Prof. Jerome Reiter

Institute of Statistics and Decision Sciences, Duke University, North Carolina

Multiple Imputation for Disclosure Limitation: Why Rubin's 1987 Rules Are Not Appropriate

Abstract:

Multiple imputation has been proposed as a technique for limiting the risks of disclosures in public use files. Agencies can release datasets with fully synthetic data, i.e. all released values are simulated from statistical models. Or, they can release the original records from the file with only some values replaced with multiple imputations, i.e. partially synthetic. Researchers at IAB have suggested implementing fully or partially synthetic data in two stages, as shall be discussed in the talk. Surprisingly, even though all of these methods rely on the concepts of multiple imputation, they re-quire different methods for combining the point and variance estimates from the multiple copies than those developed for the missing data context by Rubin (1987). In this talk, I review how multiple imputation can be used for disclosure limitation and explain why Rubin's (1987) combining rules are not appropriate in the disclosure limitation context.