Multilevel modeling of zero-inflated count data from complex surveys using pseudo-likelihood and Bayesian methods

Abstract

Multilevel complex survey data are obtained from study designs that involve multiple stages of sampling where sampling units are drawn at each stage. The features of such survey design often include clustering, stratification, multilevel sample selection, and unequal probability of selection of observations. Typically, specific methods that account for these features are needed to estimate and make inference on parameters of interest. For example, multilevel models that account for sampling weights have become popular for the analysis of such type of data. Recently, multilevel pseudo-likelihood (MPL) methods with scaled weights are gaining popularity for the analysis of Gaussian and Binomial data. However, there are no studies that assess the performance of pseudo-likelihood and scaling methods on models for count data that are characterized by point mass at zero. The literature on Bayesian modeling of count data from complex surveys is also limited. There are limited studies that assessed the performance of MPL estimations when only overall weights at the lower unit level are given in the survey data. However, sampling weights could be available at all or some levels of the sampling design. Thus, we propose to develop and assess the performance of MPL and Bayesian methods for the analysis of count data from complex surveys under several scenarios of sampling weights. Another common issue that arises with complex surveys is the aggregation of outcomes and covariates from lower level to higher level. But, there are no studies that are developed for dealing with how to address the aggregation of sampling weights which is a subject of interest in this proposal.
This work accomplished three aims: i) we developed a multilevel pseudo maximum likelihood estimate for count data from multilevel complex surveys and assessed its performance under several weight scaling approaches. ii) we developed a Bayesian approach for the analysis of count outcomes from complex surveys comparing different weight approaches, iii) we developed and assessed an aggregate data model for complex survey data, which allows for multilevel weight among disease rates across clusters. We apply the proposed analysis strategies of three aims to the real survey data, the multi-country data from DHS (Demographic health survey) to demonstrate the methods.