

Klairung Samart, PhD student University of Wollongong

Title: Analysis of Probabilistically Linked Data

Abstract: Probabilistic matching of records from different data sets is often used to create linked data sets for use in research in health, epidemiology, economics, demography and sociology. Clearly, this type of matching can lead to linkage errors, which in turn can lead to bias and increased variability when standard statistical estimation techniques are used with the linked data. In this research we develop unbiased regression parameter estimates when fitting a linear mixed model to probabilistically linked data. Furthermore, since estimation of variance components is also an important objective when fitting a mixed model, we develop appropriate modifications to standard methods of variance components estimation in order to account for linkage error. In particular, we focus on three widely used methods of variance components estimation: analysis of variance, maximum likelihood and restricted maximum likelihood. Simulation results show that our estimators performed reasonably well compared to the naive weighted least square estimator that just uses the linked data.