Laura O'Sullivan

Abstract

This paper examines why different linking approaches are used depending on the data being linked, the reason it is being linked, and the country's rules and regulations. We compare our linking approach in the Integrated Data Infrastructure (IDI) with other linking at Statistics New Zealand and at the Australian Bureau of Statistics. We describe the new methods we have developed to select cut-offs, decide when to do a clerical review, and determine the quality of the links in the IDI. We explain how, during our development, dividing the links into near-exact and non-exact links has helped us to select cut-offs. Human intervention is an important part of our process, although we minimise the amount required where possible. We examined run-times, and found the biggest factor affecting these was block sizes.