Research Design

September 28, 2022

Today

- Data Collection
  - Different Approaches & overview of data types
  - Observational Research vs. Experimental Designs
- Population & Sample—using sample data to make inferences about a population
- Fundamentals of (controlled) experimental design
  - Randomization
  - Matched pair designs & block designs
  - Strengths & limitations of experimentation
- Sampling & the foundations of statistical inference

Sample & Population

- Hypotheses are predictions about effect of X on Y in the **population** of cases
  - Examples—country years, individuals, registered voters, etc.
- We generally only examine relationship between X and Y in some **sample** of cases
- Population—the universe of subjects the researcher wants to describe
  - **Parameter**—the actual (unobserved) value in the population
- Sample—a select number of cases, or observations, drawn from a population
  - **Sample Statistic**—estimate of the population parameter from the sample data

Parameter and Statistic

- We are interested in estimating a population parameter using a sample statistic
- Example—Gallup Presidential Job Approval Rating (September 1-16, 2022)
  - "In general, do you approve or disapprove of the way Joe Biden is handling his job as President"
    - 42% approve, 56% disapprove, 2% not sure
- What are the actual values in the population?
  - Don't know, because we don't observe the population
- We want to generate the best estimate possible
- Research design is key to generating a sample statistic that gives the best (and unbiased) estimate of some population parameter

Data Collection

- Test hypotheses using sample data
- Data can come from multiple sources
- Two general approaches:
  - Observational research: researcher records data on individuals without attempting to influence responses
  - Experimental study: impose a "treatment" on individuals and record their responses
- Primary difference—ability of researcher to control the values (or application) of IV
  - If unable to (as in observational research), need to worry (a lot) about control variables

Data Collection in Observational Research

- Many ways to obtain data for observational research:
  - Existing Datasets
  - News reports
    - Human coding
    - Machine coding
  - Archival research
  - Expert surveys
  - General surveys
- Standard for data collection: **replication**
  - Other researchers should be able to use same the rules and get the same data
  - Important to assess inter-coder reliability if there is any discretion, or judgement, involved in the coding

Experiments vs. Observation

- For establishing (or at least approaching) *causal* inferences, experimental designs are far superior (than observational studies)
- A central challenge—how to rule out the potential impact of "confounding" factors
- Two variables are "confounded" when their effects on a response variable cannot be distinguished from each other
- Well-designed experiments take steps to manage, or *perhaps* eliminate, this problem

Basics of Experiments

- Individuals in an experiment: **experimental units**
  - E.g. human subjects
- Intervention—a **treatment**, or factor
  - Example: one group of people placed on a diet/exercise program for six months (**treatment**)
  - Compare treatment subjects' blood pressure (i.e., the DV) to other subjects who did not diet or exercise
- If the experiment involves two different degrees of the diet/exercise program, you are testing two levels of the treatment
- A response to a treatment is "statistically significant" if it is different than one would expect by mere chance (i.e. random variation among the subjects)
  - This insight is key—we will discuss it in a few weeks

Randomization

- It is critical in an experiment to randomly assign individuals to a treatment vs. control category
- Randomization means any unit has equal chance of ending up in the control or treatment group
- Randomization produces two groups of subjects that we expect to be similar in all respects before treatments are applied
  - Comparability of the experimental groups, *but for the treatment*, allows one to isolate any causal effect of the treatment
    - Should eliminate potential for confounding effects

Comparative Experiments

- Experiments are comparative in nature (across subject groups)
  - Compare the response to a treatment to:
    - Another treatment
    - No treatment (i.e., a control group)
    - A placebo
    - Or, some combination of the above
  - Control—no treatment is administered
    - Serves as reference mark for an actual treatment
  - Placebo—a fake treatment is administered
    - Tests the effect of an actual treatment, not subject's apparent/perceived treatment
  - Double-blind: Neither researcher nor subject knows whether a subject is receiving placebo or treatment

Comparative Experiments

- Design of a study is **biased** if it systematically favors certain outcomes
  - Our sample statistic is no longer a good estimate of the population parameter
- Best way to exclude biases from experiment is through random design
- Best way to assure the robustness of an experiment is to replicate it
  - Ensures that particular results not due to uncontrolled factors, specifics of experimental settings, or errors of manipulation

