IAB-Colloquium
zur Arbeitsmarkt- und Berufsforschung

Generating Synthetic Datasets for Small Geographic Areas in the American Community Survey

Joseph Sakshaug

University of Michigan, Institute for Social Research, Ann Harbour

Sample surveys provide a critical source of information about the well-being of populations around the world. Increasingly, this information is being demanded in the form of small area statistics to support policy decisions at the local level. However, data disseminators are prevented from releasing small area identifiers in public-use micro-level datasets due to disclosure risks. This dissertation research tests a new method for creating public-use datasets that enhances data confidentiality and permits the release of geographical details for small area estimation. The proposed method, based on methodology described in Rubin (1993), replaces observed data values with multiply-imputed, or synthetic, data values. A Bayesian hierarchical model, fitted using the observed data, is used to simulate the synthetic values while accounting for each level of geography (e.g., counties, states). Public-use data from the 2005-2007 American Community Survey, a nationwide survey that collects information on population and housing characteristics, is used to demonstrate the methodology. Comparisons between synthetic and actual data inferences will be presented for several descriptive and analytic statistics.

Donnerstag, 15. Juli 2010
11:00 Uhr
Sitzungssaal 126a