Methods Supplementary Lecture 3: Experimentation

Department of Government
London School of Economics and Political Science

Physical causality

- Action and reaction
- Features: Observable and deterministic
- Example:
  - Picture a ball resting on top of a hill
  - What happens if I push the ball?
- Physical causality is easy to see

Correlation I

- Correlation is the non-independence of two variables for a set of observations

Correlation II

- *Observation*: A case or unit (e.g., person, country)
- *Variable*: A dimension that describes an obseration (e.g., income)
- *Independence*: Variables are unrelated to one another
  - Independent: Height and value on a fair dice roll
  - Non-independent: Height and weight

Correlation III

- Synonyms: correlation, covariation, relationship, association
  - "Effect" is frequently used to mean correlation
  - We'll reserve that term for a *causal effect*
- Any correlation is a potential cause
  1. X might cause Y
  2. Y might cause X
  3. X and Y might be caused by Z
  4. X and Y might cause Z
  5. There may be no causal relationship

Mill's methods[1]

- Agreement
- Difference
- Agreement and Difference
- Residue
- Concomitant variations

[1] Discussed in Holland (1982)

Difference

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or an necessary part of the cause, of the phenomenon.

Four (or five) principles of causality

1. Correlation
2. Nonconfounding
3. Direction ("temporal precedence")
4. Mechanism
5. (Appropriate level of analysis)

Counterfactual Thinking

- *Counterfactual*: relating to what has not happened or is not the case
- Causal inference involves inferring *what would have happened* in a counterfactual reality *where the potential cause took on a different value*

"A Christmas Carol"

- 1843 novel by Charles Dickens
- Ebenezer Scrooge is shown his own future by the "Ghost of Christmas Yet to Come"
- Has the choice to either:
  - stay on current path (one counterfactual), or
  - change his ways (take a different counterfactual)

Causation

- *Causal effect*: The difference between two "potential outcomes"
  - The outcome that occurs if $X = x_1$
  - The outcome that occurs if $X = x_2$
- The causal effect of Scrooge's lifestyle is seen in the differences between two potential futures

Fundamental problem of causal inference

We can only observe any given unit in one reality!

Two solutions![2]

1. Scientific Solution
   - All units are identical
   - Each can provide a perfect counterfactual
   - Common in, e.g., agriculture, biology

[2] From Holland

Two solutions![2]

1. Scientific Solution
   - All units are identical
   - Each can provide a perfect counterfactual
   - Common in, e.g., agriculture, biology
2. Statistical Solution
   - Units are not identical
   - Random exposure to a potential cause
   - Effects measured on average across units
   - Known as the "Experimental ideal"

[2] From Holland

In Political Science

- Causal inference is about searching for appropriate counterfactuals

In Political Science

- Causal inference is about searching for appropriate counterfactuals
  - *Causal effect*: Difference in an outcome variable between two counterfactuals