Three big ideas of experimental design

- **Control** the effects of lurking variables on the response, simply by comparing two or more treatments
- **Randomize** group assignment-use chance to assign subjects to treatments
- **Replicate** each treatment on enough subjects to reduce chance variation in the results

Design Types

- Completely randomized experimental design
  - Individuals are randomly assigned to groups, then the groups are randomly assigned to treatments
- Block, or stratified, design
  - Subjects are divided into groups (or blocks)—prior to randomizing the control vs. treatment—to test hypotheses about differences between the groups
    - E.g. demographic groups
- Matched pairs:
  - Choose pairs of subjects that are closely matched
    - e.g., same sex, height, weight, age, and race.
  - Within each pair, randomly assign who will receive which treatment.

Experiments and Social Sciences

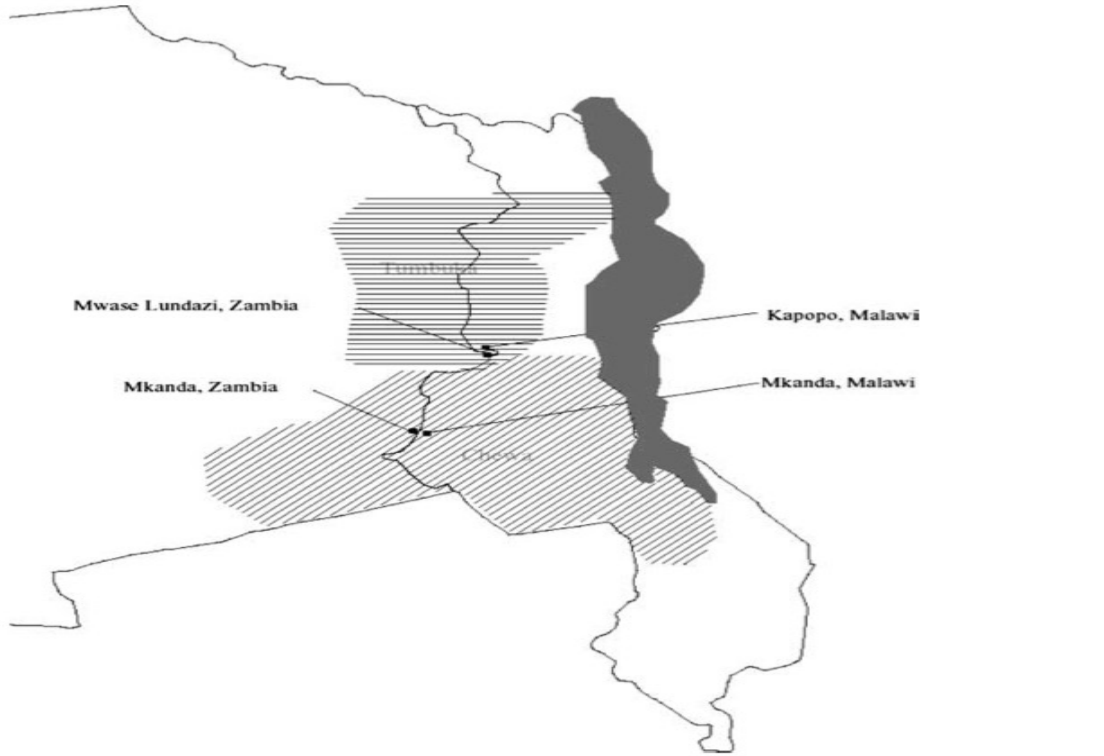
- True randomized experiments are the "gold standard" for research
  - Best means to isolate the effect of a treatment on a response variable
- But, they can be difficult in social sciences
  - Often not possible
  - If possible, often not feasible
    - E.g. Expensive, temporal concerns—might be interested in long-term effects (or, do the effects persist?)
  - If feasible, often not ethical
  - Even if feasible, can be hard to generalize—external validity
    - Ex. Crisis bargaining among college students
- Much of political science research uses observational data

