Welcome and First Lecture

Department of Political Science and Government
Aarhus University

February 3, 2015

Scientific Method

Scientific Method

1. Research question

What makes for an interesting question?

What makes for an interesting question?

1. Politically important

What makes for an interesting question?

1. Politically important
2. Contribute to scientific understanding/literature

What makes for an interesting question?

1. Politically important
2. Contribute to scientific understanding/literature
3. Personally interesting

What makes for an interesting question?

1. Politically important
2. Contribute to scientific understanding/literature
3. Personally interesting
4. Unresolved

Three Types of Research Questions

1. Prevalence
2. Changes
3. Causal Effects

Causal Questions

- Forward causal questions
- Backward causal questions

Causal Questions

- Forward causal questions
  - What effect(s) does X have?
- Backward causal questions

Causal Questions

- Forward causal questions
  - What effect(s) does X have?
  - "What if?" questions
- Backward causal questions

Causal Questions

- Forward causal questions
  - What effect(s) does X have?
  - "What if?" questions
- Backward causal questions
  - What causes Y?

Why not backward causal questions?

Why not backward causal questions?

- The set of potential X's is infinite

Why not backward causal questions?

- The set of potential X's is infinite
- We can only test a few at a time

Why not backward causal questions?

- The set of potential X's is infinite
- We can only test a few at a time
- Some X's might be unobservable or unknown

Why not backward causal questions?

- The set of potential X's is infinite
- We can only test a few at a time
- Some X's might be unobservable or unknown
- Showing that X1 causes Y doesn't tell us anything about whether X2 causes Y

Scientific Method

1. Research question

Scientific Method

1. Research question
2. Theory development

Concepts

"The empiricist perspective seems reasonable on the face of things. And yet we are unable to talk about questions of fact without getting caught up in the language that we use to describe these facts. To be sure, things exist the world separate from the language we use to describe them. However, we cannot talk about them unless and until we introduce linguistic symbols." (Gerring 2012)

Concepts

- Term/label
- Attributes (i.e., definition)
- Indicators (i.e., operationalization)

Concepts

- Term/label
- Attributes (i.e., definition)
- Indicators (i.e., operationalization)
- *Example: Democracy*

Evaluating Concepts

1. Resonance: is it intuitive?
2. Domain: where does it apply?
3. Extension: how many referants?
4. Fecundity: is it useful?
5. Utility: can we observe it? how?

Theory Development

- Caveat: Not the focus of this course
- Theories are arguments *and* explanations
- Theories are causal in nature
- Often involve claims about mechanisms
- Rooted in past evidence or observation

Scientific Method

1. Research question
2. Theory development

Scientific Method

1. Research question
2. Theory development
3. Hypotheses

Scientific Method

1. Research question
2. Theory development
3. Hypotheses
   - Expectations about differences in outcomes across levels of a putatively causal variable

Scientific Method

1. Research question
2. Theory development
3. Hypotheses
   - Expectations about differences in outcomes across levels of a putatively causal variable
4. Design

From Hypotheses to Design

- Definitions
  - Hypothesis: Expectations about differences in outcomes across levels of a putatively causal variable
  - Design: Selection and arrangement of evidence (Gerring 2012)

From Hypotheses to Design

- Definitions
  - Hypothesis: Expectations about differences in outcomes across levels of a putatively causal variable
  - Design: Selection and arrangement of evidence (Gerring 2012)

- Design must reveal evidence about our hypotheses

From Hypotheses to Design

- Definitions
  - Hypothesis: Expectations about differences in outcomes across levels of a putatively causal variable
  - Design: Selection and arrangement of evidence (Gerring 2012)

- Design must reveal evidence about our hypotheses
  - Selection of observations/units

From Hypotheses to Design

- Definitions
  - Hypothesis: Expectations about differences in outcomes across levels of a putatively causal variable
  - Design: Selection and arrangement of evidence (Gerring 2012)

- Design must reveal evidence about our hypotheses
  - Selection of observations/units
  - Observation of outcome(s)

From Hypotheses to Design

- Definitions
  - Hypothesis: Expectations about differences in outcomes across levels of a putatively causal variable
  - Design: Selection and arrangement of evidence (Gerring 2012)
- Design must reveal evidence about our hypotheses
  - Selection of observations/units
  - Observation of outcome(s)
  - Observation of variation on the causal variable

From Hypotheses to Design

- Definitions
  - Hypothesis: Expectations about differences in outcomes across levels of a putatively causal variable
  - Design: Selection and arrangement of evidence (Gerring 2012)

- Design must reveal evidence about our hypotheses
  - Selection of observations/units
  - Observation of outcome(s)
  - Observation of variation on the causal variable
  - *Causal identification*

History of Political Science Methods

- Early methods
  - Philosophy
  - History
  - Formal legalism
- Behavioral revolution (1950s)
  - Sample surveys
  - Basic quantitative methods
- Credibility revolution (1960s, 1980s, 2000s)
  - Matching
  - Experiments
  - Quasi-experiments (see Week 11)
- Post-positivists (21st century)
  - Qualitative and interpretive methods

