INSTITUTE *for* RESEARCH *on* POVERTY
UNIVERSITY OF WISCONSIN–MADISON

Quantitative Methods

Lonnie Berger

Defining Quantitative and Qualitative Research

- **Quantitative methods:** systematic empirical investigation of observable phenomena via statistical, mathematical, or computational techniques generally with the purpose of often uncovering generalizable and/or causal patterns or relations; can be descriptive or explanatory in nature
  - emphasizes "objective" measurement, highly structured data collection and analytic techniques, and data that is analyzed in numerical form
- **Qualitative methods:** systematic empirical investigation that can be exploratory or explanatory in nature and intended to gain understanding of (often) subjective phenomena (e.g., underlying reasons, ideas, opinions, and motivations) that may be difficult to directly observe and/or quantify
  - Data collection tends to be semi-structured and often with a nonrandom sample of limited size, data is typically analyzed in textual form (though some aspects may be counted/quantified)

Quantitative Research

- Quantitative research is most often focused on describing or explaining a phenomenal in terms of direction, magnitude, and precision of estimation
- **Descriptive**: identifying and quantifying trends, group differences, correlations, associations
- **Explanatory**: implies identifying and quantifying causal effects
- Language is important: causal language (effects, impacts, etc.) should only be used in the context of rigorous causal identification; descriptive language (correlations, associations) should be the default
- Determining whether a relation is correlational or causal is crucial for informing policy

What do we mean by causality?

- increased probability of some outcome as a direct result of some condition (treatment)
- What is a causal effect? ideally, the difference in some outcome for the exact same individual observed in two different conditions at the exact same time, all else equal
- Rubin-Rosenbaum(-Holland) theory (in short): Assume a causal variable or treatment with 2 possible values: experimental (E) and control (C) and an outcome (Y). If an individual receives E, we observe $Y_i(E)$; if the individual receives C, we observe $Y_i(C)$. The causal effect of the experimental treatment for that person (relative to the control condition) is the difference between these 2 outcomes: $\Delta_i = Y_i(E) - Y_i(C)$.

INSTITUTE for RESEARCH on POVERTY
UNIVERSITY OF WISCONSIN–MADISON

Rubin-Rosenbaum(-Holland) theory: Key take-aways

- the causal effect (Δ_i) is defined **uniquely for each individual** (i.e., the impact of the treatment varies across individuals); that is, it is the difference between how that individual would respond under a given condition versus under another condition – therefore, it varies across individuals;
- the causal effect **cannot be observed** (or directly computed) because an individual can only be assigned to one condition (note that the counter-factual outcome is that which is not observed);
- it must be possible to imagine a scenario in which the individual **could have received either E or C** (thus, a fixed attribute of an individual cannot typically be a "cause" – we cannot realistically know how a particular girl would be different in some regard if she were a boy, etc.)

INSTITUTE for RESEARCH on POVERTY
UNIVERSITY OF WISCONSIN–MADISON

Causal estimation (1)

- According to RRH, the problem associated with causal inference is basically a **missing data problem** – we only observed one potential outcome per individual (if we observed both we could simply calculate the causal effect), such that the counterfactual is always missing.
- The primary barrier to producing causal estimates is establishing a plausible and observable counterfactual.
- **Randomized studies** ensure that the missing **counterfactual is always missing completely at random** ensuring that the decision about which outcome will be observed ($Y_i(E)$ or $Y_i(C)$) is **determined by chance alone** – assuming randomization worked and sample size is large enough.

Causal estimation (2)

- In a perfectly estimated random experiment, we cannot estimate the causal effect for each individual, but we can calculate an **unbiased estimate of the average causal effect**: $\Delta = E[Y_i(E) - Y_i(C)]$; the population average difference between potential outcomes (i.e., the difference between sample means of the E and C groups).
- The E-C difference is an intent-to-treat effect and may be heavily influenced by take up; we therefore often also estimate treatment-on-the-treated effect
- Randomization is often impossible
- Thus, we often rely on quasi-experimental methods, such as natural experiments, for causal inference
- Focus is on isolating effect of treatment variable by establishing an appropriate counterfactual

INSTITUTE for RESEARCH on POVERTY
UNIVERSITY OF WISCONSIN–MADISON

Observational studies

- In observational (non-experimental) studies, we are worried about **confounders** or **pretreatment characteristics** that are **related to both the propensity to receive a treatment and the potential outcome(s)**. This is sometimes characterized as a social selection or endogeneity problem.
- Valid causal inference requires that the propensity to receive the treatment is **independent of the outcomes conditional on observable characteristics – in this way, we can presumably isolate the effect of the treatment**.
- Randomized studies ensure independence. Nonrandomized studies do not. In non-randomized studies, knowledge of predictors of the propensity to receive the treatment are essential for generating causal estimates. Thus, in observational studies it is the researcher's burden to show that relevant confounders have been accounted for.

Non-experimental approaches

- Things to think about:
  - What kinds of variation can you leverage? Between or within unit? Exogenous/random or only endogenous?
  - What kinds of confounders can you adjust for? Observable? Unobservable?
- Approaches:
  - Controlling for observable confounders
  - Adjusting for unobservables (often with longitudinal data or multiple observations within a unit, such as family, state, year, etc.)
  - Leveraging exogenous variation via natural experiment
  - Multiple identification strategies, sensitivity analyses/robustness checks
- See Duncan et al. (2004); Berger et al. (2009); Berger et al. (2017)—on website