In Political Science

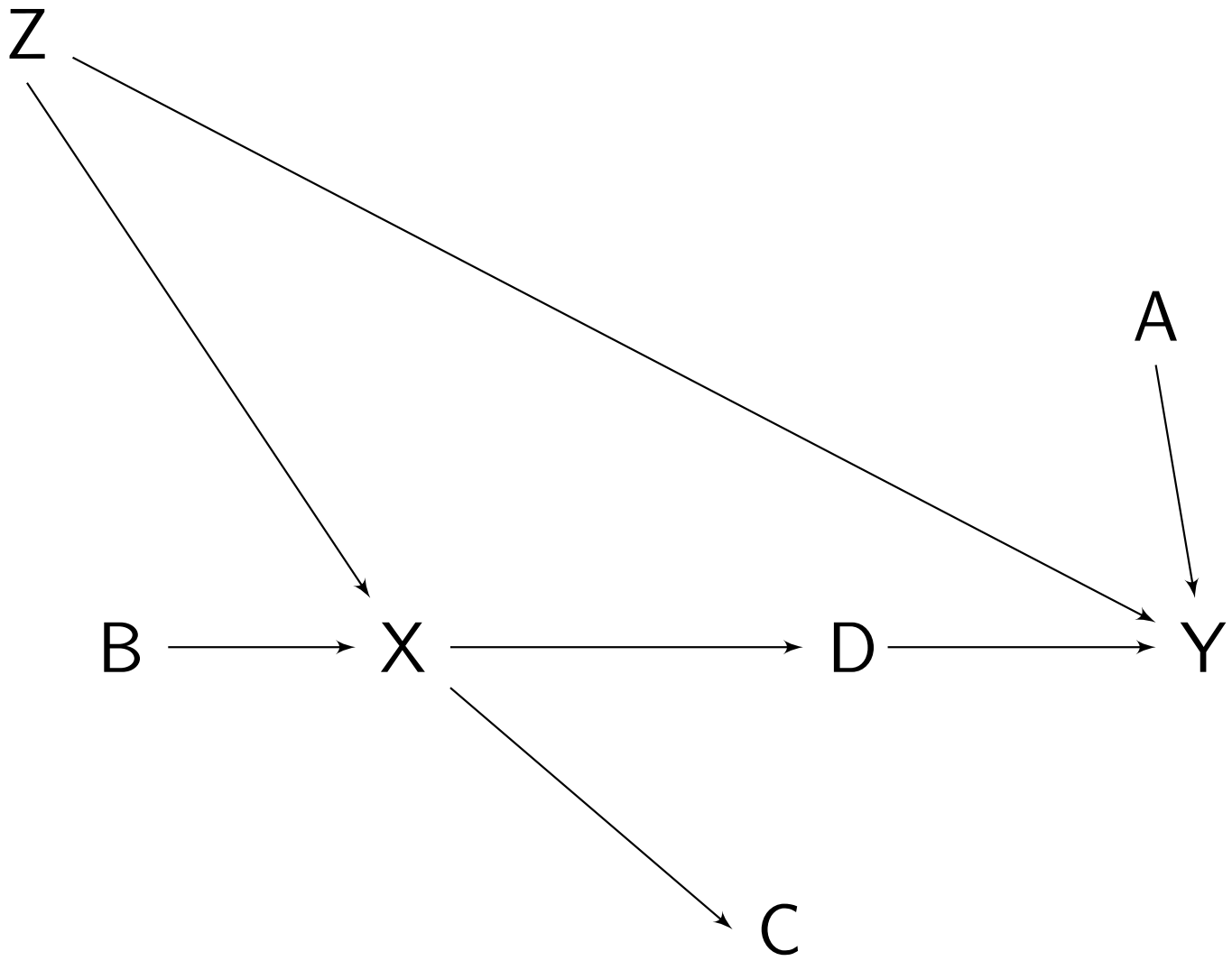
- Causal inference is about searching for appropriate counterfactuals
  - *Causal effect*: Difference in an outcome variable between two counterfactuals
  - *Causal inference*: A belief that an event or variable exerts a causal effect on an outcome

In Political Science

- Causal inference is about searching for appropriate counterfactuals
  - *Causal effect*: Difference in an outcome variable between two counterfactuals
  - *Causal inference*: A belief that an event or variable exerts a causal effect on an outcome
- Where can we look for counterfactuals?

An Example

- For example, if we think smoking might cause lung cancer, how would we know?
- How would we know if smoking caused lung cancer for an individual who smoked?
  - What's the relevant counterfactual?
- How would we know if smoking causes lung cancer on average across many individuals?
  - What's the relevant counterfactual?


Z
A
B
X
D
Y
C

Observational Causal Inference

- Find all variables Z
- Control for influence of Z to identify effect of $X \rightarrow Y$

Observational Causal Inference

- Find all variables Z
- Control for influence of Z to identify effect of $X \rightarrow Y$
- One common strategy is *matched sampling* or *matching*

Observational Causal Inference

- Find all variables Z
- Control for influence of Z to identify effect of $X \rightarrow Y$
- One common strategy is *matched sampling* or *matching*
- Regression is similar idea of "conditioning on observables"

Matching I

- Example: Effect of Education on Ideology
- Our design involves:
  - Measure outcome (ideology)
  - Measure putative cause (education; university degree)
  - Correlate outcome and cause
- Is that correlation a valid causal inference?

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant
- What about confounding?