Causal Inference

- Full discussion of causality next week

Causal Inference

- Full discussion of causality next week
- "Correlation does not prove causation"

Causal Inference

- Full discussion of causality next week
- “Correlation does not prove causation”
- Correlation is causation when:

Causal Inference

- Full discussion of causality next week
- “Correlation does not prove causation”
- Correlation is causation when:
  - X temporally precedes Y

Causal Inference

- Full discussion of causality next week

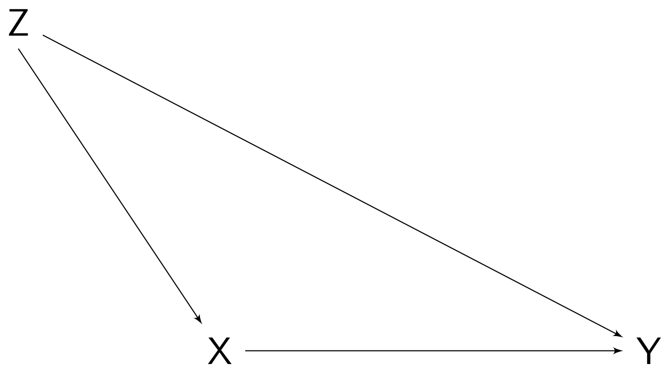
- "Correlation does not prove causation"

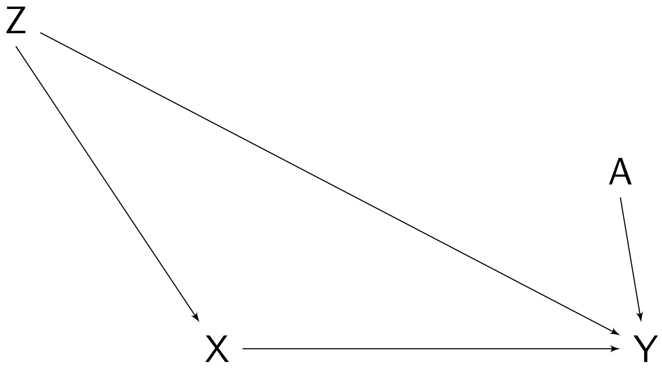
- Correlation is causation when:
  - X temporally precedes Y
  - No confounding

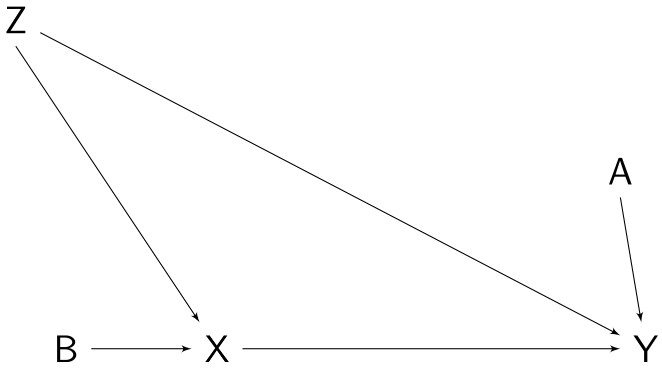
Causal Inference

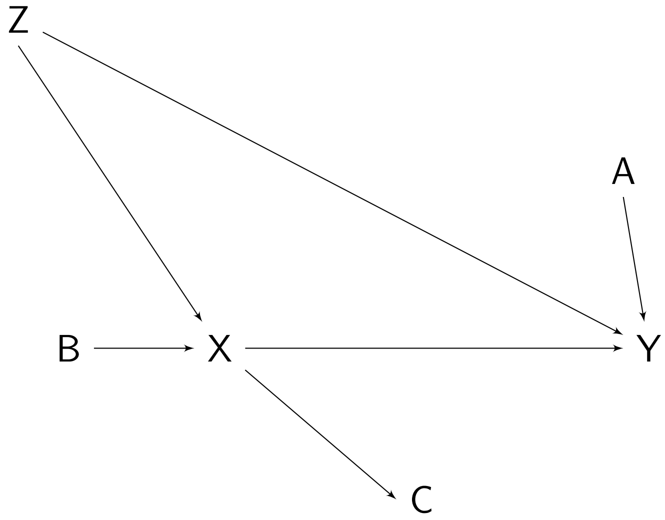
- Full discussion of causality next week
- "Correlation does not prove causation"
- Correlation is causation when:
  - X temporally precedes Y
  - No confounding
- How do we know there is no confounding?

