THE IMPORTANCE OF CLASS PREDICTION IN ZERO-INFLATED MODELS

M.S. Gilthorpe[1†], Y. Cheng[1], M. Frydenberg[2] and V. Baelum[2]

[1]*University of Leeds, UK*; [2]*University of Aarhus, Denmark.*

[†] E-mail: *m.s.gilthorpe@leeds.ac.uk*

In a variety of research domains, data are generated as a consequence of the count process and may possess an 'excess' of zeros. There have been many attempts to analyse such data using different statistical methods, including the zero-inflated Poisson (ZiP) and zero-inflated binomial (ZiB) models. The interpretation of these models is however problematic if the covariates considered for the non-zero distribution part of the model are omitted in the conditional part as class predictors. Although the practice of including covariates as both outcome predictors and class predictors is suggested as an option in the literature (not as a necessity), the argument for this practice to become the norm needs to be made more firmly.

This argument is illustrated by a hypothetical example of a ZiP model using simulated data based on a genuine clinical dataset in the public domain (http://www.blackwellpublishers.co.uk/rss/). These dental epidemiological data were used by Böhning *et al.* to demonstrate the utility of the ZiP model. We additionally explore both ZiP and ZiB models where class membership is predicted by the covariates, as originally suggested by Böhning *et al.* for the ZiP model, though not actually reported. We flag the necessity of using covariates from the non-zero distribution part of the model to predict class membership, to make meaningful interpretation of model coefficients. As previously demonstrated by Skrondal and Rabe-Hesketh, we show how the use of a Poisson distribution in this instance is questionable, since the counts represent the number of 'successes' out of a total of eight 'trials', thereby unsurprisingly yielding better models using the binomial distribution, since Poisson model predicts inappropriate long tails.

Finally, we propose the use of a visual method to aid the interpretation of model coefficients for all covariates that are simultaneously used as outcome predictors and conditionally as class predictors. This provides additional insight into the role of covariates in ZiP and ZiB models.

References

D. Böhning, E. Dietz, and P. Schlattmann. The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *Journal of the Royal Statistical Society, Series A* 162:195-209, 1999.

A. Skrondal and S. Rabe-Hesketh. Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models, London:Chapman & Hall, 2004.