# Introduction to Empirical Methods: RBSI

Lecture 1, Part 1: Science, Theory, and Causation

# Political "Science"

- Causal Inference vs. Description
- (Policy) Implications require causal explanation.

# How do we know? (aka Epistemology)

- Authority
- Tradition/Precedent
- Mutual agreement
- Intuition
- "Common sense"
- Observation (empirics)

# Observation's Benefits

- Verifiable & Repeatable
- Falsification
- Probabilistic Claims

# Political Science Defined:

- Study of politics based on theory and observation.
- We seek to generalize from our observations.
  - Example: Notice that individuals with more years of formal education turn out to vote more often than those with less education.
  - May generalize that education increases one's propensity to turn out to vote.

# Goal of Scientific Research

- Inference (both Causal and Descriptive)
- Definition: best "guess" about an unknown given known information.
  - Causal Inference: Unknown causal relation between two or more variables.
  - Descriptive Inference: Unknown fact about a single variable.
- Inference is, by definition, uncertain (it contains error)

# Causal Inference

- Causation cannot be observed: it can only be inferred.
- Correlation: when the values of two or more variables tend to move together.
- Valid Causal Inference requires:
  - theory that expects X to affect Y, and
  - evidence of correlation between X and Y.
- We *infer* causal relationship in this case.
- All causal inferences are uncertain.

# Causation versus Correlation versus Prediction

- Correlation: the values of two variables tend to move together.
- Prediction: knowing the value of X helps us predict the future value of Y.
- Causation: change in the value of one concept tends to produce change in the value of another concept.
- Causation (X causes Y) implies correlation (values of X and Y tend to move together), but correlation does not imply causation!

# Deterministic versus Probabilistic Causation

- Deterministic causal relation: cause (X) is always present when outcome (Y) occurs
  - Unit of interest (individuals, states, countries)
  - Variation
  - Both quantitative and qualitative judgments
- Probabilistic causal relation: cause (X) usually present when outcome (Y) occurs
  - Outcome occurs with some likelihood when cause is present
  - Political science seeks to identfy probabilistic causal relationships

# Descriptive Inference

- Good description offers us a profile of interest, including:
  - Unit of interest (individuals, states, countries)
  - Variation
  - Both quantitative and qualitative judgments
- Trying to *infer* characteristics of a population from those of a sample.
- Description is not the goal of science: explanation of causal relationships is the goal.

# Why can't we just observe as scientists?

- Observation without theory is inadequate for causal inference. Why?
  - Inaccurate observations
  - Tendency to overgeneralize
  - Selective observation of phenomena
  - Illogical reasoning
- To draw a valid causal inference we require both logical and empirical support.

# Criteria of Scientific Method

- Evidence
  - Are the relevant data consistent with the expectations of your theory?
- Reason
  - Do the hypotheses from your theory follow from your assumptions?
- Emotion and passion influence scientists, but do not overrule evidence and reason if one adopts the scientific method.
- Mutual agreement about the criteria of the scientific method prohibits emotion and passion undermining accumulation of knowledge.

# Requirements for Scientific Study of Politics

- Analyze the real world (not counterfactuals)
- Think in terms of concepts (or constructs), not proper nouns
- Propose testable causal relationships among concepts.
  - Must be possible to falsify the proposed relationship.
  - Must know what evidence is needed to possibly falsify hypothesis.

# Fallacy of Affirming the Consequent

- Why can't we "prove" our theory with evidence?
- Using evidence to "prove" a hypothesis is a logical fallacy:
    - If A, then B ← Hypothesis
    - Observe B ← Evidence
    - Conclude A is true ← Fallacy of Affirming the Consequent
- If the evidence is consistent with our theory, we infer (or conclude) that our theory is not wrong: we failed to falsify the theory.

# A Science of Politics

- The political and social worlds produce observable pheonomena that we can conceptualize and then measure: we can produce political and social data.
- Political scientists seek to identify, and then explain the occurrence of, patterns among political and social phenomena.
- To do so we develop causal explanations (aka theory) about the cause and effect relationships among concepts, and then test the hypotheses from those theories by examining relevant data.

