outcomes (and not, for example, from variance in the treatment effect)

Var(ATE)

You can also estimate variance straight from the data. From Freedman Prop 1 (using combinatorics!) we have:

$$V(\widehat{ATE}) = \frac{1}{n-1}\left[\frac{n_C}{n_T}V(Y(1)) + \frac{n_T}{n_C}V(Y(0))\right] + 2C(Y(1), Y(0))$$

Usefully rewritten as:

$$V(\widehat{ATE}) = \frac{n}{n-1}\left[\frac{V(Y(1))}{n_T} + \frac{V(Y(0))}{n_C}\right] - \frac{1}{n-1}\left[V(Y(1)) + V(Y(0)) - 2C(Y(1), Y(0))\right]$$

. . . where V denotes variance and C covariance

Note:

- We can use the sample estimates $s^2(\{Y_i\}_{i \in C})$ and $s^2(\{Y_i\}_{i \in T})$ for the first part.
- But $C(Y(1), Y(0))$ cannot be estimated from data.
- The "**Neyman**" estimator ignores the second part (and so is conservative).
- Tip: for STATA users, use ", robust" (see Samii and Aronow: On equivalencies between design-based and regression-based variance estimators for randomized experiments)

ATE and Var(ATE)}

For the case with blocking, the conservative estimator is:

$$V(\widehat{ATE}) = \sum_{s \in \mathcal{S}} \left(\frac{n_s}{n} \right)^2 \left(\frac{s_{s1}^2}{n_{s1}} + \frac{s_{s0}^2}{n_{s0}} \right)$$

Skip to ▸ Covariate Adjustment or ◂ Big Ideas

Illustration of Neyman Conservative Estimator

An illustration of *how* conservative the conservative estimator of variance really is (numbers in plot are correlations between $Y(1)$ and $Y(0)$.

We confirm that:

1. the estimator is conservative
2. the estimator is more conservative for negative correlations between $Y(0)$ and $Y(1)$ — eg if those cases that do particularly badly in control are the ones that do particularly well in treatment %, and
3. with τ and $V(Y(0))$ fixed. high positive correlations are associated with highest variance.

Illustration of Neyman Conservative Estimator

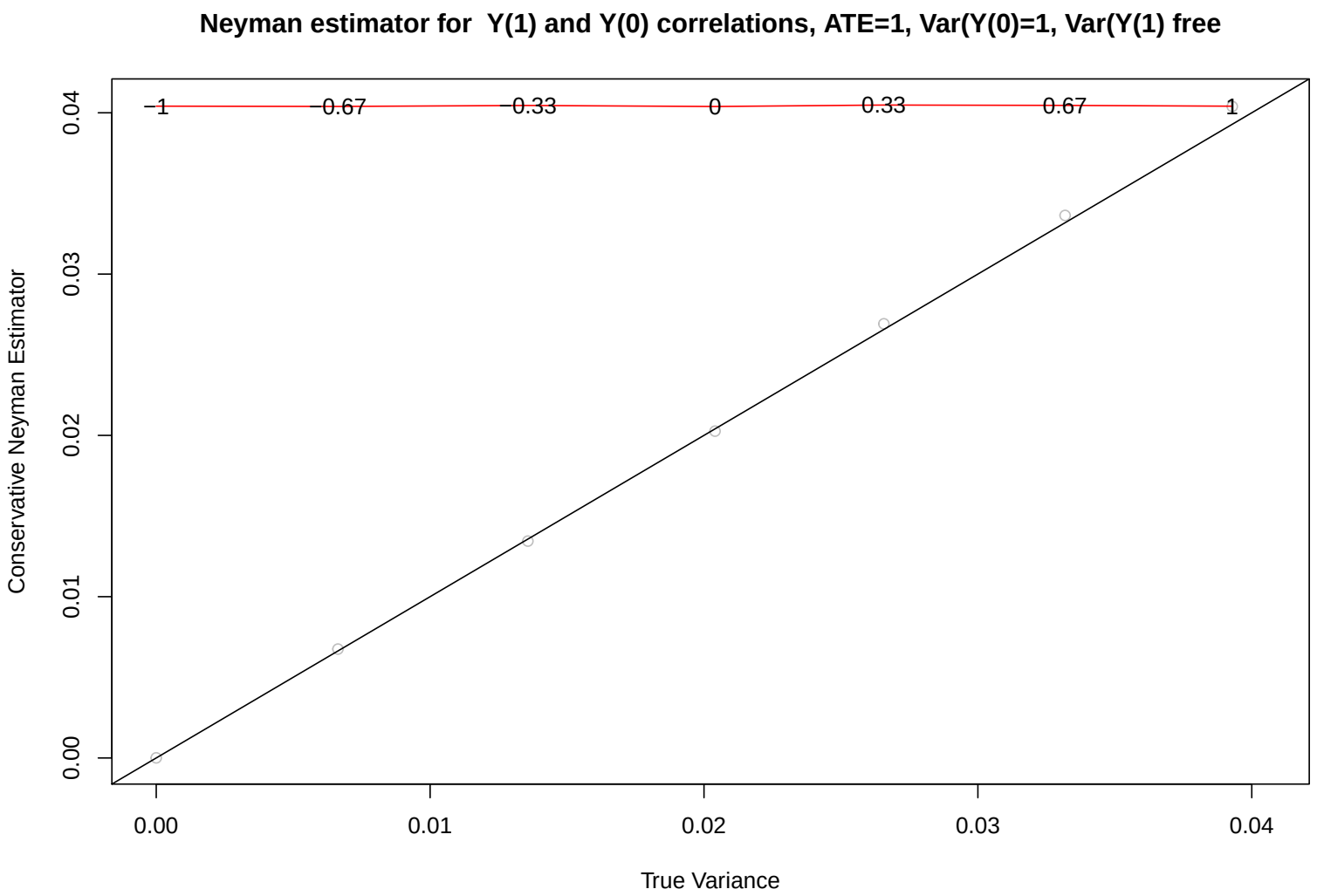


Illustration of Neyman Conservative Estimator}

τ	ρ	$\sigma^2_{Y(1)}$	Δ	σ^2_τ	$\widehat{\sigma^2_\tau}$	$\widehat{\sigma^2}_{\tau(\text{Neyman})}$
1.00	-1.00	1.00	-0.04	0.00	-0.00	0.04
1.00	-0.67	1.00	-0.03	0.01	0.01	0.04
1.00	-0.33	1.00	-0.03	0.01	0.01	0.04
1.00	0.00	1.00	-0.02	0.02	0.02	0.04
1.00	0.33	1.00	-0.01	0.03	0.03	0.04
1.00	0.67	1.00	-0.01	0.03	0.03	0.04
1.00	1.00	1.00	0.00	0.04	0.04	0.04

Here ρ is the unobserved correlation between $Y(1)$ and $Y(0)$; and Δ is the final term in the sample variance equation that we cannot estimate.

Tighter Bounds On Variance Estimate

The conservative variance comes from the fact that you do not know the covariance between $Y(1)$ and $Y(0)$.

- But as Aronow, Green, and Lee (2014) point out, you *do* know something.
- Intuitively, if you know that the variance of $Y(1)$ is 0, then the covariance also has to be zero.
- This basic insight opens a way of calculating bounds on the variance of the sample average treatment effect.

Tighter Bounds On Variance Estimate

Example:

- Take a million-observation dataset, with treatment randomly assigned
- Assume $Y(0) = 0$ for everyone and $Y(1)$ distributed normally with mean 0 and standard deviation of 1000.
- Note here the covariance of $Y(1)$ and $Y(0)$ is 0.
- Note the true variance of the estimated sample average treatment effect should be (approx) $\frac{Var(Y(1))}{\sqrt{1000000}} + \frac{Var(Y(0))}{\sqrt{1000000}} = 1$.
- But using the Neyman estimator (or OLS!) we estimate (approx) $\frac{Var(Y(1))}{\sqrt{1000000/2}} + \frac{Var(Y(0))}{\sqrt{1000000/2}} = \sqrt{2}$.
- But we can recover the truth knowing the covariance between $Y(1)$ and (0) is 0.

