

Analysis of Overdispersed Antenatal Health Care Count Data

Zakir Hossain¹ and Enamul Kabir²

¹Department of Statistics, University of Dhaka, Bangladesh

E-mail: zakir.hossain@du.ac.bd

²School of Agricultural, Computational and Environmental Sciences

University of Southern Queensland, Toowoomba, QLD, Australia

E-mail: Enamul.Kabir@usq.edu.au

Abstract

Overdispersion (or greater variability) in count data analysis is very common in many practical fields of health sciences. Ignorance of the presence of overdispersion by the researchers in such data analysis may cause misleading inferences and thus lead to incorrect interpretations of the results. Researchers should account for the consequences of overdispersion and need to select the correct choice of models for the analysis of such data. In this paper, Generalized Linear Models (GLMs) are applied in modelling and analysis of antenatal care (ANC) count data extracted from the Bangladesh Demographic and Health Survey (BDHS) 2014. Pearson chi-square and different score tests are used to investigate the effect of overdispersion in the analysis. Overdispersion is found to be significant in the antenatal health care count data and so appropriate modelling is used to produce valid inferences for the regression parameters. The zero-truncated negative binomial regression (0-NBR) is found to be the best choice for analysing such data while excluding zero counts. Study findings reveal that place of residence, order of birth, exposure to mass media, wealth index and education of mother have significant impacts on the ANC status of women during pregnancy in Bangladesh.

Keywords: Overdispersion, Pearson chi-square, Antenatal care, Zero-truncated negative binomial regression and Incidence rate ratio.