# Moving from Description to Causal Explanation

- We develop models, or theories, to explain patterns we observe.
- Model defined:
  - Simplified "snapshot" of reality
  - Provides explanation of causal relations
  - Composed of interrelated concepts.

# Causal Theory, I

- Answers the questions: What causes something, and why?
- A theory is a set of interconnected statements that identify what causes something and why.
  - Dependent Variable: the "something" being caused (the effect or outcome)
  - Independent variable: the suspected cause

# Causal Theory, II

- Provides a story about how and why change in the value of one concept influence the values of another concept
    - Should also offer additional implications (or hypotheses)

# Deduction versus Induction

- Deduction: theory first, then empirical testing.
- Induction: Observation first, then theory construction to explain what was observed.
  - Need to test further implications of theory in this case.

# Hypothesis

- A hypothesis is an expectation about what we will observe.
  - More specifically, it is a statement, derived from theory, that describes the expected co-variation between the values of two (or more) concepts.
    - Positive relationship: the values of x & y move in the same direction (x↑ & y ↑, or x↓ & y ↓)
    - Negative relationship: the values of x & y move in opposite directions (x↑ & y ↓, or x↓ & y ↑)
- Put another way: a hypothesis is a testable (i.e., falsifiable) proposition.
- Specify nature of expected relationship (not just the sign):
  - More explicit hypotheses are easier to reject (rule out).

# Null Hypothesis & Hypothesis Testing

- The expectation of no relationship between two concepts ($x\uparrow$ & y varies randomly or does not change value).
- A null hypothesis is implied by every hypothesis.
- Statistical hypothesis testing relies on rejection of the null hypothesis (i.e., if we are able to reject the null, that provides additional support for the (research) hypothesis).

# Falsifiability

- It must be possible, in principle, to collect and evaluate evidence that permits us to demonstrate that the hypothesis is not consistent with the evidence.
- If the data permit us to reject the null hypothesis, then we can infer that our hypothesis is not (yet) wrong. That implies that we continue to use the theory.
- This implies that all theory is provisional: future evidence may lead us to accept the null hypothesis
- Put another way: there is no number of tests a theory can "pass" and then become "proven."

# Operationalization (measurement)

- To test hypotheses we must measure our concepts.
  - Measurement permits us to examine data and evaluate co-variation.
- First, we need an abstract definition of our concept.
- Second, we produce an operational definition: a set of instructions that describe how to measure the value of that concept.

# Evaluating Operational Indicators

- Validity: does it match the concept well?
- Reliability: do we get the same value when we repeat measurement?
- Error can be broken into two parts:
  - Systematic: tendency to assign values that are either too high or low
  - Random: equal likelihood of assigning too high and too low values

# Model Building

- Identify a question or puzzle.
- Identify the dependent variable
  - Propose a causal explanation (theory)
- Identify the independent variable(s)
  - State the hypothesis
- Identify the null hypothesis
- Operationalize the concepts (collect data)
- Test the hypothesis
- Re-evaluate the theory given the results.

# Voter Turnout Example

- *Question*: Why do some people vote?
- *Causal Theory*:
  - Education exposes people to more information.
  - Exposure to political information increases awareness of problems.
  - Awareness increases likelihood that someone votes.
- *Hypothesis*: The more education a person has, the more likely she is to vote.
- *Null Hypothesis*: The amount of education a person has is unrelated to the likelihood that she votes.
- *Operational Indicators*: conduct a survey and record
  - Vote (y): 0 if the person did not vote in most recent election and 1 if the person did.
  - Education (x): the highest school grade the person completed.

