STATISTICAL DATA INTEGRATION

Prof. Dr. Partha Lahiri
Joint Program in Survey Methodology and Department of Mathematics
University of Maryland, College Park, USA

E-mail: plahiri@umd.edu

There is a growing demand to produce reliable estimates of various socio-economic and health characteristics at both national and sub-national levels. To avoid high costs of collecting new data in such cases, there is a growing need to combine different existing structured and unstructured databases such as surveys, administrative/register records, social media data, mobile phone data, sensor data, satellite date, etc. We shall discuss various issues and classical and Bayesian methods in statistical data integration. In particular, we shall cover various methods that use a few common variables in combining multiple data sources with no or negligible overlapping units. In another important data integration situation, we have a complete or significant overlap in units from different data sources and the common variables in different datasets are often misreported. The goal of this pre-conference workshop is to provide a broad overview of the subject and will avoid derivations of complex results.

Proposed course length: Two hours

Course Text and Materials: The course will be based on the presenter's lecture notes.

Target Audience and Prerequisites: The course is intended for researchers with a Master's degree in statistics.