

# Small Sample Characteristics of Generalized Estimating Equations for Categorical Repeated Measurements

Donguk Kim

*Department of Statistics, Sungkyunkwan University*

*53, Myungnyun-dong 3-ga, Chongno-gu*

*Seoul 110-745, Korea*

*dkim@skku.ac.kr*

Jaejik Kim

*Department of Statistics, Sungkyunkwan University*

*53, Myungnyun-dong 3-ga, Chongno-gu*

*Seoul 110-745, Korea*

*Keywords:* Generalized estimating equations; Repeated measurements; Ordinal data; Small sample size.

## 1. Introduction

Longitudinal studies, where the individual's responses are measured over the time, are frequently used. Since multiple measurements are obtained on a single subject, correlation among the observations should be considered. Generalized estimating equations(GEE) were proposed by Liang and Zeger(1986) for analyzing repeated responses, and Prentice(1988) extended their estimating equations. The estimating equations can be thought as extension of quasi-likelihood approach to correlated data, which was first proposed by Wedderburn(1974).

For the simulation studies to study GEE methods, Lipsitz, Laird and Harrington(1991) proposed using the odds ratio as the measure of association between repeated measures and simulated binary data. Lipsitz, Fitzmaurice, Orav and Laird(1994) and Gunsolley, Getchell and Chinchilli(1995) simulated binary data. Furthermore, the GEE can be extended for use with repeated multinomial responses. Miller, Davis and Landis(1993) extended GEE to polytomous responses, and Lipsitz, Kim and Zhao(1994) also extended GEE model to models for the repeated categorical responses. Stiger, Kosinski, Barnhart and Kleinbaum(1998) generated repeated ordinal data by using multivariate normal distribution and compared performance of ANOVA, MANOVA, WLS and GEE.

For models for repeated categorical data, GEE estimator has asymptotic multivariate normal distribution in large sample size. But GEE is based on large sample asymptotic theory. In this paper, we investigate the performance of GEE estimators for repeated ordinal data, especially in small sample size, and also are interested in the performance of variance for the GEE estimators. We generate ordinal repeated measurements for two groups using two methods. By using the method of Gange(1995), we can generate random correlated ordinal data from the known covariance of repeated measurements. Through Monte Carlo simulation studies we investigate the empirical type 1 error rates, powers, relative efficiencies of the GEE estimators for polytomous ordinal response variables, especially in small sample size. We also study the effect of unequal sample size for two groups and the behavior of variance estimators.

## ACKNOWLEDGEMENT

This Paper was Supported by 63 Research Fund, Sungkyunkwan University, 1999.

## REFERENCE

- [1] Gange, S.J. (1995), "Generating Multivariate Categorical Variates Using the Iterative Proportional Fitting Algorithm," American Statistical Association, Vol 49, No 2, 134-138.
- [2] Gunsolley, J.C. Getchell, C. and Chinchilli, V.M. (1995), "Small Sample Characteristics of Generalized Estimating Equations," Communications in Statistics, Part B - Simulation and Computation, Vol 24, 869-878.
- [3] Liang, K. and Zeger, S.L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," Biometrika, Vol 73, No 1, 13-22.
- [4] Lipsitz, S.R. Laird, N.M. and Harrington, D.P. (1991), "Generalized Estimating Equations for Correlated Binary Data : Using the Odds Ratio as a Measure of Association," Biometrika, Vol 78, No 1, 153-160.
- [5] Lipsitz, S.R. Kim, K. and Zhao, L. (1994), "Analysis of Repeated Categorical Data Using Generalized Estimating Equations," Statistics in Medicine, Vol 13, 1149-1163.
- [6] Lipsitz, S.R. Fitzmaurice, G.M. Orav, E.J. and Laird, N.M. (1994), "Performance of Generalized Estimating Equations in Practical Situations," Biometrics, Vol 50, 270-278.
- [7] Miller, M.E. Davis, C.S. and Landis, J.R. (1993), "The Analysis of Longitudinal Polytomous Data: Generalized Estimating Equations and Connections with Weighted Least Squares," Biometrics, Vol 49, 1033-1044.
- [8] Prentice, R.L. (1988). "Correlated Binary Regression with Covariates Specific to Each Binary Observation," Biometrics, Vol 44, 1033-1048.
- [9] Stiger, T.R. Kosinski, A.S. Barnhart, H.X. and Kleinbaum, D.G. (1998), "ANOVA for Repeated Ordinal Data with Small Sample Size? A Comparison of ANOVA, MANOVA, WLS and GEE Methods by Simulation," Communications in Statistics, Part B - Simulation and Computation, Vol 27, 357-375.
- [10] Wedderburn, R.W.M. (1974), "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," Biometrika, Vol 61, No 3, 439-447.
- [11] Zeger, S.L. and Liang, K. (1986), "Longitudinal Data Analysis for Discrete and Continuous Outcomes," Biometrics, Vol 42, 121-130.

## RESUME

Liang et Zeger ont proposé les équations d'estimation généralisées (GEE) pour analyser des données répétées qui sont séparées ou continues. Le modèle GEE peut être étendu au modèle de données catégorielles répétées et le système d'estimation GEE a une *distribution normale asymptotique multivariées* en cas d'échantillons à grande échelle. Mais le GEE est basé sur la théorie asymptotique d'échantillons à grande échelle. Dans ce document, nous étudions les propriétés du système d'estimation GEE pour des données ordinales répétées en cas d'échantillons à petite échelle. Nous créons des mesures ordinales répétées pour deux groupes utilisant deux méthodes. A travers les études de simulation Monte Carlo, nous étudions le type 1 empirique de taux d'erreur, les puissances, l'efficacité relative des systèmes d'évaluation GEE et le comportement des systèmes d'estimation variable pour les réponses ordinales polytomes variables, surtout en cas d'échantillons à petite échelle.