Dr James Chipperfield

Australian Bureau of Statistics (ABS)

Using the Bootstrap to make Inference with Estimating Equations and Probabilistically-Linked Data

Record linkage is the act of bringing together records that are believed to belong to the same unit (e.g. person or business) from two or more files. It is a very common way to enhance dimensions such as time and breadth or depth of detail. Record linkage is often not an error-free process and can lead to linking a pair of records that do not belong to the same unit. This is because linking fields on the files, which ideally would uniquely identify each unit, are often imperfect. For example, linking fields may legitimately change over time, or may contain errors. There is an explosion of record linkage applications, particularly involving government agencies and in the health area, yet there has been little work on making correct inference using such linked files. Naively treating such a linked file as if it were linked without errors can lead to biased inferences. This paper develops a method of making inference with estimating equations when record linkage is not an error-free process. In particular, this paper develops a parametric bootstrap approach to estimation which can accommodate sophisticated probabilistic record linkage techniques that are widely used in practice. The paper demonstrates the effectiveness of the method in a simulation and in a real application.

Bio:

James has worked in the Methodology Division of the Australian Bureau of Statistics since 1997 as a statistician. During this time he has worked on a range of statistical problems facing the ABS including: data access and confidentiality, split questionnaire designs, and analysis of linked data. He is currently Assistant Director in the Data Access and Confidentiality Unit and is Principal Research Fellow at NIASRA.