Comparing estimation methods for categorical marginal models

Kuijpers, R.E.; Bergsma, W.P.; van der Ark, L.A.; Croon, M.A.

DOI
10.1007/978-3-319-07503-7_23
Publication date
2015
Document Version
Final published version
Published in
Quantitative Psychology Research

Citation for published version (APA):
Kuijpers, R. E., Bergsma, W. P., van der Ark, L. A., & Croon, M. A. (2015). Comparing estimation methods for categorical marginal models. In R. E. Millsap, D. M. Bolt, L. A. van der Ark, & W-C. Wang (Eds.), *Quantitative Psychology Research: The 78th Annual Meeting of the Psychometric Society* (pp. 359-375). (Springer Proceedings in Mathematics & Statistics; Vol. 89). Springer. https://doi.org/10.1007/978-3-319-07503-7_23

Chapter 23
Comparing Estimation Methods for Categorical Marginal Models

Renske E. Kuijpers, Wicher P. Bergsma, L. Andries van der Ark, and Marcel A. Croon

Abstract Categorical marginal models are flexible models for modelling dependent or clustered categorical data which do not involve any specific assumptions about the nature of the dependencies. Categorical marginal models are used for different purposes, including hypothesis testing, assessing model fit, and regression problems. Two different estimation methods are used to estimate marginal models: maximum likelihood (ML) and generalized estimating equations (GEE). We explored three different cases to find out to what extent the two types of estimation methods are appropriate for investigating different types of research questions. The results suggest that ML may be preferred for assessing model fit because GEE has limited fit indices, whereas both methods can be used to assess the effect of independent factors in regression. Moreover, ML is asymptotically efficient, while GEE loses efficiency when the working correlation matrix is not correctly specified. However, for parameter estimation in regression GEE is easier to apply from a computational perspective.

23.1 Introduction

In the social and behavioral sciences, researchers frequently collect data that are correlated or dependent, such as longitudinal data, dyadic data, and data obtained from psychological or educational testing in which each respondent answers several

R.E. Kuijpers (✉)
Department of Methodology and Statistics, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: r.e.kuijpers@tilburguniversity.edu

W.P. Bergsma
London School of Economics, London, UK

L.A. van der Ark
University of Amsterdam, Amsterdam, The Netherlands

M.A. Croon
Tilburg University, Tilburg, The Netherlands

R.E. Millsap et al. (eds.), *Quantitative Psychology Research*, Springer Proceedings in Mathematics & Statistics 89, DOI 10.1007/978-3-319-07503-7_23

items. Although the dependencies are not always of main interest for the research, they cannot be ignored. Ignoring the dependencies in the analysis may produce incorrect standard errors and p-values. Categorical marginal models (Bergsma et al. 2009) are flexible models for categorical data that take these dependencies into account without making assumptions about their nature. These models are useful when researchers investigate research questions concerning the marginal distributions of a set of variables instead of testing hypotheses with respect to the joint distribution for all variables in a certain data set.

Categorical marginal models are used to answer various types of research questions. Two types of research questions we encountered in the literature are research questions that involve hypothesis testing and research questions that involve parameter estimation. An example of a research question that involves hypothesis testing is provided by Kuijpers et al. (2013a). They proposed fitting categorical marginal models to test the hypothesis that Cronbach's alpha is equal for two or more subgroups. Other examples include testing marginal models for scalability coefficients (Van der Ark et al. 2008; Kuijpers et al. 2013b), marginal homogeneity (Bergsma et al. 2009), and ordinal association measures (e.g., Lang 2004).

For the second type of research question, the main interest lies in the values of the estimated regression parameters. For example, Molenberghs and Verbeke (2005) used marginal models to investigate the effect of two types of vaccinations from two different companies on the presence/absence of headaches and respiratory problems in two trial periods. Other examples include (1) modelling the effect of different demographic variables on the relation between smoking and drinking behavior in different subgroups of the Belgian Interuniversity Research on Nutrition and Health study (Kesteloot et al. 1989) and (2) investigating whether different (combinations of) variables such as gender, age, education, and religiosity have a significant effect on the attitude towards women's roles (Bergsma et al. 2009, pp. 168–171).