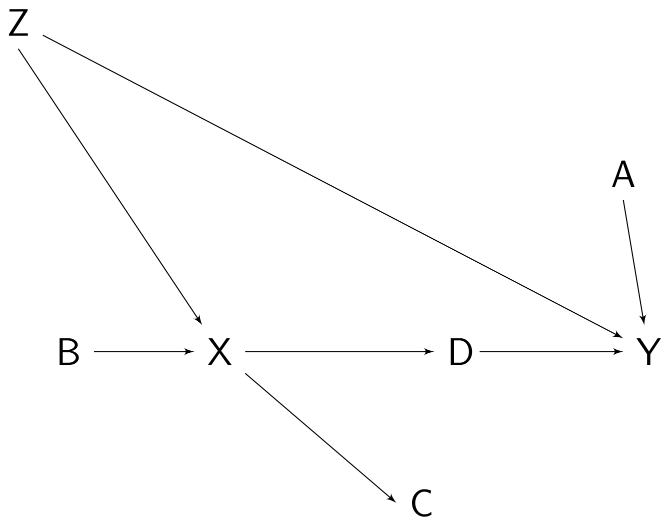
$$X \longrightarrow Y$$

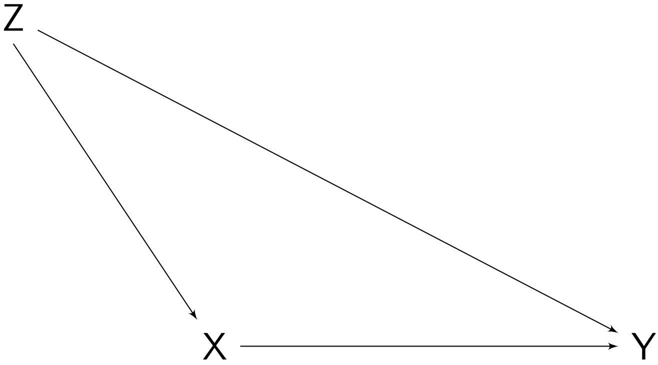

Z
X
Y


Z
A
X
Y


Z
A
B
X
Y


Z
A
B
X
Y
C


Z
A
B
X
D
Y
C


Z
X
Y

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, but we'll talk about that later

Matching I

- Example: Effect of Education on Participation
- Our design involves:
    - Measure outcome (participation)
    - 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–Pair–Share

- Does matching always get us to a clear and uncontroversial causal inference?
- Think for 15 seconds to yourself
- Then discuss with the person sitting next to you

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:

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

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

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

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 attempts to eliminate those confounds, such that:

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

Questions?

Scientific Method

1. Research question
2. Theory development
3. Hypotheses
   - Expectations about differences in outcomes across levels of a putatively causal variable
4. Design

Scientific Method

1. Research question
2. Theory development
3. Hypotheses
   - Expectations about differences in outcomes across levels of a putatively causal variable
4. Design
5. Analysis

Questions?

Course Objectives

1. Describe politically relevant research questions and hypotheses
2. Evaluate and deduce observable implications from political science theories
3. Explain statistical procedures and their appropriate usages
4. Apply statistical procedures to relevant research problems
5. Synthesize results from statistical analyses into well-written and well-structured essays
6. Demonstrate how to use Stata for statistical analysis

Software for the Course

- Stata 13
- Purchase online (200kr): http://studerende.au.dk/en/selfservice/local-it-services-and-support/it-at-bss/analytics-tools/stata/
- Available in the lab (1341–315)

Why Stata?

- SPSS is outdated
- SAS is too expensive (and too ugly)
- R is perceived as having a steep learning curve
- Stata is popular in political science and does everything we need

Textbooks and Readings

- Cameron and Trivedi
- Sønderskov
- Angrist and Pischke
- Long
- Allison
- Berry
- Chapters in compendium
- Everything else can be found online

Exam

- 7-day written exam
- Answer 1 or 2 questions applying techniques learned in the course
- Choice of 2–3 research topics and datasets
- Write-up a mini research paper

Four Assignments

- Required!
- Purpose
  - Practice techniques learned in the course
  - Receive feedback before the exam
- Schedule:
  - Essay 1 due February 27 to David
  - Essay 2 due March 20 to Thomas
  - Essay 3 due April 10 to David
  - Essay 4 due May 8 to Thomas
- On Blackboard about two weeks before deadline

Weekly Activities

- Lecture
- Partner/group activities
- Laboratory sessions

Laboratory Sessions

- Hands-on practice with Stata
- Wednesday, 14:00–16:00
- Building/Room 1341-315

Course Outline

- Research design (2 weeks)
- OLS Regression (2 weeks)
- Data handling (1 week)
- Causal inference (1 week)
- Panel and multi-level regression (2 weeks)
- GLMs (5 weeks)

Your Instructors

1. Thomas
   - Office 1340-232
   - tleeper@ps.au.dk
2. David
   - Office 1341-124
   - david.hendry@ps.au.dk

Introductions

- Please post an introduction on Blackboard

- Go to:
  Materials from students >
  Discussion Board >
  The Facebook >
  Introductions

- Post a photo and give us a short presentation of yourself