Experiments in Political Science

- If true experiments are impossible, we can use then logic of experimentation to approximate experimental design in observational research
- Natural (or, quasi-) experiments:
  - Nature, not the researcher, controls the distribution of the IV
  - But, can be useful if the argument can be made that distribution of the treatment is unrelated to other factors
- Example: Chewas & Tumbukas in Malawi/Zambia
  - See Dan Posner (APSR, 2004) "The Political Salience of Cultural Difference: Why Chewas and Tumbukas Are Allies in Zambia and Adversaries in Malawi"
    - Country borders as the "quasi-treatment" & assessing the impact of cultural cleavages on different sides of the border

Random Assignment?

FIGURE 1. Research Sites



Experiments in Political Science

- How to combat concerns over external validity?
- An alternative approach: field experiments
  - Conduct experiments with subjects who more closely resemble those of interest
  - Tradeoff—better external validity, but often less control over confounders
- Many examples in American political behavior research—e.g., canvassing & turnout
  - E.g. Gerber & Green (American Political Science Review, 2000), "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment"

Experiments and Political Science

- Experiments and observational research each have advantages
  - Experiments are ideal for testing the effect of *X* on *Y*—i.e., better test of causality
  - But, experiments are often not feasible, and may have problems of external validity
  - Observational research allows us to address a wider range of questions
- Can be ideal to combine methods (when feasible)
  - If different methods point to similar conclusions, that suggests greater support for an argument

Sampling

- Important to draw a sample that represents population
  - If sample is *systematically* different, hard to generalize to population
- Consider the potential for biased sampling & avoid additional error
  - E.g. Literary Digest Poll & the 1936 presidential election—a large, but bad, sample
- Three sources of error:
  1. **Selection/sampling bias**—some individuals are more likely to be included in the sample than others; some may be excluded altogether
  2. **Response bias**—some individuals are more likely to be measured than others
  3. **Random sampling error**—the extent to which a sample statistic differs by chance from the population parameter

Sampling Methods

- Convenience sampling: Just ask whoever is around (e.g. college students)
  - Potential bias: Opinions limited to individuals present
  - Likely to have sampling bias, sample not representative of population
- Voluntary Response Sampling: Individuals choose to be involved
  - Bias: Sample design systematically favors a particular outcome
  - Some individuals more likely to be measured than others
- Probability or random sampling: Individuals are randomly selected; no one group should be over-represented.

Simple Random Sample

- A Simple Random Sample (SRS) is made up of randomly selected individuals
  - Each individual in population has the same probability of being in the sample
  - All possible samples of size N have the same chance of being drawn
- Difficult to do in practice:
  - Internet surveys?
  - Mail surveys?
  - Phone surveys?
  - In-person surveys?

Stratified Sample

- A stratified random sample is a series of SRSs performed on subgroups (i.e. strata) of a given population
  - Subgroups are chosen to contain all the individuals with a certain characteristic
- Multistage samples use multiple stages of stratification.
  - Often used by government to obtain information about the U.S. population.

Challenges for Survey Sampling

- Undercoverage
  - Some groups in the population are left out of the process of choosing the sample
- Nonresponse
  - People who feel they have something to hide or don't like their privacy being invaded may not answer
    - But, they are still part of the population
- Response bias:
  - Lying
  - May happen when questions are personal, related to the past, or refer to taboo activities
- Wording Effects: Some questions may prompt people to give a particular answer
  - Can be unintentional

Toward Statistical Inference

- The techniques of inferential statistics allow us to draw inferences, or conclusions, about a population from a sample
  - Your estimate of the population is only as good as your sampling design
- Work hard to eliminate biases
  - A sample only produces one estimate
  - If randomly sampled again, we would probably get a somewhat different result
- The larger the sample, the better (all else equal)
  - But, again, sample size cannot compensate for poor sampling

