Flexible matching in sample surveys containing quasiexperiments

Ben Hansen, University of Michigan

In education research, medicine, administrative studies of class equity, and other areas, one finds sample surveys aiming to compare two predesignated groups along measures to be obtained from the sample. As a quasiexperiment, the comparisons thus enabled call for statistical adjustment to account for confounding variables; but as the result of a probability sample, the data being used in these comparisons call for an analysis that respects the survey's design. These demands are often at odds. This talk discusses advantages of anticipating a binary comparison at the design stage, using frame data and a flexible matching routine (optimal full matching or matching with a variable number of controls) to create matched sets that will in turn be treated as clusters in the selection of the sample. The advantages include increased likelihood of obtaining comparable subgroups, even in a self-weighting sample; implicit adjustment for potentially identifying variables that avoids the using them directly, thus easing both IRB approval and the creation of public-use files; and, relative to sampling plans that oversample the smaller of the predesignated groups, more precise inference about population means or totals. The argument draws on the speaker's experiences with a small gender-equity study.