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant

- What about confounding?
  1. Hide outcome data

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant
- What about confounding?
  1. Hide outcome data
  2. List all potential confounds (Z)

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant

- What about confounding?
  1. Hide outcome data
  2. List all potential confounds (Z)
  3. Match observations so sample consists of pairs of observations (1 with degree; 1 without) that are identical (or at least similar) on all variables

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant

- What about confounding?
  1. Hide outcome data
  2. List all potential confounds (Z)
  3. Match observations so sample consists of pairs of observations (1 with degree; 1 without) that are identical (or at least similar) on all variables
  4. Discard all observations that cannot be matched

Matching II

- Temporal ordering is correct here
  - Note: Timing of measurement may be unimportant

- What about confounding?
  1. Hide outcome data
  2. List all potential confounds (Z)
  3. Match observations so sample consists of pairs of observations (1 with degree; 1 without) that are identical (or at least similar) on all variables
  4. Discard all observations that cannot be matched
  5. Estimate $Corr(X, Y)$

Think To Yourself

- Does matching always get us to a clear and uncontroversial causal inference?

The Experimental Ideal

- Randomized experiment, or randomized control trial
  - *The observation of units after, and possibly before, a randomly assigned intervention in a controlled setting, which tests one or more precise causal expectations*
- A correctly executed experiment always provides clear causal inference
- It solves both the temporal ordering and confounding problems
  - Treatment (X) is applied by the researcher before outcome (Y)
  - Randomization means there are no confounding (Z) variables

Experiments

- American Political Science Association president A. Lawrence Lowell (1909): *"We are limited by the impossibility of experiment. Politics is an observational, not an experimental science..."*
- First political science experiment: Gosnell (1926)
- Experiments prominent in psychology and the physical sciences
- King, Keohane, and Verba (1994) only mentions experiments once

Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*

Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*
- A potential outcome is the value of the outcome (Y) for a given unit (i) after receiving a particular version/level/amount of the treatment (X)

Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*
- A potential outcome is the value of the outcome (Y) for a given unit (i) after receiving a particular version/level/amount of the treatment (X)
- Each unit has multiple *potential* outcomes, but we only observe one of them

Causal Inference in Experiments I

- Causal inference is a comparison of two *potential outcomes*
- A potential outcome is the value of the outcome (Y) for a given unit (i) after receiving a particular version/level/amount of the treatment (X)
- Each unit has multiple *potential* outcomes, but we only observe one of them
- A *causal effect* is the difference between two potential outcomes (e.g., $Y_{X=1} - Y_{X=0}$), all else constant

Causal Inference in Experiments II

- We cannot see individual-level causal effects

Causal Inference in Experiments II

- We cannot see individual-level causal effects
- We can see *average causal effects*
  - Ex.: Average difference in participation between those with and without university degrees

Causal Inference in Experiments II

- We cannot see individual-level causal effects
- We can see *average causal effects*
  - Ex.: Average difference in participation between those with and without university degrees
- We want to know: $TE_i = Y_{1i} - Y_{0i}$

Causal Inference in Experiments III

- We want to know: $TE_i = Y_{1i} - Y_{0i}$

Causal Inference in Experiments III

- We want to know: $TE_i = Y_{1i} - Y_{0i}$
- We can average:
  $ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$

Causal Inference in Experiments III

- We want to know: $TE_i = Y_{1i} - Y_{0i}$
- We can average:
  $ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$
- But we still only see one potential outcome for each unit:

  $ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0]$

Causal Inference in Experiments III

- We want to know: $TE_i = Y_{1i} - Y_{0i}$
- We can average:
  $ATE = E[Y_{1i} - Y_{0i}] = E[Y_{1i}] - E[Y_{0i}]$
- But we still only see one potential outcome for each unit:

  $ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0]$
- Is this what we want to know?

Causal Inference in Experiments IV

- What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \quad (1)$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (2)$$

Causal Inference in Experiments IV

- What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \tag{1}$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \tag{2}$$

- Are the following statements true?
  - $E[Y_{1i}] = E[Y_{1i}|X = 1]$
  - $E[Y_{0i}] = E[Y_{0i}|X = 0]$

Causal Inference in Experiments IV

- What we want and what we have:

$$ATE = E[Y_{1i}] - E[Y_{0i}] \quad (1)$$

$$ATE_{naive} = E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (2)$$

- Are the following statements true?
  - $E[Y_{1i}] = E[Y_{1i}|X = 1]$
  - $E[Y_{0i}] = E[Y_{0i}|X = 0]$
- Not in general!