- "Theories are nets cast to catch what we call the world, to rationalize, to master, and to explain it. We endeavor to make the mesh ever finer & finer."
  - Karl Popper, *The Logic of Scientific Discovery*

# Before Theory

- Ask a Question or Observe/Identify a Puzzle
- Goal: General Explanation
  - What is the general phenomenon you are seeking to explain?
  - Think in terms of concepts, not specific examples
- Primary interest:
  - Explain change (a.k.a. variation) in the phenomenon of interest (a.k.a. dependent variable)

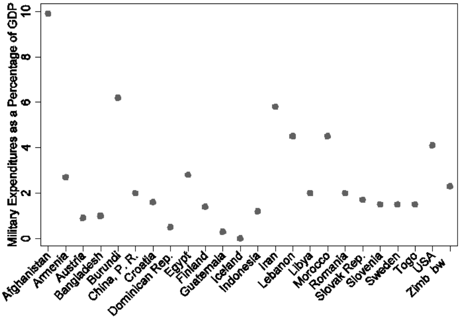
# Space & Time

- Identify the spatial dimension
  - Citizens? Students? Governments? Countries?
- Identify the temporal dimension
  - Annual? Monthly? Weekly?

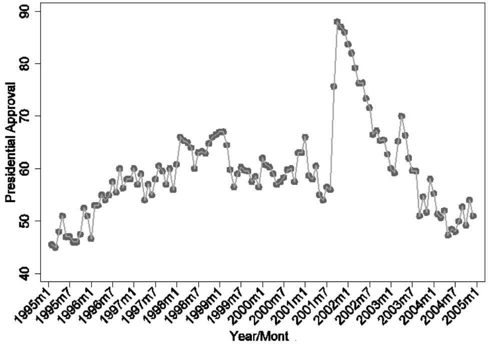
# Dimensions and Variation

- Does the variation we are interested in occur over time, across units, or both?
- Spatial Variation
  - Multiple units are measured at one moment in time
    - Cross Sectional (e.g., # of Terror Attacks in each country in the world, in 1999)
- Temporal Variation
  - Repeated measurement of one unit at different moments in time
    - Time-Series (e.g., # of Terror Attacks per year in USA, 1950-1999)

# Cross-Sectional Example



# Time-Series Example



# What do we need to understand to build our theory?

- Actors or units of interest
- What motivates the actors?
- What constrains the actors?
- How to apply logic
  - What are we willing to assume?
  - What do those assumptions imply?
- Can we generalize to a larger population?

# Define Your Concepts

- Conceptual Definitions
  - The meanings we assign to terms
  - Definitions are a shared understanding
- Do not confuse concepts with reality
  - Concepts are our perception of "things" with common characteristics
  - Concepts are tools used to describe the world

# Concepts and Measurement

- Conceptual definitions influence the way we measure the concept.

# An Example: Political Knowledge

- What do we mean by how much someone knows about politics?
    - Political Awareness
        - How attentive is the individual to politics.
        - Often measured by knowledge of Civics and Current Events
    - Political Sophistication
        - Ability to reason through political questions
        - Does the person think ideologically?
        - Often measured by questions about positions on multiple issues.

# Review the Existing Research

- How have others approached your question?
- What elements have they missed?
- Why do this?
  - Science is the accumulation of knowledge
  - Conceptual definitions: what is accepted?
  - Operational definitions: what is accepted?
  - To replicate past research and make it easier for others to replicate your research

# Thinking Formally

- Map out the process by which $x$ influences $y$
  - Focus on how and why
- Formal Theory and Social Science
  - Begin with assumptions, then use math or formal logic to determine implications.