Tighter Bounds On Variance Estimate: Code

```
sharp_var <- function(yt,yc,N=length(c(yt,yc)),upper=TRUE){
 m <- length(yt);  n <- m + length(yc)
 V <- function(x,N) {
       (N-1)/(N*(length(x)-1)) * sum((x - mean(x))^2)}
 yt <- sort(yt)
 if(upper) {yc <- sort(yc)
       } else {yc <- sort(yc,decreasing=TRUE)}
 p_i <- unique(sort(c(seq(0,n-m,1)/(n-m),seq(0,m,1)/m)))-
        .Machine$double.eps^.5
 p_i[1] <- .Machine$double.eps^.5
 yti <- yt[ceiling(p_i*m)]; yci <- yc[ceiling(p_i*(n-m))]
 p_i_minus <- c(NA,p_i[1: (length(p_i)-1)])
 return(((N-m)/m * V(yt,N) + (N-(n-m))/(n-m)*V(yc,N) +
       2*sum(((p_i-p_i_minus)*yti*yci)[2:length(p_i)]) -
       2*mean(yt)*mean(yc))/(N-1))}
```

Illustration

```
n   <- 1000000
Y   <- c(rep(0,n/2), 1000*rnorm(n/2))
X   <- c(rep(0,n/2), rep(1, n/2))
ols <- round(coef(summary(lm(Y~X)))[2,],3)
kable(t(as.matrix(ols)))
```

Estimate	Std. Error	t value	Pr(>|t|)
1.683	1.415	1.19	0.234

```
sharp <- round(c(sharp_var(Y[X==1], Y[X==0], upper = FALSE),
                 sharp_var(Y[X==1], Y[X==0], upper = TRUE)),3)
sharp
```

```
[1] 1.001 1.001
```

Principle: Keep the reporting close to the design

Design based analysis

- Report the analysis that is implied by the design.

		T2 N	T2 Y	T2 All	Diff
T1	N	$\overline{y}_{00}$ (sd)	$\overline{y}_{01}$ (sd)	$\overline{y}_{0x}$ (sd)	$d_2 \mid T1 = 0$ (sd)
	Y	$\overline{y}_{10}$ (sd)	$\overline{y}_{10}$ (sd)	$\overline{y}_{1x}$ (sd)	$d_2 \mid T1 = 1$ (sd)
	All	$\overline{y}_{x0}$ (sd)	$\overline{y}_{x1}$ (sd)	y (sd)	d_2 (sd)
	Diff	$d_1 \mid T2 = 0$ (sd)	$d_1 \mid T2 = 1$ (sd)	d_1 (sd)	$d_1\, d_2$ (sd)

This is instantly recognizable from the design and returns all the benefits of the factorial design including all main effects, conditional causal effects, interactions and summary outcomes. It is much clearer and more informative than a regression table.

{Covariate and Regression Adjustment}

Covariate Adjustment

- Even though randomization ensures no bias you may sometimes **want** to "**control**" for covariates in order to improve efficiency (see the discussion of blocking above).
- Or you may **have** to take account of the fact that the assignment to treatment is correlated with a covariate.

Covariate Adjustment

Consider for example this data.

- You randomly pair offerers and receivers in a dictator game (in which offerers decide how much of $1 to give to receivers)
- Your population comes from two groups (80% Baganda and 20% Banyankole) *so in randomly assigning partners you are randomly determining whether a partner is a coethnic or not*
- **You find that in non coethnic pairings 35% is offered, in coethnic pairings 48% is offered**

Should you believe it?

Covariate Adjustment

- Population: randomly matched Baganda (80% of pop) and Banyankole (20% of pop)
- You find: in non coethnic pairings 35% is offered, in coethnic pairings 48% is offered
- But a closer look at the data reveals . . .

		To: Baganda	To: Banyankole
Offers by	Baganda	64%	16%
	Banyankole	16%	4%

Table 17: Number of Games

		To: Baganda	To: Banyankole
Offers by	Baganda	50	50
	Banyankole	20	20

Table 18: Average Offers

So that's a problem

Covariate Adjustment

Control?

- With such data you might be tempted to 'control' for the covariate (here: ethnic group), using regression
- But, perhaps surprisingly, it turns out that regression with covariates does not estimate average treatment effects.
- It does estimate an average of treatment effects, but specifically a minimum variance estimator, not necessarily an estimator of your estimand

Compare:

- $\hat{\tau}_{ATE} = \sum_x \frac{w_x}{\sum_j w_j} \hat{\tau}_x$
- $\hat{\tau}_{OLS} = \sum_x \frac{w_x p_x (1-p_x)}{\sum_j w_j p_j (1-p_j)} \hat{\tau}_x$

Instead the formula above for $\hat{\tau}_{ATE}$ is all you need to estimate ATE — at least for discrete covariates.

OLS and its discontents: Illustration

Two strata with different potential outcomes and assignment probabilities.

- Stratum 1 has $Y(0) = Y(1) = 0$ and 1 in 4 are treated
- Stratum 2 has $Y(1) = 14, Y(0) = 0$ and 2 in 4 assigned to treatment

```
X   <- 1*((1:8)>4) - .5
Z   <- 1:8%in%c(1,7,8)
Y0  <- rnorm(8, sd = .01)
Y1  <- 14*(X==.5)
Y   <- Z*Y1 + (1-Z)*Y0
tau <- Y1-Y0
p   <- ave(Z, group_by = X)
ipw <- 1/(p*Z + (1-p)*(1-Z))
D   <- data.frame(X, Y0, Y1, Y, tau, p, ipw)
```

OLS and its discontents: Illustration

Two strata with different potential outcomes and assignment probabilities.

- Stratum 1 has $Y(0) = Y(1) = 0$ and 1 in 4 are treated
- Stratum 2 has $Y(1) = 14, Y(0) = 0$ and 2 in 4 assigned to treatment

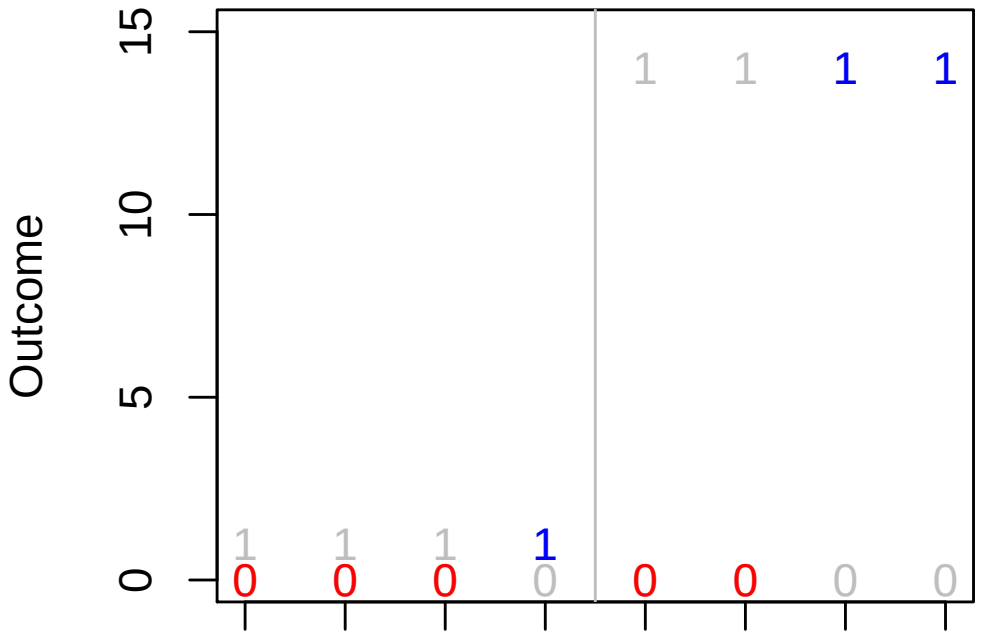
```
kable(round(D,3))
```

X	Y0	Y1	Y	tau	p	ipw
-0.5	0.018	0	0.000	-0.018	0.25	4.000
-0.5	0.003	0	0.003	-0.003	0.25	1.333
-0.5	0.000	0	0.000	0.000	0.25	1.333
-0.5	-0.008	0	-0.008	0.008	0.25	1.333
0.5	-0.021	14	-0.021	14.021	0.50	2.000
0.5	0.010	14	0.010	13.990	0.50	2.000
0.5	-0.002	14	14.000	14.002	0.50	2.000
0.5	0.011	14	14.000	13.989	0.50	2.000

OLS and its discontents

Two strata with different potential outcomes and assignment probabilities.