Causal Inference in Experiments V

- Only true when both of the following hold:

$$E[Y_{1i}] = E[Y_{1i}|X = 1] = E[Y_{1i}|X = 0] \quad (3)$$
$$E[Y_{0i}] = E[Y_{0i}|X = 1] = E[Y_{0i}|X = 0] \quad (4)$$

- In that case, potential outcomes are *independent* of treatment assignment

- If true, then:

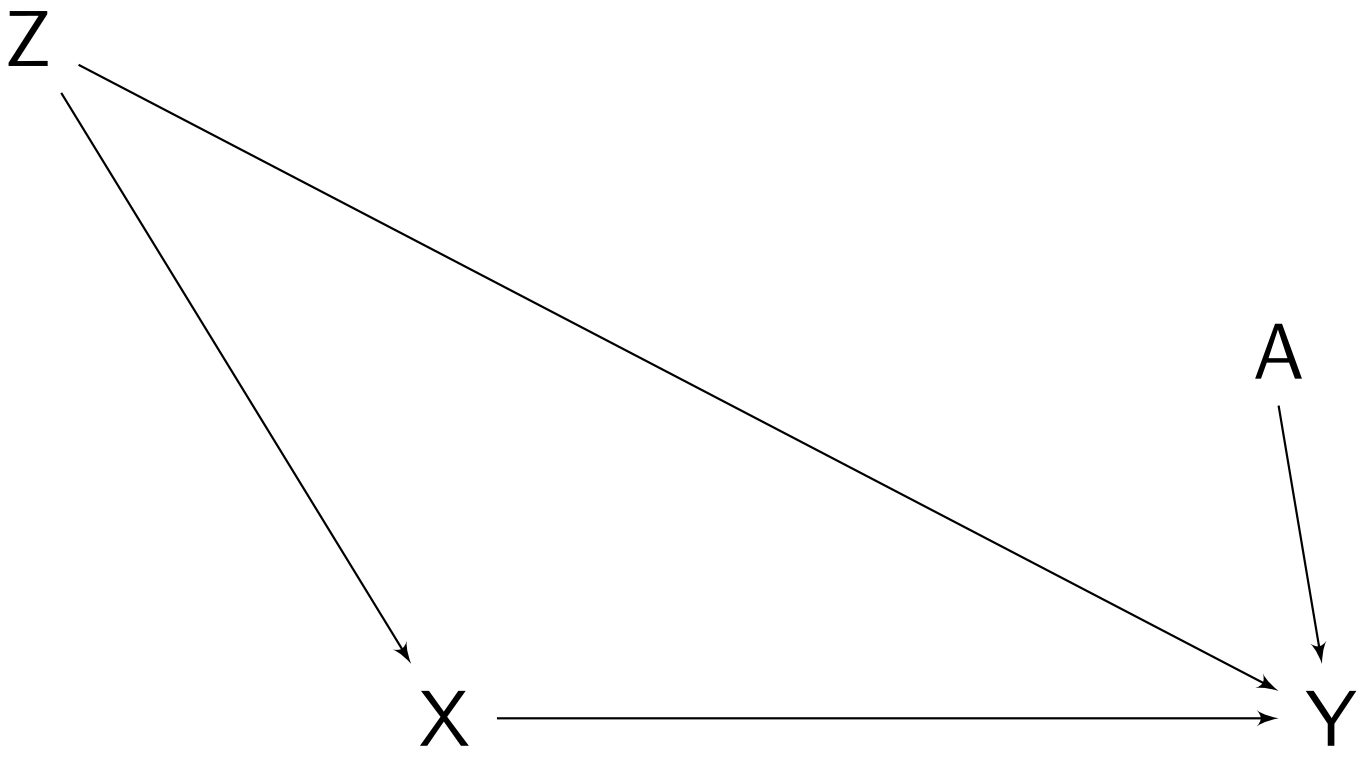
$$\begin{aligned} ATE_{naive} &= E[Y_{1i}|X = 1] - E[Y_{0i}|X = 0] \quad (5) \\ &= E[Y_{1i}] - E[Y_{0i}] \\ &= ATE \end{aligned}$$

