

Research Questions and Design

Jeffrey Arnold

April 5, 2016

Outline

1. Steps of a research Project and how 501/503 work together
2. Types of Research Questions
3. Broad Overview of Causal Inference

Steps of a Research Project

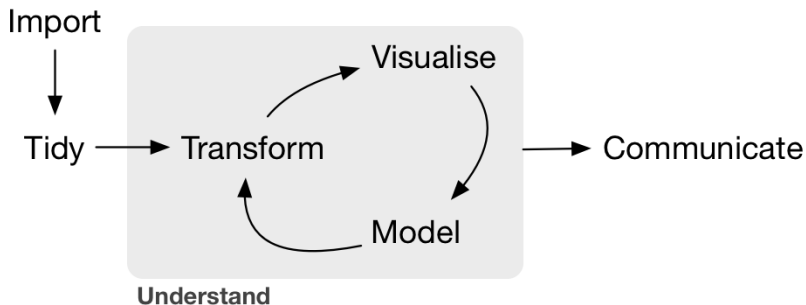


Figure 1: Hadley Wickham's Diagram Data Science

Types of Research Questions

1. Causal
2. Predictive
3. Descriptive

Casual inference vs. prediction

Consider the regression

$$Y = f(X) = \beta_0 + \beta_1 X + \beta_2 Z + \epsilon$$

1. Prediction: concerned with estimating \hat{Y} .
2. Explanation (Causal Inference): concerned with estimating β_1 .
The effect of X on Y .

Best model for prediction \neq best model for identification

Thinking about Causal Inference

something that makes a difference, and the difference it makes must be a difference from what would have happened without it. — David Lewis

Thinking About Causal Inference

1. Counterfactual: Difference between Y with and without X for each case.
2. Fundamental problem of causal inference: We don't observe the Counterfactual
3. Counterfactual model is about contribution, not attribution
4. Counterfactual models do not focus on mechanisms
5. No causation without manipulation

Broad Types of Casual Questions

Forward Effects of causes. What is the effect of smoking on the probability of lung cancer?

Reverse Causes of effects. What are the causes of lung cancer?

Statistics and econometrics focuses on *forward casual questions*

Strategies for Finding Causality (I)

1. Randomization
2. Experimental Control
3. Natural experiments (as-if randomization)
4. Before/after comparisons (fixed effects, diff-in-diff)
5. Ex-post controlling: Regression

Strategies for Finding Causality (II)

6. Ex-post Controlling II: Matching, weighting
7. Instrumental Variables (IV)
8. Regression Discontinuity Designs (RDD)
9. Process Tracing
10. Front Door Strategies (Argument by Mechanism)

Design vs. Model Based Approaches to Causal Inferences

- ▶ Design based approaches: as-if-randomization or a data-generating process that creates comparable units.
- ▶ Model based approaches: models the data-generating process
- ▶ Recent preference for design based approaches: experiments, natural experiments, *regression discontinuity*, *IV*, *diff-in-diff*
- ▶ Even in models, preferences for more “robust” models: weighting, matching.

Potential outcomes effect on political science research

- ▶ preferences for types of model
- ▶ reinterpretation of existing models
- ▶ focus on “identification strategies”

Thoughts on Causality and Methods

1. Not all research questions are causal, nor should they be!
2. Be clear about what question you are asking, and that your method is appropriate
3. Same method can be used for causal inference and prediction, but different interpretation and assumptions necessary.

This course

1. Linear Regression

1.1 Mechanics

1.2 Causal and predictive interpretations of it

2. Other approaches to causal inference

2.1 Matching and weighting

2.2 Instrumental variables

2.3 Panel data (diff and diff, fixed effects)

2.4 Regression discontinuity