- Stratum 1 has $Y(0) = Y(1) = 0$ and 1 in 4 are treated
- Stratum 2 has $Y(1) = 14, Y(0) = 0$ and 2 in 4 assigned to treatment.



OLS and its discontents: Illustration

Two strata with different potential outcomes and assignment probabilities.

```
ols1 <- coef(summary(lm(Y~Z, data = D)))[2,]
ols2 <- coef(summary(lm(Y~Z+X, data = D)))[2,]
ipw1 <- coef(summary(lm(Y~Z, weight = ipw, data = D)))[2,]
ipw2 <- coef(summary(lm(Y~Z+X, weight = ipw, data = D)))[2,]
sat  <- coef(summary(lm(Y~X*Z, data = D)))[3,]
```

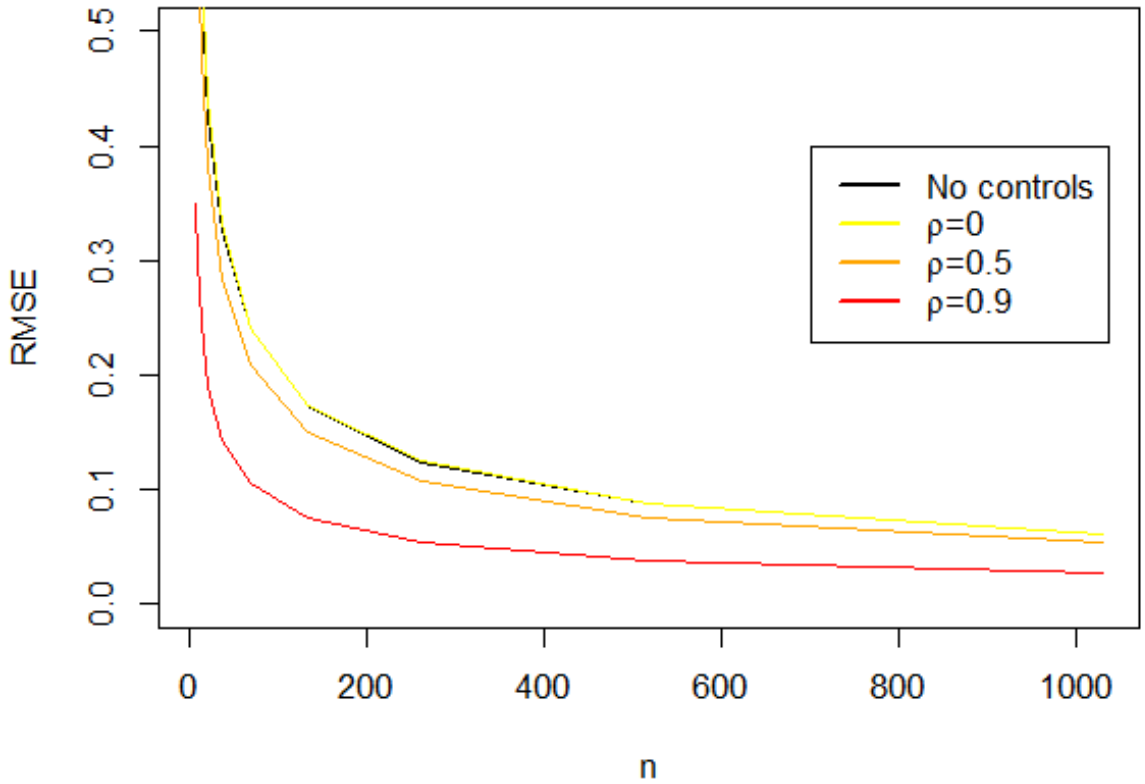
	OLS	LSDV	ipw1	ipw2	Saturated
Estimate	9.34	8.00	7.00	7.00	7.00
Std. Error	3.41	3.10	4.04	3.13	0.01
t value	2.74	2.58	1.73	2.24	794.45
Pr(>|t|)	0.03	0.05	0.13	0.08	0.00

OLS and its discontents: Illustration

- Two strata with different potential outcomes and assignment probabilities.
- So, OK on estimates but what about those varying p values in that last row? Skip to ▸ ri
- **Idea**: OLS can give the wrong answer if there is heterogeneity. But you do not need to use it.

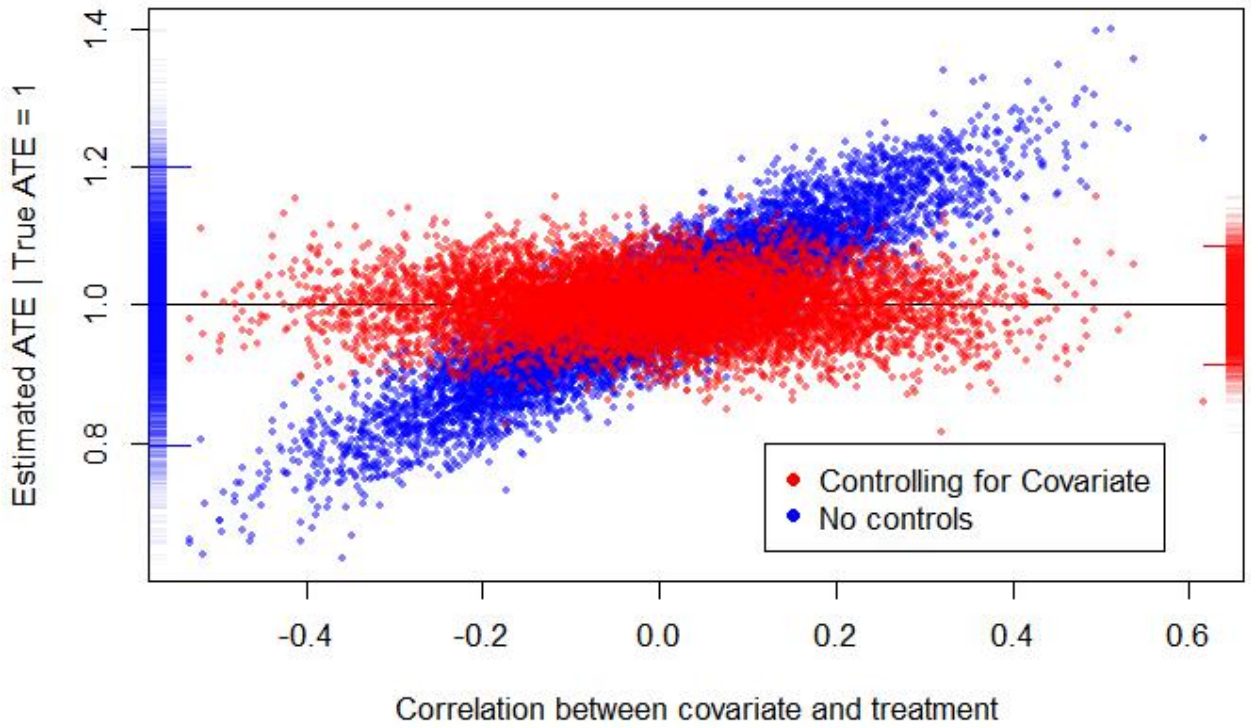
Conditional Bias and Precision Gains from Controls

What controls can do however is reduce noise and improve precision. This is an argument for using variables that are correlated with the output (not with the treatment).