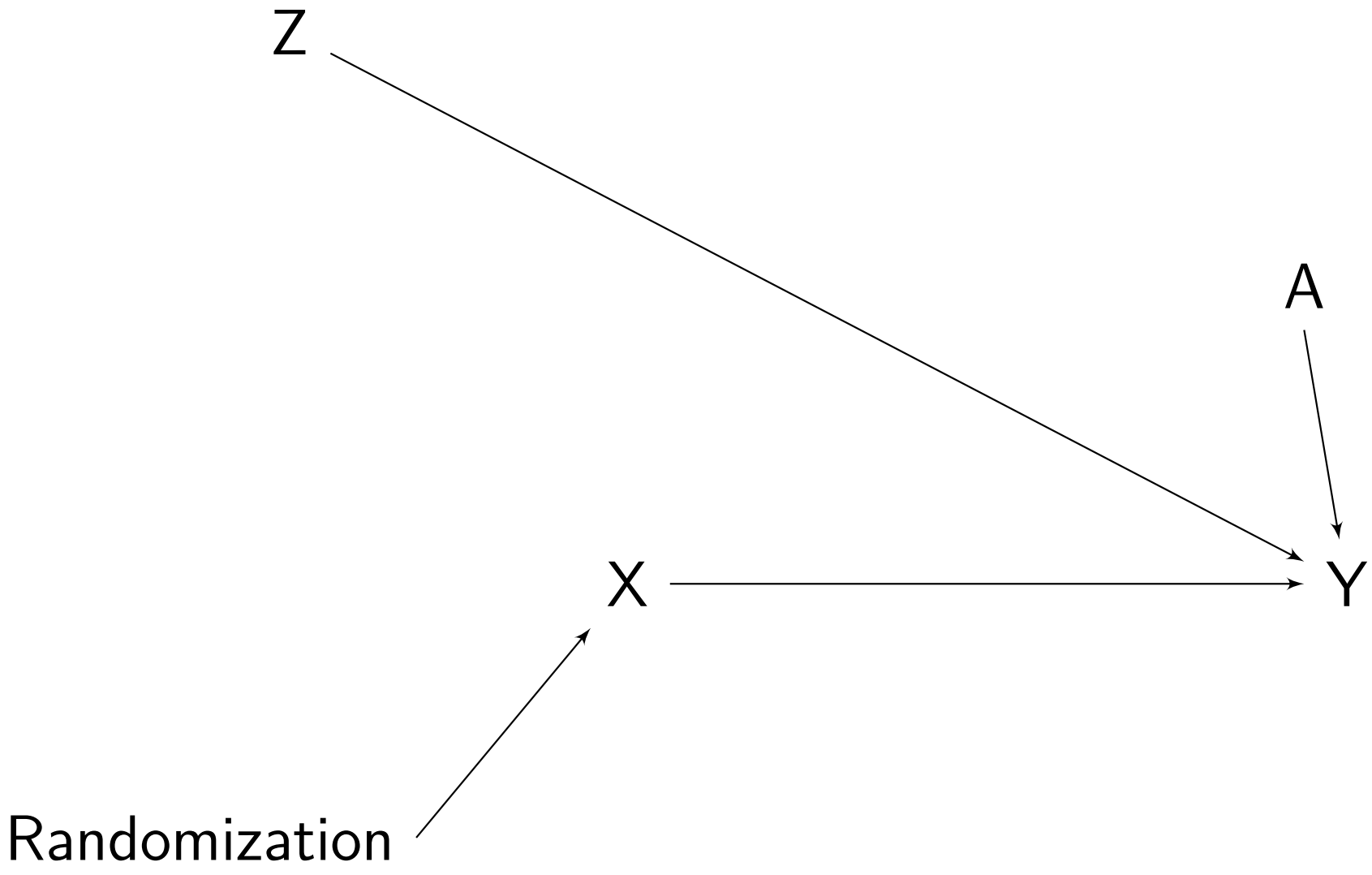
Causal Inference in Experiments VI

- This holds in experiments because of randomization
  - Units differ only in what side of coin was up
  - Experiments randomly reveal potential outcomes