Toward Statistical Inference

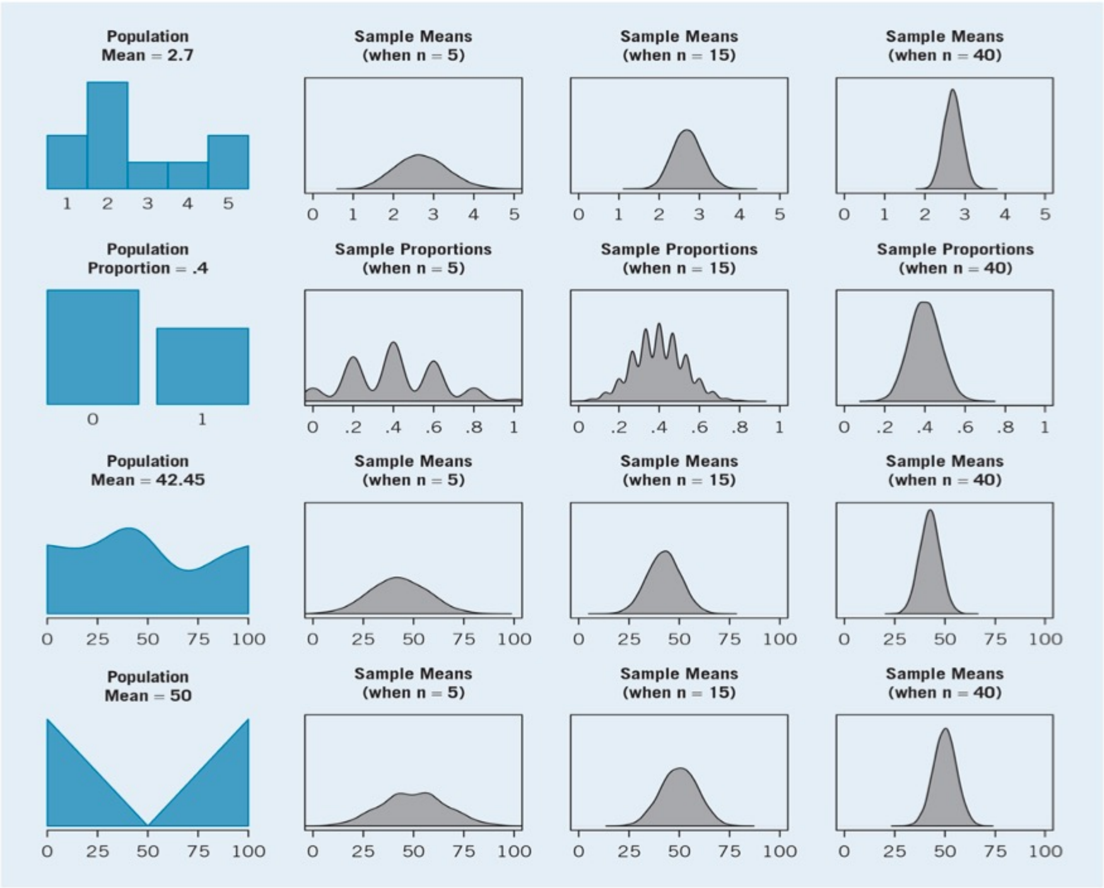
- Each time we take a random sample from a population, we are likely to get a different set of individuals and calculate a different value for the sample statistic of interest
  - Sampling variability
- Good news: If we take lots of random samples of the same size from a given population, the variation from sample to sample (the sampling distribution) will follow a predictable pattern.
  - All of our statistical inference (and the rest of the course) is based on this knowledge

Toward Statistical Inference

- The variability of a statistic is described by the spread of its sampling distribution
  - Spread depends on:
    - Sampling design
    - Population variance
    - Sample size—larger sample sizes leading to lower variability
- Statistics from large samples are almost always close estimates of the true population parameter.
  - However, this only applies to random (unbiased) samples

Sampling Size & Sampling Variability

Figure 6-3 Illustration of the Central Limit Theorem



Next Steps

- No new material next week
  - We will discuss several applications of what we have learned here
    - Please take a look at the two syllabus readings
  - Review for the midterm exam—come prepared with any questions
- Homework 3 due next week
- Midterm Exam #1 assigned next week, due Monday, October 10
- After the exam, probability theory and the foundations of statistical inference