# Rational Choice: a widely used set of assumptions

- Core assumptions: People. . .
  - Form well-behaved preferences over possible outcomes
  - Calculate the expected outcomes given different choices
  - Maximize their expected utility (i.e., select the choice they believe will make them best off)

# Why Theorize Formally?

- It forces you to make your assumptions explicit
- And establishes that implications follow logically from assumptions

# Why Vote? An Application of Formal Theory

- Is the benefit one gets from voting greater than the costs of voting?
- Costs of Voting
    - Registering
    - Going to polls
    - Information collection
- Benefits of Voting
    - Influencing government policies
    - Satisfaction of doing one's civic duty

# Should you (an Ernie supporter) vote?

**You Don't Vote**

| Bert | Ernie | Winner |
|---|---|---|
| 5 | 3 |  |
| 4 | 3 |  |
| 4 | 4 | Tie |
| 3 | 4 |  |
| 3 | 5 |  |

**You Vote**

| Bert | Ernie | Winner |
|---|---|---|
| 5 | 4 |  |
| 4 | 4 | Tie |
| 4 | 5 |  |
| 3 | 5 |  |
| 3 | 6 |  |

# What is the probability your vote will make the difference?

- If the electorate is large, almost no chance.
  - Though information is easy to obtain
- If the electorate is very small, then maybe.
  - Though information is harder to obtain
- Rational abstention
  - If your vote is unlikely to influence the outcome, it is typically "rational" to choose not to vote
  - Why? Because the costs outweigh the benefits
- Can focus on duty to vote, or
- Can develop alternative "boundedly rational" theories that illustrate why people vote.

# Non-Empirical Evaluation

- Empirical tests of hypotheses are not the only way in which we evaluate theories: we also evaluate them on logical and other grounds.

# Evaluating Your Theory, I

- Is your theory causal?
  - It should explain how and why change in the values of the independent variable change the values of the dependent variable.
- Does your theory generate testable hypotheses?
  - For a theory to be testable, it must be falsifiable

# Evaluating Your Theory, II

- Keep it simple
  - Connect $x$ & $y$ via the shortest explanatory route
  - Parsimonious theories are easier to use
  - Occam's Razor
- Is your theory novel and interesting?
  - Your theory should make new predictions
  - Your theory should not propose explanations that are obvious to all
  - This can mean it yields counterintuitive results.
  - Or that it clarifies underlying causal mechanisms that had previously been unclear.

# Puzzles

- What is a puzzle?
  - We expect two cases to have similar outcomes, but they have different outcomes.
    - or vice versa
- Why begin with a puzzle instead of a question?
  - Requires some knowledge of events.
  - Grounds our theory in concrete cases.
  - Makes us look for non-obvious causes.
- Be careful: They may also make us focus too much on rare events.

# Microfoundations

- What are the microfoundations?
  - Who are the actors?
    - What are their goals?
- What are the relevant institutions?
  - Institutions constrain actors' options
  - Institutions influence actors' preferences
  - Institutions affect actors' incentives

# Assumptions

- What assumptions inform the theory?
  - Make your assumptions explicit.
    - Write them out formally or in prose.
- Assumptions are inherently. . .
  - Simplifications
  - Generalizations
    - We want to identify central tendencies.

# Generalization

- Make your nouns and verbs more general
  - The president wants to be reelected
  - Presidents want to be reelected
  - Politicians want to be reelected
- One cautionary note: Don't overgeneralize
  - Theory and previous research should inform how general we can be

# Check the Logic

- Is the form of the argument valid? (i.e., does the conclusion follow from the premises?)
- Are the premises valid? (i.e., would more accurate assumptions lead to different conclusions?)
  - Assumptions:
    - Politicians want to retain office.
    - The number of supporters one requires depends on whether free and fair elections exist.
    - Politicians gain support by providing supporters with goods and services.
    - Public Goods are goods that cannot be denied to anyone, regardless of support.
  - Implication:
    - Politicians who face elections will provide more public goods than those who do not.
- NB: This is done *prior* to hypothesis testing and does not involve data

# Benefits of Explicitly Stating Our Assumptions

- Explicit statement of our assumptions leads us to think precisely about our concepts
  - What are the precise definitions?
- Thinking about the assumptions could lead to promising lines of research
  - Are the assumptions in a well-known theory flawed?
    - Assumptions do not always hold in all cases.
    - What are the implications if the assumptions do not hold?

