

POLS201 Spring 2019

Threats to Validity

Ian McDonald

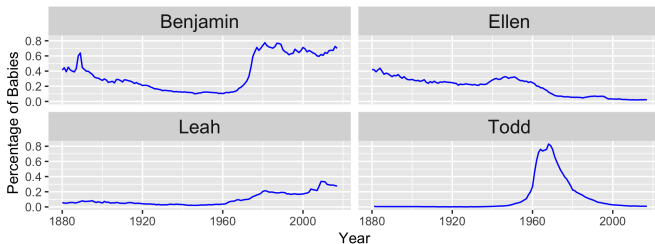
February 4

Data Visualization Tease

POLS201
Spring 2019

Ian McDonald

Percentage of Babies with Distinguished First Names
By year from 1880 - 2017



Today's Topics: Inference and the Big Threats

POLS201
Spring 2019

Ian McDonald

- Ideas to Review
 - Internal Validity: is the model/experiment consistent?
 - “External” validity means “real world relevance”
- Intervening Variables
- Confounding Variables
- Random Assignment

Today's Topics: Inference and the Big Threats

POLS201
Spring 2019


Ian McDonald

- Be warned: We have a lot of vocabulary to cover today.

The Basic Sequence of Experiment and Control

POLS201
Spring 2019

Ian McDonald

Randomize	Observe Pre-Test	Treat	Observe Post-Test
	O		O
	O	X	O

Compare
Control and
Treatment
Group
Outcomes

Compare Pre-Test with Post-Test

Threats to Internal Validity

POLS201
Spring 2019

Ian McDonald

- Internal validity
 - Are the changes in the independent variable **truly** responsible for the observed variation in the DV?
 - Threats to internal validity:
 - Violations of Random Assignment
 - Contamination Threat
 - Mortality Threat (Attrition)
 - Hawthorne Effect / Instrumentation Threat

Most Important General Threat: Endogeneity

POLS201
Spring 2019

Ian McDonald

- Contrasts with the word “exogenous”
- A factor in a causal model whose value is determined by the states of other variables in the system.
- Endogeneity ignores hidden factors and misstates causal effects.
- Endogeneity problems arise for any of several reasons.

Example: We want a model that predicts voting

POLS201
Spring 2019

Ian McDonald

- Could consider: Income, Party ID, Ideology, Education, Racial Identity
- Do some values in the model determine other values that we haven't captured?
- Is there a factor whose value is determined by other variables in the system? Or, others we haven't considered?

Exogenous Variables (The Good Kind)

POLS201
Spring 2019

Ian McDonald

- IV's that we know originate *outside* the model
- Examples:
 - Weather
 - Gender of Children
 - Unexpected "Shocks" (economic, political)
- Random Assignment will make ANY variable exogenous!!
 - Even lotteries occurring naturally.

The Important Types of Endogeneity

POLS201
Spring 2019

Ian McDonald

- Omitted Variable Bias
 - Selection Bias (Selection Effects) is one kind
 - The worst kind, in fact
- Reverse Causality

The Important Types of Endogeneity

POLS201
Spring 2019

Ian McDonald

- Endogeneity creates *bias* (i.e., results \neq reality)
 - Omitted Variable Bias
 - The bias caused by omitting a confound from your analysis
 - Example: Incumbency is a confound of a model predicting age affects candidate electoral success.
- Selection Bias (Selection Effects)
 - Observations are not categorized randomly

The Important Types of Endogeneity

POLS201
Spring 2019

Ian McDonald

- Reverse Causality
 - The dependent variable affects an independent

Endogeneity Can Produce:

POLS201
Spring 2019

Ian McDonald

- False-Positives
 - We claim there is a relationship when none really exists
 - That is: we overstate the effect of an IV
- False-Negatives
 - We claim there is no relationship when one really exists
 - That is: we understate the effect of an IV
- Both are bad; false positives are worse

False Positive Endogeneity

POLS201
Spring 2019

Ian McDonald

- A well publicized example: an experiment claims exposure to Rodin's "The Thinker" reduces self reported religiosity.
- Wrong. People in the treatment condition were aware of the control condition (contamination)

False Negative Endogeneity:

POLS201
Spring 2019

Ian McDonald

- Polarization does not exist when we consider self reported independent voters.
- Guess what? Most self reported independents are partisans in disguise. Reveal the unstated partisanship and polarization is exposed.

Colbert Bump

POLS201
Spring 2019

Ian McDonald

- The story is now 10+ years old
- House candidates appear on the old Colbert Report and see their support go up
- Consider Fowler's elaborate matching technique. Why not...
- ...just compare the votes received by candidates that appear to votes received by candidates that didn't appear?

Colbert Bump

POLS201
Spring 2019

Ian McDonald

- Consider a few questions about the Fowler article.
 - Why Study \$?
 - What was the hypothesis, IV, DV, unit of analysis?
 - Fowler offers his preferred, silly theoretical research designs. What are they? (see p. 534)
 - What does Fowler do instead, exactly?

Colbert Bump: Fowler does this:

POLS201
Spring 2019


Ian McDonald

- **Matching similar cases** A way to mimic random assignment.
- Fowler considers incumbency
- ...and donations prior to appearance (#/\$) over same time period
 - (30 days before appearing on show)
- ...and partisanship
- Experiment: Colbert appearance; Control: No Colbert
- What are the advantages and risks of this strategy? (#'s and match quality)

Colbert Bump: Pre- and Post-test Controls

POLS201
Spring 2019

Ian McDonald

Randomize	Observe Pre-Test	Treat	Observe Post-Test
	O		O
	O	X	O


Compare
Control and
Treatment
Group
Outcomes

Compare Pre-Test with Post-Test

Colbert Bump: Results (Significant or no?)

POLS201
Spring 2019

Ian McDonald

Randomize	Observe Pre-Test	Treat	Observe Post-Test
	O		O
	O	X	O

Compare
Control and
Treatment
Group
Outcomes

Compare Pre-Test with Post-Test

What's Going On With Democrats

POLS201
Spring 2019

Ian McDonald

- Possibilities:
 - Colbert targeting Democrats in trouble
 - Candidates in trouble selecting themselves
 - Candidates doing well avoid the show
- Moral: Results are interesting but endogeneity must be presumed until proven otherwise.

Two Final Perils of Modern Research

POLS201
Spring 2019

Ian McDonald

- **P-hacking** Fiddle with a model until you see a statistically “significant” result.
- There is nothing magical about the $p < .05$ standard, as you will see. It should not change your analysis.

Two Final Perils of Modern Research

POLS201
Spring 2019

Ian McDonald

- **Selecting cases on the dependent variable.**
- Suppose we're interested in a special category of voters. We naturally try to focus on those voters.
- But testing a causal theory demands we compare to the less interesting cases outside that group.
 - Sounds obvious but...
- Selecting on the dependent variable is an easy trap that creates big selection bias

The leap from research question

POLS201
Spring 2019

Ian McDonald

- For a theory and hypothesis, you now need a “take”
- Can you think of a contemporary theory about your RQ?