Generalized Estimating Equations (GEE) Approach for Clustered Binary Data with Application to COVID-19 Treatment.

Sadixa Sanjel

MS in Data Science, 2022

Minnesota State University, Mankato

Keywords: clusters, binary, correlation, logistic regression, GEE method, generalized linear models

Clustered binary data frequently occur in epidemiology and other applied fields such as clinical trial studies, where observations within the respective samples are correlated. In such situations, the standard logistic regression method is not valid as logistic regression requires the observations to be independent of one other. This situation arises when treating COVD-19 patients. Patients from certain clusters, such as geographic areas or the same family, are highly correlated, and we need to fit the model using the GEE approach. In this paper, Standard Logistic Regression (LS), Generalized Linear Models (GENMOD), and GEE procedures have been utilized for comparison purposes. Simulated data have been used for case study analysis. The GEE model's accuracy rate is superior when data are binary clustered.

For more information about the GEE and GENMOD methods, see Fitzmaurice, Laird, and Ware (2011); Hardin and Hilbe (2003); Diggle et al. (2002); Lipsitz et al. (1994).

References:

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). Applied Longitudinal Analysis. Hoboken, NJ: John Wiley & Sons.

Hardin, J. W., and Hilbe, J. M. (2003). Generalized Estimating Equations. Boca Raton, FL: Chapman & Hall/CRC.

Lipsitz, S. R., Fitzmaurice, G. M., Orav, E. J., and Laird, N. M. (1994). "Performance of Generalized Estimating Equations in Practical Situations." Biometrics 50:270–278.

Lipsitz, S. R., Kim, K., and Zhao, L. (1994). "Analysis of Repeated Categorical Data Using Generalized Estimating Equations." Statistics in Medicine 13:1149–1163.