Both likelihood methods and quasi-likelihood methods have been used to estimate marginal models. For likelihood methods, which include maximum likelihood (ML) estimation (Bergsma 1997), maximum empirical likelihood (MEL) estimation, and maximum augmented empirical likelihood (MAEL) estimation (Van der Ark et al. 2013), the full likelihood is optimized under the marginal model of interest and under the assumption that the data follow a multinomial distribution. ML, MEL, and MAEL estimation differ with respect to whether or not they use all possible item-score patterns of a set of items for the estimation of a model. For research questions that concern hypothesis testing, the authors have used ML (e.g., Kuijpers et al. 2013a,b; Van der Ark et al. 2008). For this paper, we only consider ML estimation. The most popular quasi-likelihood method is generalized estimating equations (GEE; Liang and Zeger 1986). GEE is not based on a specific probability model for the data. The estimation method assumes only a mean-variance relationship for the dependent variable. GEE is mainly used for estimating regression models (e.g., Agresti 2013; Molenberghs and Verbeke 2005; Pawitan 2001). Skrondal and Rabe-Hesketh (2004, p. 200) noted that GEE has some limitations with respect to hypothesis testing and assessing model adequacy.

In this study, we explored to what extent ML estimation and GEE are appropriate for investigating the three types of research questions. We considered three different research questions, referred to as Case 1, Case 2, and Case 3. Let θ denote a particular coefficient, and let c denote a fixed value. In this study θ can refer to either the mean (μ) or the reliability coefficient Cronbach's alpha (α). In Case 1, we investigated whether θ is equal to a fixed value c (i.e., $\theta = c$); in Case 2, we investigated whether θ is equal for two groups (i.e., $\theta_1 = \theta_2$); and in Case 3, we investigated whether θ is a linear function of independent variable X (i.e., $\theta = \beta_0 + \beta_1 X$). In each case, we investigated the two coefficients μ and α, and we compared the results obtained with ML estimation and GEE. We illustrated each case with a real-data example.

The remainder of this paper is organized as follows. First, we briefly explain categorical marginal models. Second, we discuss the two groups of estimation methods. Third, we discuss how to express μ and α in an appropriate notation for ML estimation. Fourth, using a real-data set, we compare the estimation methods for the three cases. Finally, we discuss the outcomes and provide recommendations for future research.

23.2 Categorical Marginal Models

In order to use categorical marginal models for testing hypotheses for a coefficient or for estimating parameters in a regression model, the first step is to write the coefficient or the regression model as a function of the frequencies of the item-score patterns that are observed in the data. Consider a set of J items, each item having $z+1$ ordered answer categories $(0, 1, \ldots, z)$; this produces $L = (z+1)^J$ possible item-score patterns. Let $\mathbf{n}$ be an $L \times 1$ vector containing the observed frequencies of the L possible item-score patterns. For example, a dichotomously scored test consisting of J = 3 items (denoted by a, b, and c) has $L = 2^3 = 8$ possible item-score patterns; hence, vector $\mathbf{n}$ equals

$$\mathbf{n} = \begin{pmatrix} n_{abc}^{000} \\ n_{abc}^{001} \\ n_{abc}^{010} \\ n_{abc}^{011} \\ n_{abc}^{100} \\ n_{abc}^{101} \\ n_{abc}^{110} \\ n_{abc}^{111} \end{pmatrix}, \tag{23.1}$$

where the subscripts denote the items and the superscripts the item scores. The observed frequencies of the item-score patterns in vector $\mathbf{n}$ are given in lexicographic order, running from $00\ldots0$ to $zz\ldots z$ with the last digit changing fastest and the digit in the first column changing slowest.

The expected frequencies under a categorical marginal model are collected in an $L\times 1$ vector $\mathbf{m}$. Because there may be more than one set of expected frequencies that satisfy a marginal model, $\mathbf{m}$ is as close as possible to $\mathbf{n}$. Let matrix $\mathbf{C}$ be a *marginal matrix* consisting of zeros and ones, such that $\mathbf{C}'\mathbf{m}$ produces the relevant marginals from the contingency table. Vector $\boldsymbol{\beta}$ contains the K model parameters β_k $(k=0,1,\ldots,K-1)$. Then, let $\mathbf{Z}$ be the design matrix of the marginal model that uses effect coding in order to select the right parameters from vector $\boldsymbol{\beta}$. In a categorical marginal model, a function of the relevant marginals is then written as