Conditional Bias and Precision Gains from Controls

However, including controls when treatment is correlated with covariates can reduce "conditional bias." Doing this will change your estimates so be sure not to fish!



{Randomization Inference}

Randomization Inference

- Introducing an entirely new way to think about statistical significance. . .
- Say you randomized assignment to treatment and your data looked like this.

Unit	1	2	3	4	5	6	7	8	9	10
Treatment	0	0	0	0	0	0	0	1	0	0
Healthy?	3	2	4	6	7	2	4	9	8	2

- Does the treatment improve your health?
- $p =?$

Randomization Inference

- Introducing an entirely new way to think about statistical significance. . .
- Say you randomized assignment to treatment and your data looked like this.

Unit	1	2	3	4	5	6	7	8	9	10
Treatment	0	0	0	0	0	0	0	1	0	0
Healthy?	3	2	4	6	7	2	4	8	9	2

- Does the treatment improve your health?
- $p =?$

Randomization Inference: Some code

- In principle it is very easy.
- These few lines generate data, produce the regression estimate and then an RI estimate of p:

```
X <- rep(c(FALSE,TRUE),50)
Y <- .5*X + rnorm(100)                    # DATA

b = matrix(NA,10000)                      # RI
for(i in 1:length(b)){
     Z    <- sample(X)
     b[i] <- mean(Y[Z])- mean(Y[!Z])
     }
mean(b>=mean(Y[X])- mean(Y[!X]))    # One sided p value
```

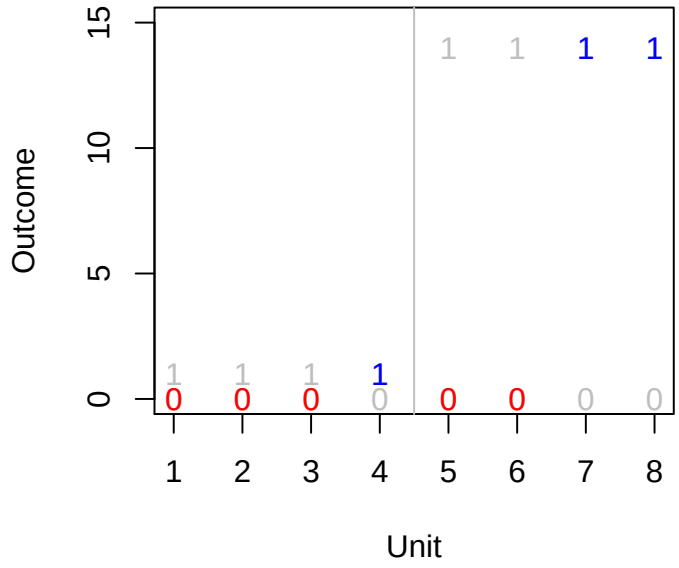
```
[1] 0.0056
```

Randomization Inference

In practice it is a good idea to create a P matrix when you do your randomization (although note: if the null is about one treatment, then you are interested only in the randomization of that treatment, not the joint randomization of all)

RI (Problem revisited)

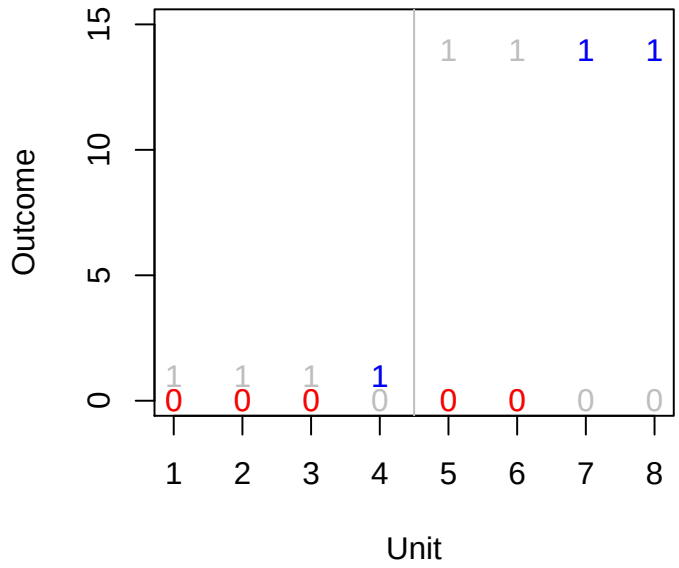
Return to this problem.



Given the strata, what are the chances that you would estimate such a big effect if in fact there was no effect for any unit?

RI (Problem revisited)

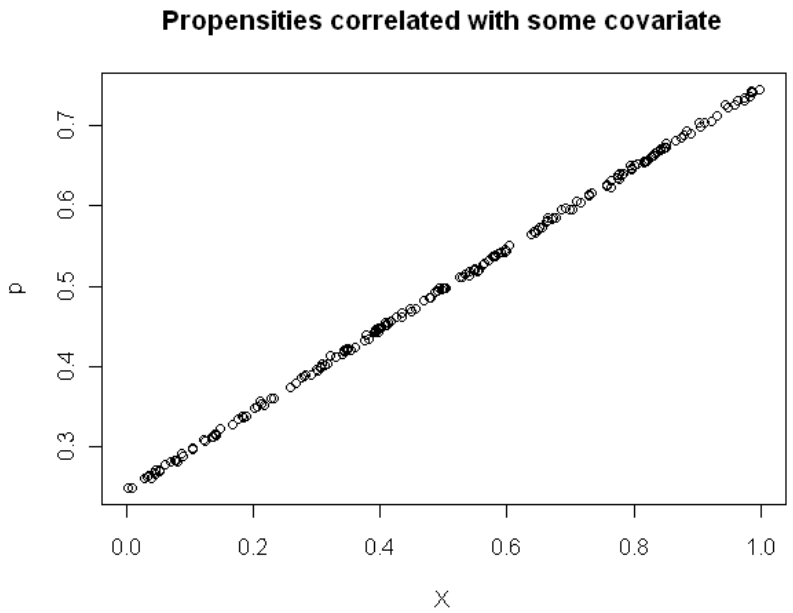
Return to this problem.

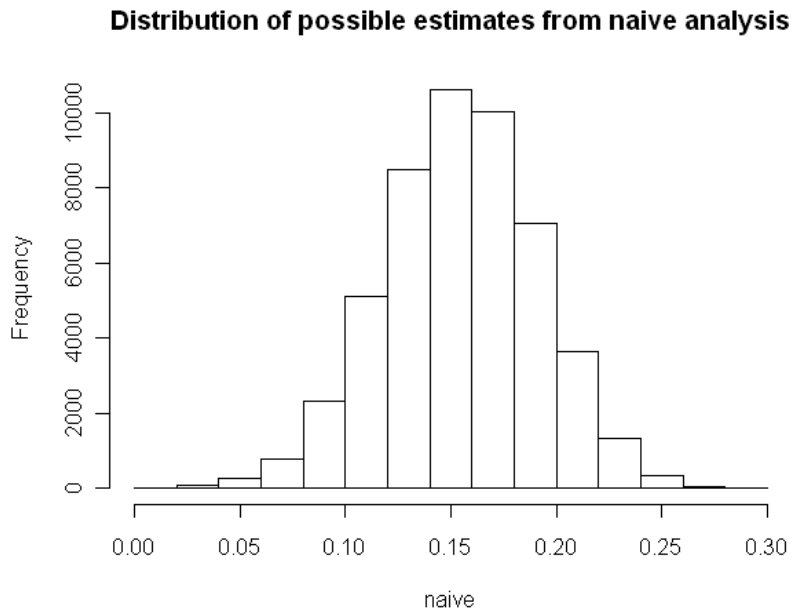


It is the probability that the two units that have Y(1)=14 both get assigned to treatment = (1/2)*(1/3)=1/6.

Randomization Inference

- Say you had a silly randomization procedure and forgot to take account of it in your estimates.





