Designing Monte Carlo Simulations for Power Analysis (and Other Things): A Hands-on Workshop Using R

Presenters: James E. Pustejovsky and Luke Miratrix

Course Description

In this course we will learn how to write general simulations in R. Monte Carlo simulations are an essential tool of inquiry for quantitative methodologists and students of statistics, useful both for small-scale or informal investigations and for formal methodological research. Our primary running example will be designing simulations for complex power analyses in multisite and cluster randomized trials (e.g. to handle varying cluster sizes or attrition), but our focus is more on the best practices of simulation design and how to use simulation to be a more informed and effective quantitative analyst.

Simulations are also critical for understanding the benefits and drawbacks of analytic methods. In many situations, more than one modeling approach is possible for addressing the same research question (or estimating the same target parameter). Comparing the costs of one vs. another using simulation is informative for guiding the design of analytic plans (such as plans included in pre-registered study protocols). As an example of the type of questions that researchers might encounter in designing analytic plans: what are the benefits and costs of using a model that allows for cross-site impact variation in practice?

Overall, we will show how simulation frameworks allow for rapid exploration of the impact of different design choices and data concerns, and how simulation can answer questions that are hard to answer using direct computation (e.g., with power calculators or mathematical formula). For example, available algebraic formulas are often based on asymptotic approximations, which might not “kick in” if sample sizes are moderate. This is a particular concern with hierarchical data structures that include 20-40 clusters, which is the range of common sample sizes in many large-scale randomized trials in education research.

Course content: Over four hours we will first begin by considering a set of general principles for designing simulations, dive into how to think about data generating processes as a core element of simulation, and then get a taste of writing our own simulations in R by modifying a provided set of code for doing power calculations. Time permitting, we will work together on problems or simulation design issues submitted by students prior to or during the course. The course will be hands-on, with students running and modifying code to solve exercises, so as to maximize the utility of the content.

Course materials: As part of the course, students will receive examples of scripts that can later be modified for individual purposes, along with a manuscript about simulation design that we will teach from.