$$\mathbf{f}(\mathbf{C}'\mathbf{m})=\mathbf{Z}\boldsymbol{\beta}, \tag{23.2}$$

where $\mathbf{f}$ is an appropriate vector function. Alternatively, the model can be written without parameter vector $\boldsymbol{\beta}$ (Agresti 2013, pp. 460–461; Aitchison and Silvey 1958; Bergsma et al. 2013). Let $\mathbf{B}$ be the orthogonal complement of $\mathbf{Z}$, then $\mathbf{B}'\mathbf{Z}=0$. By premultiplying both sides of Eq. (23.2) by $\mathbf{B}'$, the categorical marginal model can be written as a set of constraints

$$\mathbf{B}'\mathbf{f}(\mathbf{C}'\mathbf{m})=\mathbf{B}'\mathbf{Z}\boldsymbol{\beta}=\mathbf{0}.$$

Because $\mathbf{B}$ and $\mathbf{C}$ are known design matrices, we can write $\mathbf{g}(\mathbf{m})=\mathbf{B}'\mathbf{f}(\mathbf{C}'\mathbf{m})$. Then, a concise notation for a categorical marginal model, as is used throughout the literature (e.g., Bergsma 1997; Kuijpers et al. 2013a; Van der Ark et al. 2008), is

$$\mathbf{g}(\mathbf{m})=\mathbf{0}. \tag{23.3}$$

Let D be the number of constraints on the expected frequencies $\mathbf{m}$. Each constraint is a scalar function, so, for example, $g_1(\mathbf{m})=d_1$, and can be collected in the vector $\mathbf{g}(\mathbf{m})$. So $\mathbf{g}(\mathbf{m})$ contains all constraints that are placed on a vector $\mathbf{m}$. The constraints in Eq. (23.3) constitute the categorical marginal model. Some examples of constraints are $\alpha=0.80$ and $\mu_1=\mu_2$.

23.3 Estimation Methods

23.3.1 Likelihood Methods

Likelihood methods use the constraint notation in Eq. (23.3) in combination with ML estimation. The unconstrained log-likelihood function (for more details see Bergsma 1997) is

$$\ell(\mathbf{m}|\mathbf{n})=\mathbf{n}'\log\mathbf{m}.$$

The maximum likelihood estimate $\widehat{\mathbf{m}}$ maximizes $\ell(\mathbf{m}|\mathbf{n})$ subject to the constraints implied by the categorical marginal model, $\mathbf{g}(\mathbf{m}) = \mathbf{0}$ [Eq. (23.3)], and to the constraint that $\sum_i m_i = \sum_i n_i = N$, where N denotes the total sample size.

Let $\boldsymbol{\lambda}$ be a $D \times 1$ vector of Lagrange multipliers and let ν be a single Lagrange multiplier, then under some regularity conditions, the ML estimates under Eq. (23.3) are a saddle point of the Lagrangian log-likelihood

$$\ell(\mathbf{m}|\mathbf{n}, \boldsymbol{\lambda}, \nu) = \mathbf{n}' \log \mathbf{m} - \nu(\mathbf{1}'\mathbf{m} - N) - \boldsymbol{\lambda}'\mathbf{g}(\mathbf{m}). \tag{23.4}$$

Bergsma (1997) proposed a Fisher scoring algorithm to find the vector $\mathbf{m}$ in Eq. (23.4). The fit of the categorical marginal model can be assessed by means of a likelihood ratio test $G^2 = 2\mathbf{n}' \log(\mathbf{n}/\widehat{\mathbf{m}})$ or a Pearsons' chi-square test $X^2 = (\widehat{\mathbf{m}} - \mathbf{n})'\mathbf{D}_{\widehat{\mathbf{m}}}^{-1}(\widehat{\mathbf{m}} - \mathbf{n})$ with D degrees of freedom. Here, $\mathbf{D}_{\widehat{\mathbf{m}}}$ is a diagonal matrix with the elements of vector $\widehat{\mathbf{m}}$ on the diagonal. Because ML estimation is based on the likelihood function, models can be compared and statistical inferences about parameters can be made.