Causal Inference in Experiments VI

- This holds in experiments because of randomization
  - Units differ only in what side of coin was up
  - Experiments randomly reveal potential outcomes
- Potential outcomes are not independent of treatment assignment when there is confounding


Z
A
X
Y


Z
A
X
Y
Randomization

Causal Inference in Experiments VI

- This holds in experiments because of randomization
  - Units differ only in what side of coin was up
  - Experiments randomly reveal potential outcomes
- Potential outcomes are not independent of treatment assignment when there is confounding

Causal Inference in Experiments VI

- This holds in experiments because of randomization
  - Units differ only in what side of coin was up
  - Experiments randomly reveal potential outcomes
- Potential outcomes are not independent of treatment assignment when there is confounding
- Matching and regression attempt to eliminate those confounds, such that:

$$E[Y_{1i}|Z] = E[Y_{1i}|X = 1, Z] = E[Y_{1i}|X = 0, Z]$$
$$E[Y_{0i}|Z] = E[Y_{0i}|X = 1, Z] = E[Y_{0i}|X = 0, Z]$$

Experimental Analysis I

- The statistic of interest in an experiment is the *sample average treatment effect* (SATE)
- This boils down to being a mean-difference between two groups:

$$SATE = \frac{1}{n_1} \sum Y_{1i} - \frac{1}{n_0} \sum Y_{0i} \tag{5}$$

- In practice we often estimate this using:
  - t-tests
  - Linear regression

Experimental Analysis II

- We don't just care about the size of the SATE. We also want to know whether it is significantly different from zero (i.e., different from no effect/difference)
- To know that, we need to estimate the *variance* of the SATE
- The variance is influenced by:
  - Total sample size
  - Variance of the outcome, Y
  - Relative size of each treatment group

Experimental Analysis II

- Formula for the variance of the SATE is:

$$\widehat{Var}(SATE) = \frac{\widehat{Var}(Y_0)}{N_0} + \frac{\widehat{Var}(Y_1)}{N_1}$$

  - $\widehat{Var}(Y_0)$ is control group variance
  - $\widehat{Var}(Y_1)$ is treatment group variance

- We often express this as the *standard error* of the estimate:

$$\widehat{SE}_{SATE} = \sqrt{\frac{\widehat{Var}(Y_0)}{N_0} + \frac{\widehat{Var}(Y_1)}{N_1}}$$

Statistical Significance I

- To assess statistical significance, we convert the mean-difference into a t-statistic
- This is just the ratio of the effect to the SE:

$$t = \frac{\widehat{SATE}}{\hat{SE}(SATE)}$$

- A t-statistic of 1.96 or larger is considered statistically significant

Statistical Significance I

- To assess statistical significance, we convert the mean-difference into a t-statistic
- This is just the ratio of the effect to the SE:

$$t = \frac{\widehat{SATE}}{\hat{SE}(SATE)}$$

- A t-statistic of 1.96 or larger is considered statistically significant
- But we also care about *substantive* significance

t-statistic

- A measure of how large a coefficient is relative to our uncertainty about its size
- Typically used to test a formal null hypothesis:
  - No effect null: $t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{SE_{\hat{\beta}_1}}$
  - Any other null: $\frac{\hat{\beta}_1 - \alpha}{SE_{\hat{\beta}_1}}$, where α is our null hypothesis effect size

t-statistic

- A measure of how large a coefficient is relative to our uncertainty about its size
- Typically used to test a formal null hypothesis:
  - No effect null: $t_{\hat{\beta}_1} = \frac{\hat{\beta}_1}{SE_{\hat{\beta}_1}}$
  - Any other null: $\frac{\hat{\beta}_1 - \alpha}{SE_{\hat{\beta}_1}}$, where α is our null hypothesis effect size
- Note: The t-statistic from a t-test of mean-difference is the same as the t-statistic from a t-test on an OLS slope for a dummy indicator of treatment group

