

CENSUS-LIKE POPULATION SIZE ESTIMATION BASED ON ADMINISTRATIVE DATA

Li-Chun Zhang
University of Southampton, UK, L.Zhang@soton.ac.uk
Statistics Norway, Norway, lcz@ssb.no

John Dunne
Central Statistics Office, Ireland, John.Dunne@cso.ie

Register data that originate from administrative sources are increasingly being explored to generate statistical outputs directly. Using register data to produce census-like population size estimates is an on-going development at the Central Statistics Office, Ireland. It aims to reduce the cost associated with the traditional approach based on census and coverage surveys, making it feasible to update the population estimates on a yearly basis. The methodology can potentially provide a viable approach and a pioneering example for provision of population statistics, in a setting where one neither has a central population register nor deploys any additional coverage surveys.

Traditionally capture-recapture models have been used to deal with multiple list enumerations subjected to under-enumeration errors. A number of assumptions need to be checked when applying the method to the available register data, including independence between the lists, homogeneous capture probability, etc. In particular, compared to census and coverage surveys, the registers may contain non-negligible erroneous enumerations, or over-coverage error, which is a challenge to the existing capture-recapture estimation methods.

In this talk we discuss two recent developments in handling capture-recapture data with additional over-coverage errors. First, several alternative modelling assumptions regarding the erroneous enumerations are considered in details for the setting with two (or more) list enumerations that may suffer from erroneous enumerations and an additional list with only under-coverage error. This helps to determine, in a given data setting, which assumption is most useful. Next, we consider a trimmed dual-system estimation (TDSE) approach, which consists of scoring the potential erroneous records in the two lists on which the standard DSE is based, and apply the DSE after removing the flagged records. We discuss how stopping rules of the trimming can be constructed theoretically, and how the TDSE can be combined with the modelling approach, in order to verify the extent of the remaining erroneous enumerations and, if necessary, to accommodate them in the estimation.