23.3.2 Generalized Estimating Equations

GEE specifies a link function for the mean, and specifies the dependence of the variance on the mean. Furthermore, GEE replaces the often complex dependence structure by a so-called *working correlation* structure that is more straightforward to define. GEE can be used to fit any categorical marginal model expressed in terms of Eq. (23.2), but traditionally GEE is used for regression models for longitudinal data. In the case of longitudinal data, Y_{it} is the response for person i (with $i = 1, 2, \ldots, N$) on time point t (with $t = 1, 2, \ldots, T$). For GEE, for person i, the model of interest is equal to

$$h(\boldsymbol{\mu}_i) = \mathbf{Z}_i \boldsymbol{\beta}, \tag{23.5}$$

In Eq. (23.5), $h(\cdot)$ is a link function that applies element by element to vector $\boldsymbol{\mu}_i$. Vector $\boldsymbol{\mu}_i$ contains the expected responses (i.e., for person i, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})'$).

GEE links the mean μ to a linear predictor and in addition specifies a variance function that describes how the variance of Y_{it} depends on μ_{it} (Agresti 2013, p. 462). This model applies to the marginal distribution for each Y_{it}. The estimating equation used in GEE is

$$\sum_{i=1}^{N} \frac{\partial \boldsymbol{\mu}_i}{\partial} \boldsymbol{\beta} \mathbf{V}_i^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \tag{23.6}$$

where $\mathbf{y}_i$ is a vector with t observed responses (i.e., $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$), and $\mathbf{V}_i$ is an appropriately chosen working correlation matrix. The estimates of the parameters β_i in vector $\boldsymbol{\beta}$ are a solution of Eq. (23.6). For an exponential family $\mu_{it} = E(Y_{it})$.

For GEE, the particular working correlation structure needs to be specified for the relation between the t different responses of person i collected in $\mathbf{y}_i$. Different correlation structures can be chosen, depending on the nature of the dependencies between the different responses (Pawitan 2001, p. 396). Choosing a working correlation structure that approximates the true correlation structure between the dependent responses enhances the efficiency of the parameter estimates (Agresti 2013, p. 463). Commonly used specifications of the working correlation matrix are: (1) the independence structure, which treats the different responses as independent; thus, no dependency exists; (2) the exchangeable structure, which assumes constant dependency; thus, the correlations between the different responses are assumed to be equal for each observed response; (3) the autoregressive structure, which is often used for measurement over time, and treats the correlations as an exponential function of the time lag; thus, this structure assumes that observations farther apart in time have weaker correlations; and (4) the unstructured structure, which assumes a free specification of the working correlation matrix, implying a separate correlation for each pair of observations (see Agresti 2013, p. 462, and Pawitan 2001, pp. 396–397, for more details).

The choice of the working correlation structure determines the GEE estimates of the model parameters and the accompanying standard errors (Agresti 2013, pp. 462–463). However, even if the working correlation matrix is misspecified, the estimates of the parameters are consistent. In contrast, the estimates of the standard errors of the parameters are not accurate, and need to be adjusted for misspecification of the working correlation matrix by using the so-called sandwich estimator (e.g., Agresti 2013, p. 467). Liang and Zeger (1986) proposed estimating the GEE parameter estimates and the standard errors by means of a Fisher scoring algorithm.

GEE can also be used for fitting categorical marginal models that are defined by more complex functions than the link function $h()$, and by functions that have $\mathbf{n}$ rather than $\mathbf{y}$ as an argument. Here, $\mathbf{f}(\mathbf{C}'\mathbf{n})$ is a function of the observed responses and $\mathbf{Z}\boldsymbol{\beta} = \mathbf{f}(\mathbf{C}'\mathbf{m})$ is a function of the expected responses, so Eq. 23.6 becomes

$$\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{f}(\mathbf{C}'\mathbf{n}) - \mathbf{Z}\boldsymbol{\beta}) = \mathbf{0}. \tag{23.7}$$

A marginal model $\mathbf{Z}\boldsymbol{\beta}$ can represent a wide range of parameters or coefficients, with $\mathbf{f}(\mathbf{C}'\mathbf{n})$ being the corresponding sample value (Bergsma et al. 2013). Equation (23.7) can easily be solved by using

$$\boldsymbol{\beta} = (\mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}^{-1}\mathbf{f}(\mathbf{C}'\mathbf{n}), \tag{23.8}$$

which is equivalent to weighted least squares, with $\mathbf{V}^{-1}$ being a weight matrix. By means of Eq. (23.8), estimates for the parameters in $\boldsymbol{\beta}$ can be obtained.