Statistical Significance II

- Two equivalent ways to obtain a t-statistic:
  - A two-group t-test (in R: `t.test()`)
  - A regression of the outcome on an indicator for treatment group (in R: `lm(y ~ tr)`)

Standardized Effect Sizes

- In two-group experiments, we can use the *standardized mean difference* as an effect size
  - Expresses size of mean-difference in "number of standard deviations"
  - Typically referred to as Cohen's d or Hedge's g
- Definition:
  $d = \frac{\bar{x}_1 - \bar{x}_0}{s}$, where $s = \sqrt{\frac{(n_1-1)s_1^2 + (n_0-1)s_0^2}{n_1+n_0-2}}$
- Heuristic about size of effects:
  - Small: ~ 0.2; Medium: ~ 0.5; Large: ~ 0.8

Type I and Type II Errors

We often talk of Type I and Type II errors. This is a statement about whether our experimental results are "true."

	H_0 True	H_0 False
Reject H_0	Type I Error	*True positive*
Accept H_0	*True negative*	Type II Error

Power Analysis

- Definitions of experimental power:
  - "The probability of not making a Type II error"
  - "Probability of a true positive"
  - "The probability of rejecting the null hypothesis when a causal effect exists"

What increases power?

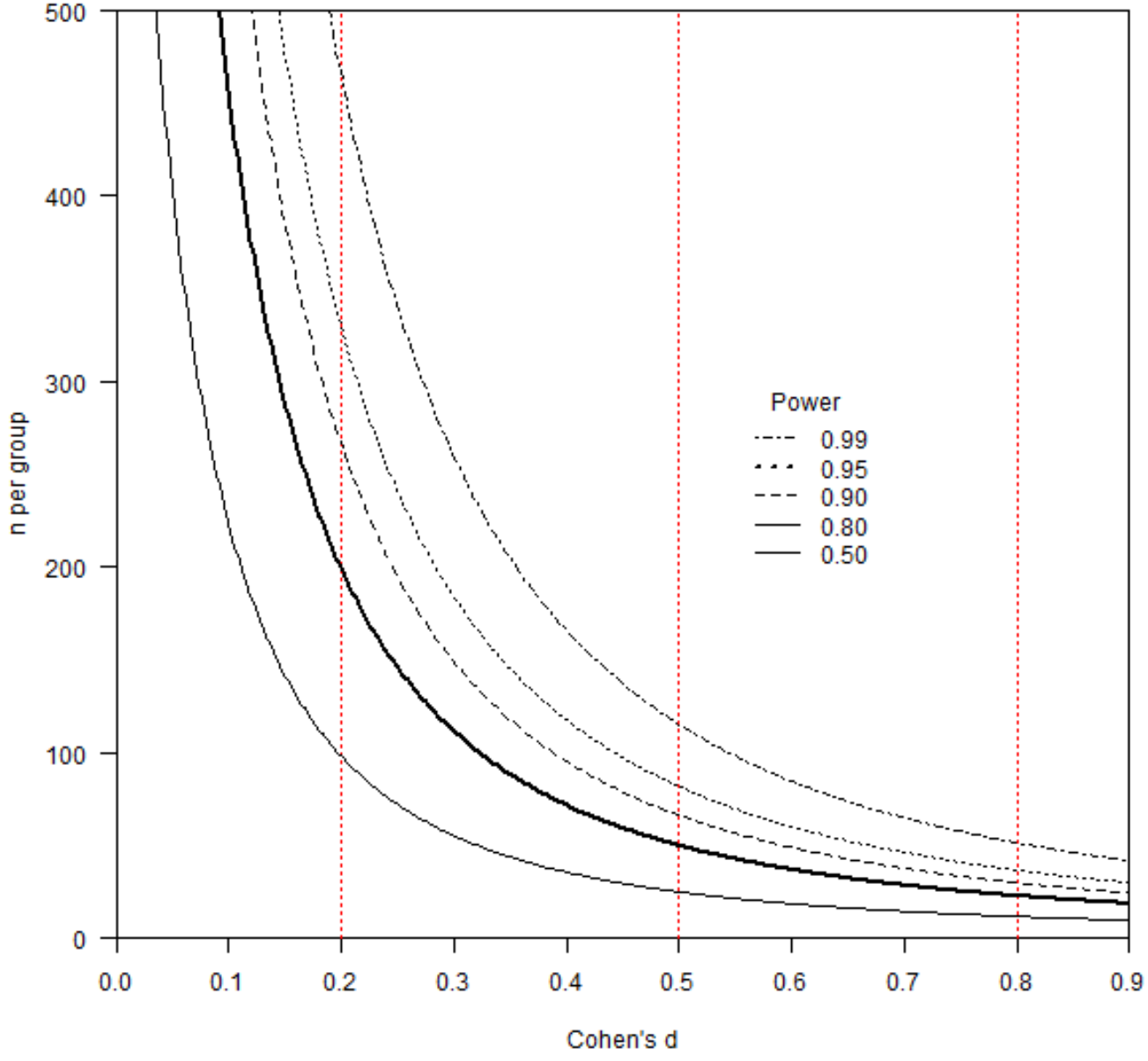
- As n increases, power increases
- As the true effect size increases, power increases (holding n constant)
- As $Var(Y)$ decreases, power increases

