

Distribution-free Model Selection for Longitudinal Zero Inflated Count Data with Missing Responses and Covariates

沈仲維

國立中正大學數學系

Abstract

In many medical and social science studies, count responses with excess zeros are very common and often the primary outcome of interest. Such count responses are usually generated under some clustered correlation structures due to longitudinal observations of subjects. To model such longitudinal count data with excess zeros, the zero-inflated binomial (ZIB) models for bounded outcomes, and the zero-inflated negative binomial (ZINB) and zero-inflated poisson (ZIP) models for unbounded outcomes all are popular methods. To alleviate the effects of deviations from model assumptions, a semiparametric (or, distribution-free) weighted generalized estimating equations has been proposed to estimate model parameters when data are subject to missingness. In this article, we further explore important covariates for the response variable. Without assumptions on the data distribution, a model selection criterion based on the expected weighted quadratic loss is proposed to select an appropriate subset of covariates, especially when count responses have excess zeros and data are subject to nonmonotone missingness in both responses and covariates. To understand the selection effects of the percentages of excess zeros and missingness, we design various scenarios for covariate selection in the mean model via simulation studies and a real data example regarding the study of cardiovascular disease is also presented for illustration.

關鍵詞：generalized estimating equations, missing at random, two-component mixture models, variable selection, zero-inflation