- You estimate .15. *Does the treatment improve your health?*
- $p =?$

Randomization Inference

- Randomization procedures are sometimes funky in lab experiments
- Using randomization inference would force a focus on the true assignment of individuals to treatments
- Fake (but believable) example follows

Randomization Inference

Table 21: Optimal assignment to treatment given constraints due to facilities

		Capacity	T1	T2	T3
Session	Thursday	40	10	30	0
	Friday	40	10	0	30
	Saturday	10	10	0	0
		90	30	30	30

Table 22: Constraints due to subjects

Subject Type	N	Available
A	30	Thurs, Fri
B	30	Thurs, Sat
C	30	Fri, Sat

Randomization Inference

If you think hard about assignment you might come up with an allocation like this.

Table 23: Assignment of people to days

			Allocation		
Subject Type	N	Available	Thurs	Fri	Sat
A	30	Thurs, Fri	15	15	
B	30	Thurs, Sat	25		5
C	30	Fri, Sat		25	5

That allocation balances as much as possible. Given the allocation you might randomly assign individuals to different days as well as randomly assigning them to treatments within days. If you then figure out assignment propensities, this is what you would get:

			Assignment Probabilities		
Subject Type	N	Available	T1	T2	T3
A	30	Thurs, Fri	0.25	0.375	0.375
B	30	Thurs, Sat	0.375	0.625	0
C	30	Fri, Sat	0.375		0.625

Randomization Inference

Even under the assumption that the day of measurement does not matter, these assignment probabilities have big implications for analysis.

			Assignment Probabilities		
Subject Type	N	Available	T1	T2	T3
A	30	Thurs, Fri	0.25	0.375	0.375
B	30	Thurs, Sat	0.375	0.625	0
C	30	Fri, Sat	0.375		0.625

- Only the type A subjects could have received any of the three treatments.
- There are no two treatments for which it is possible to compare outcomes for subpopulations B and C
- A comparison of $T1$ versus $T2$ can only be made for population $A \cup B$
- However subpopulation A is assigned to A (versus B) with probability $4/5$; while population B is assigned with probability $3/8$

- **Implications for design**: need to uncluster treatment delivery
- **Implications for analysis**: need to take account of propensities

Randomization Inference

- Randomization inference can get quite a bit more complicated when you want to test a null other than the sharp null of no effect.
- Say you wanted to test the null that the effect is 2 for all units. How do you do it?
- Say you wanted to test the null that an *interaction effect* is zero. How do you do it?
- In both cases by filling in a potential outcomes schedule given the hypothesis in question and then generating a test statistic

Observed		Under null that effect is 0		Under null that effect is 2	
Y(0)	Y(1)	Y(0)	Y(1)	Y(0)	Y(1)
1	?	1	1	1	3
2	?	2	2	2	4
?	4	4	4	2	4
?	3	3	3	1	3

Lecture 5: Complications for Design and Inference

Noncompliance and the LATE estimand

LATE—Local Average Treatment Effects

Sometimes you give a medicine but only a non random sample of people actually try to use it. Can you still estimate the medicine's effect?

	$X = 0$	$X = 1$
$T = 0$	$\overline{y}_{00}$ (n_{00})	$\overline{y}_{01}$ (n_{01})
$T = 1$	$\overline{y}_{10}$ (n_{10})	$\overline{y}_{11}$ (n_{11})

Say that people are one of 3 types:

- n_a "always takers" have $X = 1$ no matter what and have average outcome $\overline{y}_a$
- n_n never takers have $X = 0$ no matter what with outcome $\overline{y}_n$
- n_c compliers have $X = T$ and average outcomes $\overline{y}_c^1$ if treated and $\overline{y}_c^0$ if not.

LATE—Local Average Treatment Effects

Sometimes you give a medicine but only a non random sample of people actually try to use it. Can you still estimate the medicine's effect?

	$X=0$	$X=1$
$T=0$	$\overline{y}_{00}$ (n_{00})	$\overline{y}_{01}$ (n_{01})
$T=1$	$\overline{y}_{10}$ (n_{10})	$\overline{y}_{11}$ (n_{11})

We can figure something about types:

	$X=0$	$X=1$
$T=0$	$\frac{\frac{1}{2}n_c}{\frac{1}{2}n_c+\frac{1}{2}n_n}\overline{y}_c^0 + \frac{\frac{1}{2}n_n}{\frac{1}{2}n_c+\frac{1}{2}n_n}\overline{y}_n$	$\overline{y}_a$
$T=1$	$\overline{y}_n$	$\frac{\frac{1}{2}n_c}{\frac{1}{2}n_c+\frac{1}{2}n_a}\overline{y}_c^1 + \frac{\frac{1}{2}n_a}{\frac{1}{2}n_c+\frac{1}{2}n_a}\overline{y}_a$

LATE—Local Average Treatment Effects

You give a medicine to 50% but only a non random sample of people actually try to use it. Can you still estimate the medicine's effect?

	$X=0$	$X=1$
$T=0$	$\frac{n_c}{n_c+n_n}\overline{y}_c^0 + \frac{n_n}{n_c+n_n}\overline{y}_n$	$\overline{y}_a$
(n)	$(\frac{1}{2}(n_c+n_n))$	$(\frac{1}{2}n_a)$
$T=1$	$\overline{y}_n$	$\frac{n_c}{n_c+n_a}\overline{y}_c^1 + \frac{n_a}{n_c+n_a}\overline{y}_a$
(n)	$(\frac{1}{2}n_n)$	$(\frac{1}{2}(n_a+n_c))$

Average in $T=0$ group: $\frac{n_c\overline{y}_c^0+(n_n\overline{y}_n+n_a\overline{y}_a)}{n_a+n_c+n_n}$

Average in $T=1$ group: $\frac{n_c\overline{y}_c^1+(n_n\overline{y}_n+n_a\overline{y}_a)}{n_a+n_c+n_n}$