23.4 Expressing Item Means and Cronbach's Alpha in Terms of the Generalized Exp-Log Notation

Maximizing the Lagrangian likelihood in Eq. (23.4) requires the matrix of first partial derivatives of $\mathbf{g}(\mathbf{m})$ with respect to $\mathbf{m}$. This matrix, also known as the Jacobian, is usually difficult to obtain. However, if $\mathbf{g}(\mathbf{m})$ is written in the so-called exp-log notation (Bergsma 1997; Kritzer 1977), the derivation of the Jacobian is straightforward, and an automated recursive algorithm can be used to compute the Jacobian for a particular categorical marginal model (Bergsma 1997, p. 68).

23.4.1 Item Means in Exp-Log Notation

For testing hypotheses about the means in vector $\boldsymbol{\mu}$, the coefficient should first be rewritten in the generalized exp-log notation. In this recursive exp-log notation let $\mathbf{A}_1$ and $\mathbf{A}_2$ be appropriate design matrices. Then $\boldsymbol{\mu}$ is equal to

$$\boldsymbol{\mu} = \exp(\mathbf{A}_2 \log(\mathbf{A}_1 \mathbf{m})). \tag{23.9}$$

Let $\mathbf{R}$ be a $J \times L$ matrix that contains all L possible item-score patterns. The rows of $\mathbf{R}$ correspond to the J different items. The item-score patterns in $\mathbf{R}$ are from left to right in lexicographic order, running from $00\ldots0$ to $zz\ldots z$ with the digit in the last row changing fastest and the digit in the first row changing slowest, just as is the case in vectors $\mathbf{m}$ and $\mathbf{n}$. Furthermore, let $\mathbf{u}'_L$ be a $1 \times L$ unit row vector. The $[J+1] \times L$ design matrix $\mathbf{A}_1$ is a concatenation of matrix $\mathbf{R}$ and vector $\mathbf{u}'_L$; that is,

$$\mathbf{A}_1 = \begin{pmatrix} \mathbf{R} \\ \mathbf{u}'_L \end{pmatrix}.$$

For a dichotomously scored test consisting of $J = 3$ items [Eq. (23.1)] this produces

$$\mathbf{A}_1\mathbf{n} = \begin{pmatrix} 0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1 \\ 0&1&0&1&0&1&0&1 \\ 1&1&1&1&1&1&1&1 \end{pmatrix} \begin{pmatrix} n_{abc}^{000} \\ n_{abc}^{001} \\ n_{abc}^{010} \\ n_{abc}^{011} \\ n_{abc}^{100} \\ n_{abc}^{101} \\ n_{abc}^{110} \\ n_{abc}^{111} \end{pmatrix} = \begin{pmatrix} \sum X_a \\ \sum X_b \\ \sum X_c \\ N \end{pmatrix}. \tag{23.10}$$

As the first three elements of the right-hand side of Eq. (23.10) show, $\mathbf{Rn}$ produces a vector containing the sum of the scores on items a, b, and c across respondents, and $\mathbf{u}'_L\mathbf{n}$ produces the sample size N.

Let $\mathbf{I}_J$ be an identity matrix of order J. Then, the $J \times [J+1]$ design matrix $\mathbf{A}_2$ is a concatenation of matrix $\mathbf{I}_J$ and unit vector $-\mathbf{u}_J$

$$\mathbf{A}_2 = \begin{pmatrix} \mathbf{I}_J & -\mathbf{u}_J \end{pmatrix}.$$

For the three items a, b, and c, substituting the right-hand side of Eq. (23.10) for $\mathbf{A}_1\mathbf{n}$, $\exp(\mathbf{A}_2 \log(\mathbf{A}_1\mathbf{n}))$ yields

$$\exp\left[\begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix} \log\begin{pmatrix} \sum X_a \\ \sum X_b \\ \sum X_c \\ N \end{pmatrix}\right] = \begin{pmatrix} \overline{X}_a \\ \overline{X}_b \\ \overline{X}_c \end{pmatrix}. \tag{23.11}$$

Equation (23.11) shows that $\exp(\mathbf{A}_2 \log(\mathbf{A}_1\mathbf{n}))$ produces the mean score for each item in a data set.

