

Robustness of population size estimates against violation of the independence assumption

Susanna. C. Gerritse*
University Utrecht, Utrecht, Netherlands
s.c.gerritse@uu.nl

Peter. G. M. van der Heijden
University Utrecht, Utrecht, Netherlands/University Southampton, Southampton, England

Bart. F. M. Bakker
Statistics Netherlands, The Hague, Netherlands/VU University, Amsterdam, Netherlands

An important quality aspect of censuses is the degree of coverage of the population. When administrative registers are used, undercoverage can be determined by linking two or more registers and estimate the number of individuals that is missed by each of the registers. Adding this estimated number to the observed number of individuals yields a population size estimate. This is also known as the capture-recapture methodology. The standard approach is to use loglinear models that most often rely on an independence assumption, i.e. that being in the first register is statistically independent from being in the second register. If covariates are available, the independence assumption can be replaced with a less strict independence assumption conditioned on covariates. This assumption is less stringent because it can take into account heterogeneous inclusion probabilities over the levels of the covariate.

However, this independence assumption is rarely met. In this paper we will study the violation of the independence assumption. In particular, we will investigate the robustness of the population size estimate under dependence, both with and without conditioning on covariates, using real data. Thus the sensitivity analyses investigate the robustness of the capture recapture methods under violation of (conditional) independence. The results show that violation of the independence assumption could lead to seriously biased estimates. To exemplify, we used data considering people with a Polish, and people with an Afghan, Iraqi and Iranian nationality residing in the Netherlands. For these nationalities we find that, respectively, under dependence the population size estimate can either range from an underestimation of 15 % to an overestimation of 10%, or from an underestimation of 44% to an overestimation of 64%.

Key words: Capture – recapture method, population size estimate, sensitivity analysis, census