# Information Theory of Democratic Peace

- Assumptions:
  - Bargaining failures lead to war
    - Failures are caused by uncertainty about resolve
  - Democratic leaders are more accountable than non-democratic leaders
  - Accountability increases audience costs
  - As audience costs increase, uncertainty about resolve decreases
  - As uncertainty about resolve decreases, states adopt less escalatory strategies

# The Theoretical Conclusion

- Pairs of democracies are less likely to experience military conflict

# Theory

- Set of logically related symbols that represents what we think happens in the world.

# Hypothesis

- A testable statement of relationship, derived from a theory.

# Falsification Revisted

- We attempt to find evidence to be able to "reject the Null Hypothesis."
  - We do not find evidence to accept the research hypothesis.

# Does the Death Penalty Deter?

- Theory: People consider the consequences before taking action.
- Hypothesis: The death penalty should be an effective means of deterrence.
  - Research Hypothesis ($H_1$): The Death Penalty Reduces Crime.
  - Null Hypothesis: ($H_0$): The Death Penalty Has No Effect on Crime.

# A Case Study of State Education Spending

- Our theoretical proposition begins with identifying the key actors and assumptions
- Actors
    - State elected leaders make education expenditure decisions
- Assumptions about the actors
    - State elected leaders value being reelected
    - State elected leaders seek to maximize their chances of getting reelected
    - State elected leaders respond to the expenditure demands of their constituencies

# State Education Spending Continued

- More assumptions
  - The constituency is able to make expenditure demands
  - Meeting expenditure demands is not costless
  - Costs of expenditures are conditional on state policy factors and financial conditions

# Our Theory

- State educational spending is a function of political demands and financial costs.
- This theory was developed deductively.
- We began from assumptions, then deduced the implications for our dependent variable.

# Operational Definitions

- Variables to operationalize
- Dependent Variable
  - State education spending ($y$)
- Independent Variables
  - Political Demands ($x_1$)
  - Financial Costs ($x_2$)

# State Spending on Education

- Total $ spent by state on education
  - But this might include higher ed. (college, universities etc.)
- Total $ spent by state on elementary and secondary education
  - But a state with more children spends more—that doesn't mean that they are devoting more money.
- $ expenditures per pupil in elementary and secondary education across the states
  - But this might include administrative costs
- $ expenditures per pupil on instruction (remove administration expenditures) in elementary and secondary education across the states

# Independent Variables

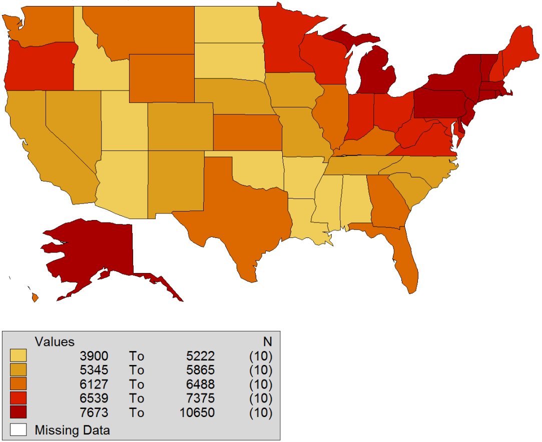
- Political Demands
    - % Congressional Delegation that is Democratic
- Financial Costs
    - State Wealth: Per capita Income

# Our Hypotheses

- $H_1$: As political demands increase, state spending on education will increase
  - As the congressional delegation is more Democratic, state spending on education will increase
- $H_2$: As financial costs decrease, spending on education will increase
  - Wealthier states will spend more on education
- What are the null hypotheses?