23.4.2 Coefficient α in Exp-Log Notation

Kuijpers et al. (2013a) used categorical marginal models for testing different hypotheses about Cronbach's alpha (Cronbach 1951). They showed that Cronbach's alpha, denoted by α, can be written as a function of $\mathbf{m}$ in the generalized exp-log notation:

$$\alpha = \mathbf{A}_5 \exp(\mathbf{A}_4 \log(\mathbf{A}_3 \exp(\mathbf{A}_2 \log(\mathbf{A}_1\mathbf{m})))), \tag{23.12}$$

where matrices $\mathbf{A}_1$ to $\mathbf{A}_5$ are appropriate design matrices. For the exact specification of the design matrices and more details about the procedure, see Kuijpers et al. (2013a).

23.5 Three Cases

23.5.1 Data

The use of the two different estimation methods to test three different cases is illustrated by means of a data set obtained by administering a questionnaire to $N = 496$ Dutch union members (Van der Veen 1992). The questionnaire measures the attitudes and opinions on general militancy, and consists of four subscales—

Table 23.1 Item means and Cronbach's alpha for each subscale

	Subscales			
Items	General attitude	Permissibility	Effectiveness	Intention
Strike	1.383	1.208	1.698	1.151
Work-to-rule	2.278	1.556	1.788	1.536
D. walkout	2.266	1.573	1.702	1.442
C. walkout	2.161	1.546	1.560	1.450
Protest meeting	2.653	2.258	1.835	1.589
Street protest	2.214	1.810	1.625	1.351
Cronbach's alpha	0.744	0.840	0.738	0.877

D. walkout demonstrative walkout, *C. walkout* collective walkout

General Attitude, Permissibility, Effectiveness, and Intention—which each contains six items. Each of the six items in a subscale refers to different actions union members can engage in, such as a strike, a protest meeting, or a street protest. For the subscales Permissibility and Intention, the answer categories range from 0 to 3, and for the subscales General Attitude and Effectiveness the answer categories range from 0 to 4. Table 23.1 shows the item means, and the values for Cronbach's alpha for each subscale.

Coefficient θ is used to express the different hypotheses. In what follows, θ will be replaced by either the mean (μ) or Cronbach's alpha (α). For ML estimation, we used the R package `cmm` (Bergsma and Van der Ark 2013), and for GEE, we used the R package `geepack` (Yan et al. 2012).

23.5.2 *Case 1:* $\theta = c$

First, we tested whether the mean value of General Attitude towards a Strike was significantly greater than 1 (sample value 1.383, Table 23.1). Second, we tested whether Cronbach's alpha of the subscale Permissibility was significantly greater than 0.80 (sample value 0.84, Table 23.1). Nunnally (1978, pp. 245–246) argued that tests used for making decisions about groups should have at least a reliability of 0.80. The research question is of the form $\theta > c$, and the associated null hypothesis is $\theta = c$.

For investigating $\theta = c$ by means of ML estimation, $\theta = c$ should be written in the constraint notation, $g(\mathbf{m}) = \theta - c = 0$. In the generalized exp-log notation, $g(\mathbf{m}) = \theta - c$ equals

$$g(\mathbf{m}) = \begin{bmatrix} 1 & -c \end{bmatrix} \exp\left(\begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \log\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \theta\right)\right). \tag{23.13}$$

The categorical marginal model estimates vector $\mathbf{m}$ under the constraint $\theta = c$.

Replacing θ in Eq. (23.13) by μ [Eq. (23.9)] and letting $c = 1$ yields the hypothesis $\mu = 1$. In general, G^2 pertains to a two-sided test. Here, the hypothesis is one-sided, so for a significance level of 0.05 the value of G^2 at the 2×0.05 significance level is used. Comparing the observed and expected frequencies allowed us to reject the hypothesis ($G^2 = 77.662$, $df = 1$, $p \leq 0.000$), and conclude that $\mu > 1$. Replacing θ in Eq. (23.13) by α [Eq. (23.12)], and letting $c = 0.80$ yields the hypothesis $\alpha = 0.80$. Comparing the observed and expected frequencies allowed us to reject the hypothesis ($G^2 = 9.489$, $df = 1$, $p = 0.002$), and conclude that $\alpha > 0.8$. This example illustrates that likelihood methods can be used to investigate research questions of the type $\theta = c$.