Difference: $ITT = (\overline{y}_c^1 - \overline{y}_c^0)\frac{n_c}{n}$

So: $LATE = ITT \times \frac{n}{n_c}$

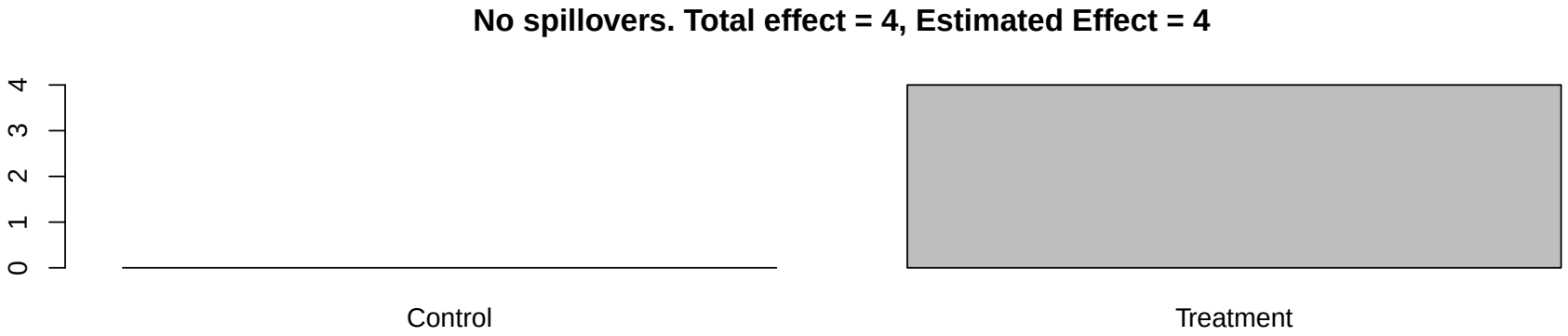
The good and the bad of LATE

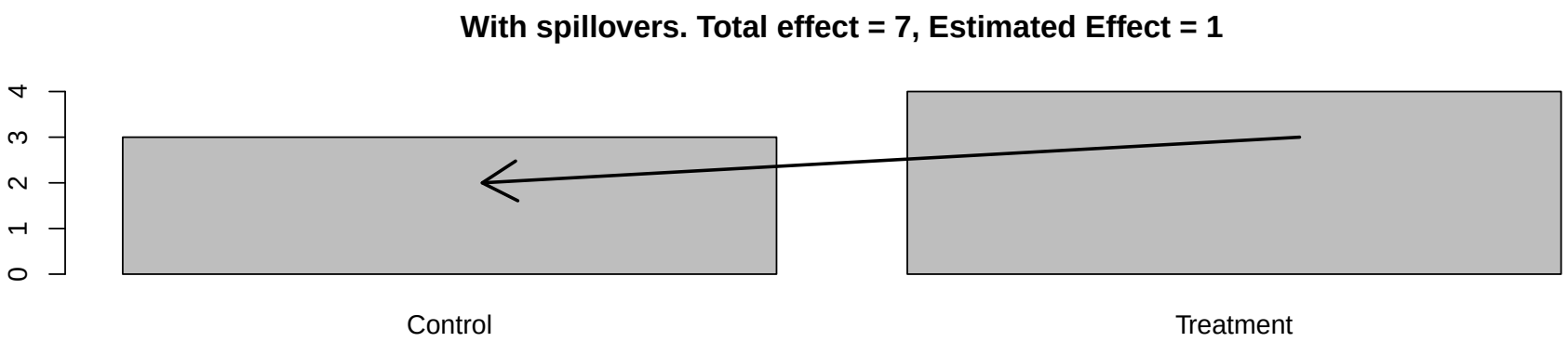
- You get a well defined estimate even when there is non random take up
- May sometimes be used to assess mediation or knock-on effects
- But:
  - You need assumptions (monotonicity and the exclusion restriction – *where were these used above*?)
  - Your estimate is only for a subpopulation
  - the subpopulation is not chosen by you and is unknown
  - Different encouragements may yield different estimates since they may encourage different subgroups

Spillovers

SUTVA violations (Spillovers)

Spillovers can result in the estimation of weaker effects when effects are actually stronger.





The key problem is that $Y(1)$ and $Y(0)$ are not sufficient to describe potential outcomes

SUTVA violations

More completely specified potential outcomes (and estimands)

		0		1		2		3		4	
Unit	Location	$D_\emptyset$	$y(D_\emptyset)$	D_1	$y(D_1)$	D_2	$y(D_2)$	D_3	$y(D_3)$	D_4	$y(D_4)$
A	1	0	0	1	3	0	1	0	0	0	0
B	2	0	0	0	3	1	3	0	3	0	0
C	3	0	0	0	0	0	3	1	3	0	3
D	4	0	0	0	0	0	0	0	1	1	3
$\bar{y}_{\text{treated}}$			-		3		3		3		3
$\bar{y}_{\text{untreated}}$			0		1		4/3		4/3		1
$\bar{y}_{\text{neighbors}}$			-		3		2		2		3
$\bar{y}_{\text{pure control}}$			0		0		0		0		0
ATT (direct effect)			-		3		3		3		3
ATT (indirect effect)			-		3		2		2		3

Table 24: Potential outcomes for four units for different treatment profiles, D_1-D_4. D_i represents an allocation to treatment and $y_j(D_i)$ is the potential outcome for (row) unit j given (column) allocation i.

SUTVA violations

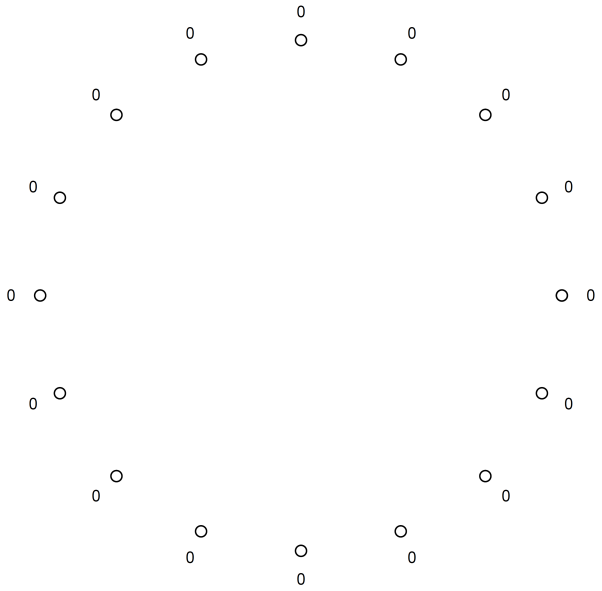
		0		1		2		3		4	
Unit	Location	$D_\emptyset$	$y(D_\emptyset)$	D_1	$y(D_1)$	D_2	$y(D_2)$	D_3	$y(D_3)$	D_4	$y(D_4)$
A	1	0	0	1	3	0	1	0	0	0	0
B	2	0	0	0	3	1	3	0	3	0	0
C	3	0	0	0	0	0	3	1	3	0	3
D	4	0	0	0	0	0	0	0	1	1	3

Table 25: Potential outcomes for four units for different treatment profiles, D_1-D_4. D_i represents an allocation to treatment and $y_j(D_i)$ is the potential outcome for (row) unit j given (column) allocation i.

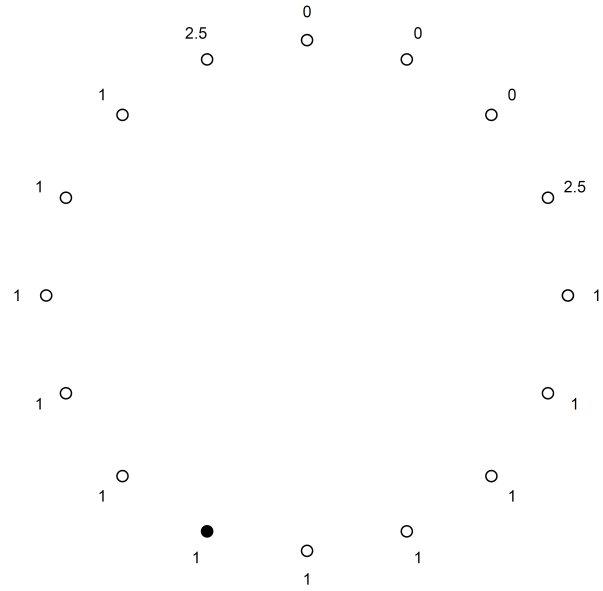
- The key is to think through the structure of spillovers.
- Here immediate neighbors are exposed
- In this case we can **define a direct treatment** (being exposed) and **an indirect treatment** (having a neighbor exposed) and we can work out *the **propensity** for each unit of receiving each type of treatment*
- These may be non uniform (here central types are more likely to have teated neighbors); but we can still use the randomization to assess effects

SUTVA violations}

Even still, to estimate effects you need some SUTVA like assumption.