# State Education Expenditures

$PER PUPIL -- 1998: EXPENDITURE PER PUPIL IN AVERAGE DAILY ATTENDANCE IN PUBLIC ELE
SCHOOLS (SA, 1999)



# Evaluating Our Theory

- We take each hypothesis separately
- Can we reject the first null hypothesis?
  - How would we know?
    - If spending significantly increased as political demands increased
- Can we reject the second null hypothesis?
  - How would we know?
    - If spending significantly decreased as financial costs increased

# Models are incomplete simplifications

- Theoretical models: often bivariate
  - District Opinion → Representative's Vote
- Reality: multivariate
  - District Opinion → Representative's Vote
  - Party stance → Representative's Vote
  - Personal preferences → Representative's Vote
- Or interactive:
  - Preferences alter effect of District Opinion on Vote

# Models are probabilistic I:

- independent variable (x) contributes to change in the dependent variable (y)
    - DV: variable w/ values influenced by other variables
    - IV: causally prior to DV; influences values of DV
        - Ex: As frustration (IV) increases, so does individual aggression (DV)

# Models are probabilistic II:

- x may increase or decrease the probability of y occurring (or changing)
- we cannot directly claim that x causes y based on observation
- by eliminating rival explanations, we can *infer* that x causes or contributes to y
- any effect is *on average*—a change in x need not lead to a change in y in every single case

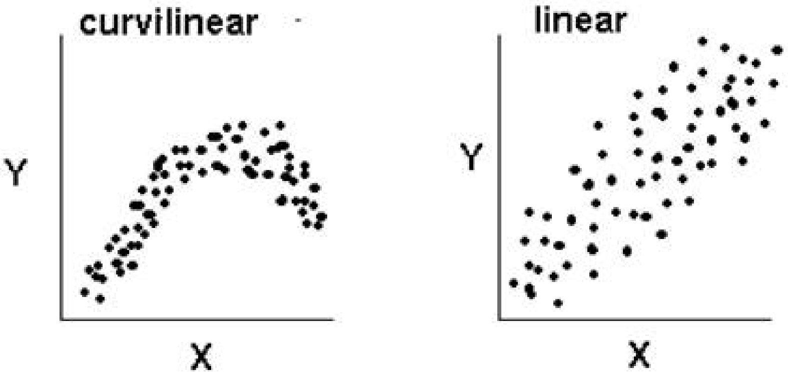
# The Primacy of Inference

- Causal relationships can, at best, be inferred:
  - Never observed
  - Never "proven"
    - Good research design: best tool for overcoming threats to causal inference

# Theory must specify type of causal relationship:

- Direct: $x \rightarrow y$
- Indirect: $x \rightarrow z \rightarrow y$
- Spurious? $z \rightarrow x$ and $z \rightarrow y$
- Linear or Curvilinear (see figure below)?

Figure: Curvilinear v Linear Relationship between x and y



# Inferring Causation: Four Hurdles

- A credible causal mechanism connecting values of X to values of Y
- Could Y cause X?
- Do values of X covary with values of Y?
- Does a third variable, Z, influence both the values of X and Y?

# Non-causal theory

- Does not explain how and why x moves or changes y, so no causality to assess
- It must be possible for x to cause y
  - e.g., change in x must precede observed change in y
- Solution: Provide a theory to explain how and why changes in the value of x will tend to produce changes in the value of y.

# Endogeneity: (reverse causation: not x $\rightarrow$y but y $\rightarrow$x)

- If x contributes to change in y, then y should not also contribute to change in x
  - Ex: States with more interest groups (x) tend to have more citizens' initiatives (y). However, states with more citizens' initiatives (y) foster the formation of additional interest groups (to sponsor countermeasures).

# No covariation (between x & y)

- if x causes y, y must tend to change when x changes
- necessary but not sufficient for inferring causation

# Spurious Correlation

- x covaries with y, but true causal mechanism overlooked
- Another variable, not considered, is moving both x & y
- Unfortunately, cannot control for all possible intervening variables
- However, must attempt to control for as many competing (theoretically relevant) explanations for DV as possible
    - Ex: In 1960s higher coffee consumption was associated with higher probability of developing lung cancer.
    - Does coffee cause lung cancer?
    - Spurious correlation: coffee drinkers more likely to be smokers in 1960s; smokers more likely to develop lung cancer