

Disclosure-Protected Inference with Linked Micro-data using a Remote Analysis Server

Large amounts of micro-data are collected by data custodians in the form of Censuses and administrative sources. Often, data custodians will collect different information on the same individual. Many important questions can be answered by linking micro-data collected by different data custodians. For this reason, there is very strong demand from analysts, within government, business and universities, for linked micro-data. However, many data custodians are legally obliged to ensure the risk of disclosing information about a person or organisation is acceptably low. Different authors have considered the problem of how to facilitate reliable statistical inference from analysis of linked micro-data while ensuring that the risk of disclosure is acceptably low. This paper considers the problem from the perspective of an Integrating Authority that is trusted to link the micro-data and to facilitate analysts' access to the linked micro-data via a remote server, which allows analysts to fit models and view the statistical output without being able to observe the underlying linked micro-data itself. One disclosure risk that must be managed by an Integrating Authority is that one data custodian may use the micro-data it supplied to the Integrating Authority and statistical output released from the remote server to disclose information about a person or organisation that was supplied by the other data custodian. This paper considers analysis of only binary variables. The utility and disclosure risk of the proposed method are investigated in a simulation and with a real example.

Dr. James Chipperfield

Assistant Director, Australian Bureau of Statistics

Senior Research Fellow, National Institute for Applied Statistics Research Australia

University of Wollongong