Outcomes given no unit treated



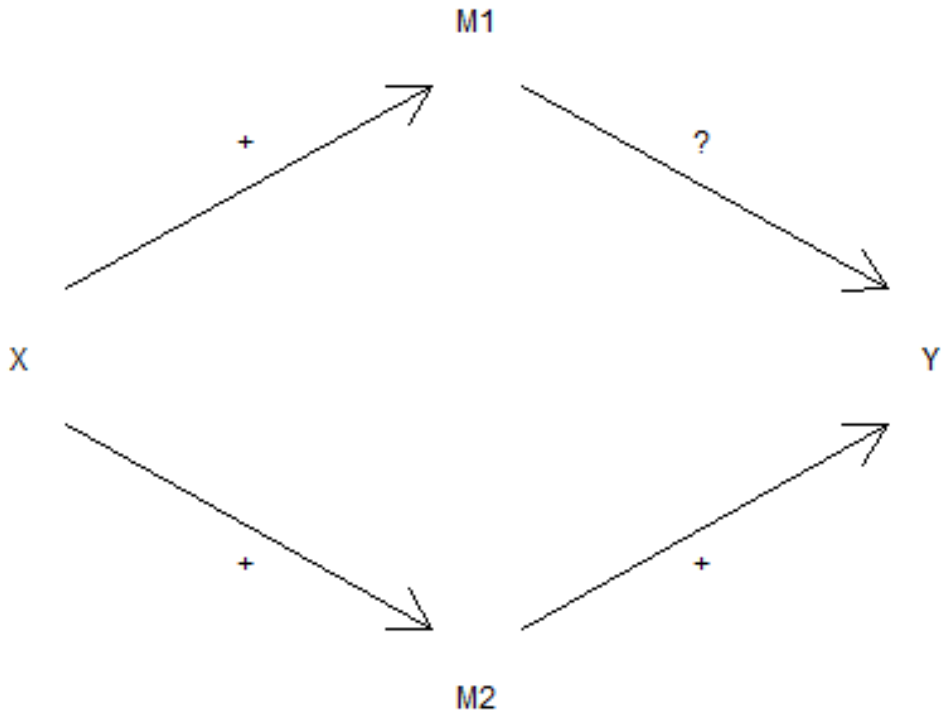
Outcomes given black unit treated

But NB: Estimates of treatment effects are sensitive to assumptions of spillover structures. In this example if one compared the outcome between treated units and all control units that are at least n positions away from a treated unit you will get the wrong answer unless $n \geq 7$.

Mediation

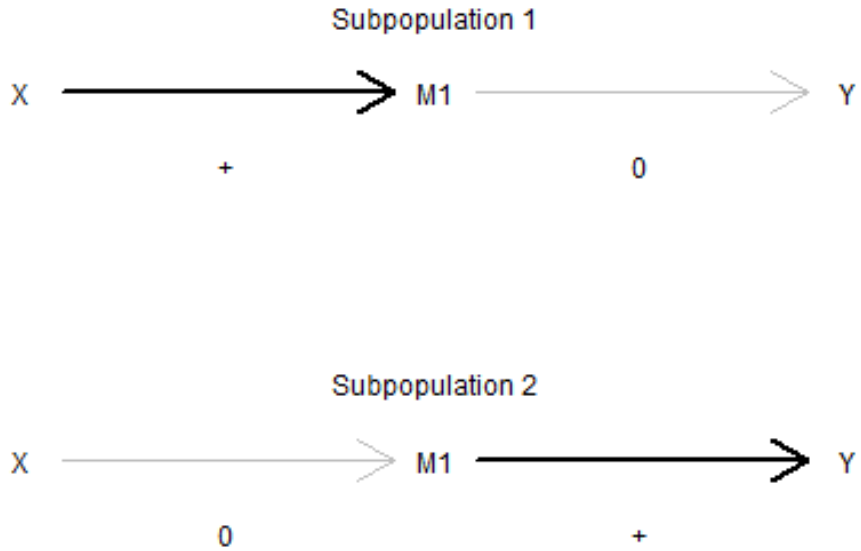
The problem of unidentified mediators

- Consider a causal system like the below.
- The effect of X on M1 and M2 can be measured in the usual way.
- But unfortunately if there are multiple mediators the effect of M1 (or M2) on Y is not identified.
- The 'excluson restriction' is obviously violated when there are multiple mediators (unless you can account for them all).



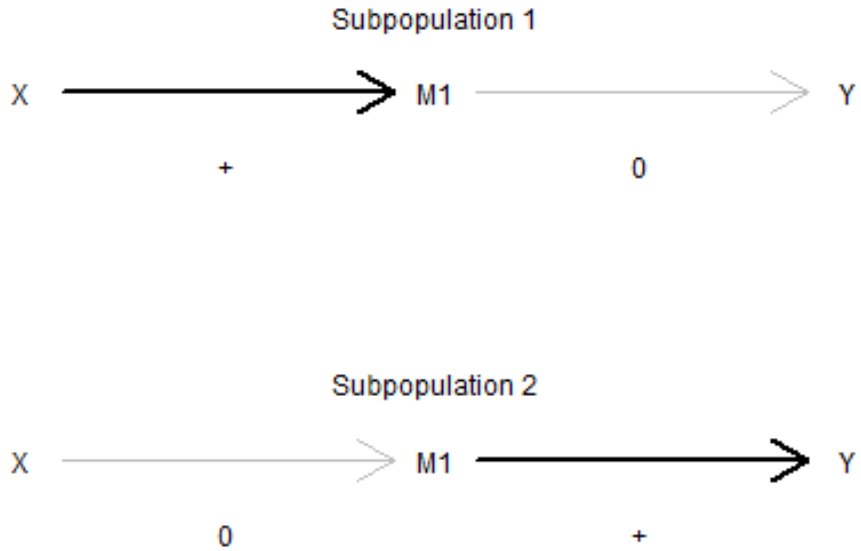
The problem of unidentified mediators}

- An obvious approach is to first examine the (average) effect of X on M1 and then use another manipulation to examine the (average) effect of M1 on Y.
- But **both of these average effects may be positive (for example) even if there is no effect of X on Y through M1**.



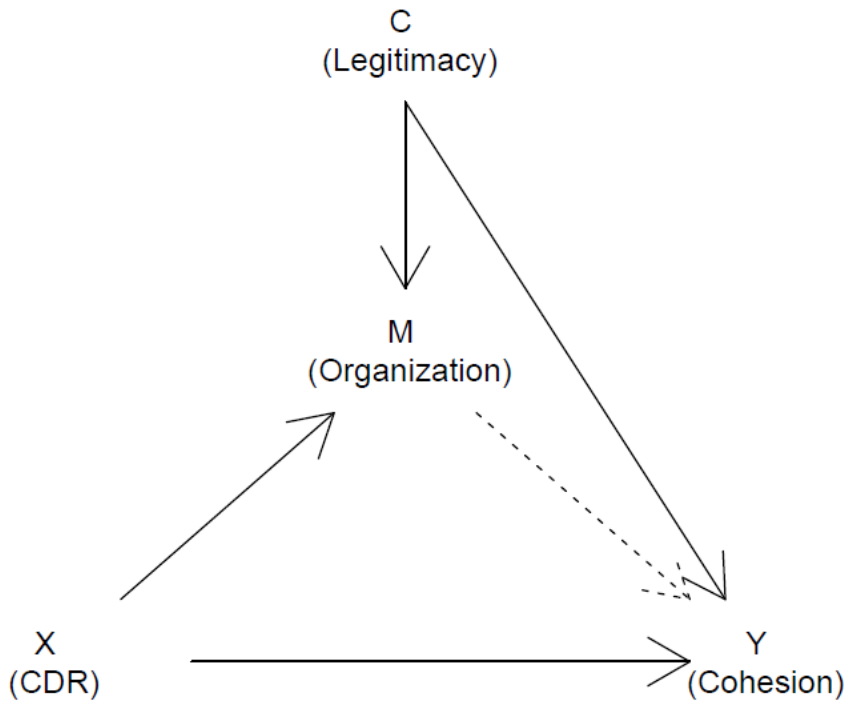
The problem of unidentified mediators}

- An obvious approach is to first examine the (average) effect of X on M1 and then use another manipulation to examine the (average) effect of M1 on Y.
- Similarly **both of these average effects may be zero even if X affects on Y through M1 for every unit!**.



The problem of unidentified mediators}

- Another somewhat obvious approach is see how the effect of X on Y in a regression is reduced when you control for M. If the effect of X on Y passes through M then surely there should be no effect of X on Y after you control for M.
- But this common strategy is also not guaranteed to produce reliable results
- See Imai on better ways to think about this problem and designs to address it



The problem of unidentified mediators: Quantities

- In the potential outcomes framework we can describe a **mediation effect** as (see Imai et al):

$$\delta_i(t) = Y_i(t, M_i(1)) - Y_i(t, M_i(0)) \text{ for } t = 0, 1$$

- The **direct effect** is:

$$\psi_i(t) = Y_i(1, M_i(t)) - Y_i(0, M_i(t)) \text{ for } t = 0, 1$$

- This is a **decomposition**, since:

$$Y_i(1, M_i(1)) - Y_1(0, M_i(0)) = \frac{1}{2}(\delta_i(1) + \delta_i(0) + \psi_i(1) + \psi_i(0))$$

- If (and a big if), there are no interaction effects—ie $\delta_i(1) = \delta_i(0), \psi_i(1) = \psi_i(0)$, then

$$Y_i(1, M_i(1)) - Y_1(0, M_i(0)) = \delta_i + \psi_i$$

- The bad news is that although a single experiment might identify the total effect, it can not identify these elements of the direct effect.