Doing a Power Analysis I

- Formal definition of power:
  $$\textit{Power} = \phi\left(\frac{|\mu_1 - \mu_0|\sqrt{N}}{2\sigma} - \phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$
  - μ is treatment group mean
  - N is total sample size
  - σ is outcome standard deviation
  - α is statistical significance level
  - ϕ is Normal distribution function
- Power of experiment is determined *a priori* by guessing at these values
- Conventionally, 0.80 is a reasonable power level

Doing a Power Analysis II

- A "minimum detectable effect size" flips the power formula to solve for effect size rather than power
- In R, we can determine how large of a sample we need for given effect sizes:
  `power.t.test(n = NULL, delta = x, power = 0.50)$n`,
  where we set `x` as the minimum detectable standardized mean-difference


n per group
Cohen's d
Power
0.99
0.95
0.90
0.80
0.50
0
100
200
300
400
500
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Some final considerations

1. Ethics
2. Compliance
3. Effect Moderation
4. Mediation/mechanisms

Ethics

- Experiments raise lots of ethical considerations
- Because we are intervening in peoples' lives, we have to weight harm and benefits of our interventions
- A big question relates to "deception" (are we deceiving our experimental participants? is that a problem?)

Compliance I

- Compliance is when individuals receive and accept the treatment to which they are assigned
- Non-compliance is when participants receive the wrong treatment (cross-over) or simply fail to receive the treatment to which they are assigned
- This causes problems for our analysis because factors other than randomization explain why individuals receive their treatment
- Lots of methods for dealing with this, but the consequence is generally reduced power

Asymmetric Noncompliance

- If noncompliance only occurs in one group, it is *asymmetric*
- We can ignore non-compliance and analyze the "intention to treat" effect, which will underestimate our effects because some people were not treated as assigned
  $ITT = \overline{Y}_1 - \overline{Y}_0$
- We can use "instrumental variables" to estimate the "local average treatment effect" (LATE) for those that complied with treatment:
  $LATE = \frac{ITT}{PercentCompliant}$
- We can ignore randomization and analyze data "as-treated", but this makes our study no longer an experiment

Two-Sided Noncompliance

- Two-sided noncompliance is more complex analytically
- Stronger assumptions are required to analyze it and we won't discus them here
- Best to try to develop a better design to avoid this rather than try to deal with the complexities of analyzing a broken design

Effect Moderation

- Sometimes effect might vary across individuals or contexts
- This is called *moderation* or *effect heterogeneity*
- If we suspect this happens, we should design a complex experiment in which we manipulate the moderator
- This way we can estimate "conditional average treatment effects" (CATEs) for each subgroup

Effect Mediation

- Sometimes we care about *why* an effect comes about (i.e., what is the mechanism?)
- This is called *mediation*
- If we suspect this happens and we care about the mediation process, we should try to manipulate the treatment and the suspected mediator
- If we cannot manipulate the mediator, there is basically no credible way of estimating the "mediation effect" of the treatment group a given mediator