Adjusting for linkage errors to analyse coverage of the Integrated Data Infrastructure (IDI) and the administrative population (IDI-ERP)

Hochang Choi
Statistical Analyst
Stats NZ
Christchurch, New Zealand

Abstract

The Integrated Data Infrastructure (IDI) is a linked data environment combining a variety of administrative and survey datasets. It is anchored by a spine which defines an "ever-resident" population constructed from the union of births, tax and visa data. Stats NZ has constructed an experimental New Zealand resident population (the IDI-ERP) from the IDI by defining rules to identify people who are likely to be usual residents and exclude people likely to no longer be usual residents in New Zealand at some reference date. The IDI-ERP may be of interest as a reference population for small area population estimation, and other social and population research. It is therefore of interest to analyse the coverage, or representativeness, of the IDI-ERP with respect to the true usual resident population of New Zealand. Linkage of the 2013 Census data to the IDI spine provides an opportunity for detailed analysis of the population coverage patterns of the IDI-ERP. While straightforward in principle, this analysis is complicated by error in the linkage of the census to the IDI spine and the fact the census itself is subject to under-coverage with respect to the true usual resident population. In this paper, we outline our approach for adjusting IDI-ERP estimates for these sources of error and present preliminary estimates of the population coverage of the IDI-ERP, by sex, age and ethnicities.