LITERATURE REVIEW: Generalizing Estimating Equations

STA 6257 - Group Members: Ashley Dosch, Amy Muller and Anthony Stawiery
July 2023

A review of the literature confirms the application of GEE to longitudinal data sets (Burton & Sly 1998, Hanley et al. 2003, Liang & Zeger 1986, Gayen & Kumar 2018, Shao et al. 2023, Wang et al. 2012, Salazar et al. 2016, Ito & Sugasawa 2022). The literature also suggests GEE improves accuracy of estimating functional dependence and cluster analysis (D'Angelo, et al. 2011). GEE methods are also less susceptible to correlation structure (Koper & Manseau 2009), and can be extended to higher dimensionality to improve correlation structure analysis in the presence of covariances (Crespi et al. 2009).

Burton and Sly (1998) contend that repeated measurements over a period of time contain a correlation between observations. Standard t-testing, linear regression, and chi-squared analyses (to name a few) fail to address these correlations because the methods assume that repeated measurements are independent from one another. Hanley et al. (2003) attempt to use GEE with weighted combinations of data to extract important correlation information. Hanley et al. found GEE is able to produce accurate standard errors. Further, the use of GEE reduces computation complexity by focusing on the largest cluster, rather than the number of clusters. Shao et al. (2023) conduct an experiment to compare different methods, including GEE, on a longitudinal data set. The authors confirmed that GEEs are successful at determining direct relationships between predictors and observations. However, there is no indication of which method is better since it greatly depends on the needs of the resulting analysis.

D'Angelo et al. (2011) attempt a two-stage GEE approach to longitudinal data: GEE Marginal and GEE Transition models. The marginal approach estimates regression and correlation as separate pieces. The transition model uses previous models to estimate the expectation of a current model. The authors found that a transition model resulted in less connectivity differences and also produced more desirable statistical properties. The authors conclude that GEE is beneficial because it does not require knowledge of the distribution of the variables, however, the method does need a correlation structure to be specified.

Koper and Manseau (2009) compare generalized mixed-effects models and GEEs in an attempt to accurately model animal habitat selection utilizing resource selection functions. The authors confirm what has been found elsewhere in the literature that GEEs reduce the complexity of data analysis. The authors concluded that both methods were able to adequately develop resource selection functions, but GEEs were less susceptible to correlation structure and more likely to show convergence between correlated data. Generalized mixed-effects did prove to be more useful with large differences in variances. Crespi et al. (2009) explored using second-order GEEs because the authors posit that one-dimensional GEEs are inefficient when identified correlation structures contain errors and that second-order GEEs are more efficient at identifying correlation structures when covariances are present. However, if the number of randomized clusters are small, second-order GEEs become less effective since some predictors can be discarded as inconsequential.