The problem of unidentified mediators: Solutions?

- Check **formal requirement** for identification under single experiment design ("sequential ignorability"—that, conditional on actual treatment, it is as if the value of the mediation variable is randomly assigned relative to potential outcomes). But this is strong (and in fact unverifiable) and if it does not hold, bounds on effects always include zero (Imai et al)
- You can use **interactions** with covariates **if you are willing to make assumptions on no heterogeneity of direct treatment effects** over covariates. eg you think that money makes people get to work faster because they can buy better cars; you look at the marginal effect of more money on time to work for people with and without cars and find it higher for the latter. This might imply mediation through transport but only if there is no direct effect heterogeneity (eg people with cars are less motivated by money).

The problem of unidentified mediators: Solutions?

- Weaker assumptions justify '**parallel design**'
  - Group A: T is randomly assigned, M left free.
  - Group B: divided into four groups $T \times M$ (requires two more assumptions (1) that the **manipulation** of the mediator only affects outcomes through the mediator (2) **no interaction**, for each unit, $Y(1, m) - Y(0, m) = Y(1, m') - Y(0, m')$.)

Idea 5: Understanding mechanisms is harder than you think. Figure out what assumptions fly.

Skip to ▸ Spillovers or ◂ Big Ideas

Lecture 6: Prospects and Limits

Prospects

Prospects

- Whenever someone is uncertain about *something* they are doing (all the time)
- Whenever someone hits scarcity constraints
- When people have incentives to demonstrate that they are doing the right thing (careful...)
- **Advice 1**: If you can, **start from theory** and find an intervention, rather than the other way around.
- **Advice**: If you can, go for *structure* rather than *gimmicks*
- **Advice**: In attempts to parse, beware of generating unnatural interventions (how should a voter think of a politician that describes his policy towards Korea in detail but does not mention the economy? Is not mentioning the economy sending an unintended message?)

Prospects & Potential

- Randomization of where police are stationed (India)
- Randomization of how government tax collectors get paid (do they get a share?) (Pakistan)
- Randomization of the voting rules for determining how decisions get made (Afghanistan)
- Random assignment of populations to peacekeepers (Liberia)
- Random assignment of excombatants out of their networks (Indonesia)
- Randomization of students to ethnically homogeneous or ethnically diverse schools (anywhere?)

Ethics

Constraint: Is it ethical to manipulate subjects for research purposes?

- There is no foundationless answer to this question. So lets take some foundations from the Belmont report and seek to ensure:
  1. Respect for persons
  2. Beneficience
  3. Justice
- Unfortunately operationalizing these requires further ethical theories. Lets assume that (1) is operationalized by informed consent (a very liberal idea). We are a bit at sea for (2) and (3) (the Belmont report suggests something like a utilitarian solution).
- The major focus on (1) by IRBs might follow from the view that if subjects consent, then they endorse the ethical calculations made for 2 and 3 —*they* think that it is good and fair.
- This is a little tricky though since the study may not be good or fair because of implications for non-subjects.

Is it ethical to manipulate subjects for research purposes?

- The problem is that many (many) field experiments have nothing like informed consent.
- eg Whether the government builds a school in your village, whether an ad appears on your favorite radio show, and so on.
- Consider three cases:
  1. You work with a nonprofit to post (true?) posters about the crimes of politicians on billboards to see effects on **voters**
  2. You hire confederates to offer bribes to **police officers** to see if they are more likely to bend the law for coethnics
  3. The British government asks you to work on figuring out how the use of water cannons helps stop **rioters** rioting

Is it ethical to manipulate subjects for research purposes?}

- Consider three cases:
  - You work with a nonprofit to post (true?) posters about the crimes of politicians on billboards to see effects on **voters**
  - You hire confederates to offer bribes to **police officers** to see if they are more likely to bend the law for coethnics
  - The British government asks you to work on figuring out how the use of water cannons helps stop **rioters** rioting
- In all cases there is **no consent** given by subjects
- In 2 and 3 the treatment is **possibly harmful** for subjects and the results might also be harmful. But even in case 1 there could be major unintended harmful consequences.
- In cases 1 and 3 however the "intervention" is within the sphere of **normal activities** for the implementer.

Constraint: Is it ethical to manipulate subjects for research purposes?

- Sometimes possible to use this point of difference to make a "spheres of ethics" argument for "embedded experimentation"
- Spheres of Ethics Argument: experimental research that involves manipulations that are not normally appropriate for researchers may nevertheless be ethical if:
  - Researchers and implementers agree on a **division of responsibility** where implementers take on responsibility for actions
  - Implementers have **legitimacy** to make these decisions within the sphere of the intervention
  - Implementers are indeed **materially independent** of researchers (no swapping hats)
- Difficulty with this arguments:
  - Question begging: How to determine legitimacy of implementer? (Can we rule out Nazi doctors?)

Otherwise keep focus on consent and desist if this is not possible

Transparency & Experimentation

Contentious Issues

Experimental researchers are deeply engaged in the movement towards more transparency social science research.

Contentious issues (mostly):

- **Analytic replication.** This should be a no brainer. Set everything up so that replication is easy. Use rmarkdown, or knitr or sweave. Or produce your replication code as a package.

Contentious Issues

Experimental researchers are deeply engaged in the movement towards more transparency social science research.

Contentious issues (mostly):

- **Data.** How soon should you make your data available? **My view**: as soon as possibe. Along with working papers and before publication. Before it affects policy in any case. Own the ideas not the data.
- Hard core: no citation without (analytic) replication. Perhaps. Non-replicable results should not be influencing policy.
- **Where should you make your data available?** Dataverse is focal for political science. Not personal website (mea culpa)
- **What data should you make available?** Disagreement is over how raw your data should be. **My view**: as raw as you can but at least post cleaning and pre-manipulation.

Contentious Issues

Experimental researchers are deeply engaged in the movement towards more transparency social science research.

Contentious issues (mostly):

- **Should you register?**: Hard to find reasons against. But case strongest in testing phase rather than exploratory phase.
- **Registration**: When should you register? **My view:** Before treatment assignment. (Not just before analysis, mea culpa)
- **Registration**: Should you deviate from an preanalysis plan if you change your mind about optimal estimation strategies. **My view:** Yes, but make the case and describe both sets of results.

Contentious Issues

Experimental researchers are deeply engaged in the movement towards more transparency social science research.

Contentious issues (mostly):

- **Registration**: When should you register? **My view:** Before treatment assignment. (Not just before analysis, mea culpa)
- **Registration**: Should you deviate from an preanalysis plan if you change your mind about optimal estimation strategies. **My view:** Yes, but make the case and describe both sets of results.

Hard Constraints

There remain hard constraints:

- Real Time
- History has Happened
- Scale
- Variables as Attributes
- The assignment process matters

These may make experimentation hard or impossible. That's OK. Let the question determine the method.

Backmatter

About these slides

- **Credits**: DeclareDesign code is developed with Graeme Blair, Jasper Cooper, and Alex Coppock; DeclareDesign applications draw on material developed by Anne Wilke and Tara Slough. Much thanks to them.
- These slides are written in Rmarkdown. Rmarkdown is a simple markup language that lets you integrate R with LaTeX. The great advantage is that all code written in these slides is run on compilation and so all output is generated live. Not all the code is displayed, but you can always inspect the Rmd file to see the backend of these slides.