Michigan Program in Survey Methodology

Dissertation Defense

Qi Dong

Monday, April 23, 2012

2:00pm-4:00pm

368 ISR and JPSM Conference Room

Combining Information from Multiple Complex Surveys

Increasingly many substantive research questions require a degree of information not adequately collected in a single survey. Fortunately, survey organizations often repeatedly draw samples from the same population for different surveys and collect data on a considerable number of overlapping variables. This dissertation presents a new method for combining multiple surveys from a missing data perspective. Two major improvements of the new method include: 1) it adjusts for the incomparability among different data sources especially the complex sampling design features and 2) it can combine any number of surveys.

The basic proposal is to simulate synthetic populations from which the respondents of each survey have been selected. In this process, different sampling designs of the multiple surveys will be taken into account. Once we have the synthetic populations, we could treat them as simple random samples with no complex sampling design features and borrow information across surveys to adjust for nonsampling errors or fill in the variables that are lacking in one or more surveys. Then, we can analyze each synthetic population with standard complete-data software for simple random samples and obtain valid inference by combining the point and variance estimates first across synthetic populations within each survey using the existing combining rules for synthetic data and then across multiple surveys using the methods developed in this dissertation.

The first study develops the combining rule when multiple surveys present and proposes a model-based method to impute the unobserved population. The 2006 National Health Interview Survey (NHIS) and Medical Expenditure Panel Study (MEPS) are combined to estimate health insurance coverage. The second study develops a nonparametric method to impute the unobserved population, which is used to generate synthetic populations for the 2006 NHIS and MEPS and produce combined estimates of health insurance coverage. The third study extends the new method to combine surveys with missing variables. A new two-stage combining rule is developed to account for the uncertainty due to simultaneously imputing the missing variables and generating synthetic populations. The 2006 Behavioral Risk Factor Surveillance System (BRFSS) is combined with the NHIS and MEPS to estimate health insurance coverage.

Associate Professor Michael R. Elliott, Co-chair and Professor Trivellore E. Raghunathan, Co-chair