For testing whether $\theta = c$ by means of GEE, $\theta = c$ should be written as $\theta = \mathbf{Z}\boldsymbol{\beta}$. It trivially follows that $\mathbf{Z}$ equals the scalar 1, and $\boldsymbol{\beta} = c$, so $\hat{\theta}$ is trivially fixed to c, and the standard error is zero. The software did not provide goodness of fit statistics. Because $\hat{\theta}$ is fixed to c and no model fit statistics are available, we could not use GEE to meaningfully answer research questions that can be cast into $\theta = c$. This is in accordance with Skrondal and Rabe-Hesketh (2004, p. 200), who stated that GEE has limitations with respect to hypothesis testing and assessing model fit.

23.5.3 Case 2: $\theta_1 = \theta_2$

In this example, we considered whether the population means of the two items General Attitude towards a Demonstrative Walkout and General Attitude towards a Collective Walkout were equal. The sample means for the items were 2.266 and 2.161, respectively (see Table 23.1). Furthermore, we investigated whether the alphas of the two subscales Permissibility and Intention were equal. For the subscale Permissibility $\widehat{\alpha} = 0.840$, for subscale Intention $\widehat{\alpha} = 0.877$ (see Table 23.1). This categorical marginal model can be useful when one wants to compare the alphas of two subscales or tests, or for assessing change in reliability over time. Differences between the reliabilities of two alternate test forms can indicate that the two forms differ in content and measure slightly different traits (Nunnally 1978, p. 231).

For investigating this model by means of ML estimation, $\theta_1 = \theta_2$ has to be rewritten in the constraint notation, $g(\mathbf{m}) = \theta_1 - \theta_2 = 0$. Because the two coefficients we compared are dependent, vector $\mathbf{n}$ first should be premultiplied by $\mathbf{A}_0$, a marginal matrix (Bergsma et al. 2009, pp. 52–56). Multiplication by matrix $\mathbf{A}_0$ yields the marginal frequencies of the item-score patterns for both sets of items separately. Let L_1 and L_2 be the number of possible item-score patterns for which coefficients θ_1 and θ_2 are computed, respectively. Let $\otimes$ denote the Kronecker product. The general form of the $(L_1 + L_2) \times (L_1 L_2)$ matrix $\mathbf{A}_0$ is

$$\mathbf{A}_0 = \begin{pmatrix} \mathbf{I}_{L_1} \otimes \mathbf{u}'_{L_2} \\ \mathbf{u}'_{L_1} \otimes \mathbf{I}_{L_2} \end{pmatrix}.$$

After premultiplying vector $\mathbf{n}$ by $\mathbf{A}_0$, the two coefficients for the two sets of items are computed using design matrices that are constructed as follows. Let design matrix $\mathbf{A}_q$, with $q = 1,\ldots,q$, be the particular qth design matrix constructed for the particular coefficient. For testing the equality of two coefficients, design matrices $\mathbf{A}_1$ to $\mathbf{A}_q$ are the direct sum of $\mathbf{A}_q$ and $\mathbf{A}_q$. Since for each design matrix $\mathbf{A}_q$ the procedure is the same, it can be expressed in a general form

$$\mathbf{A}_q^* = \mathbf{A}_q \oplus \mathbf{A}_q = \begin{pmatrix} \mathbf{A}_q & 0 \\ 0 & \mathbf{A}_q \end{pmatrix}.$$

For more details, see Kuijpers et al. (2013a).

In the generalized exp-log notation, $g(\mathbf{m}) = \theta_1 - \theta_2$ equals

$$g(\mathbf{m}) = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}. \tag{23.14}$$

The categorical marginal model estimates vector $\mathbf{m}$ under the constraint $\theta_1 - \theta_2 = 0$. Then, vectors $\mathbf{m}$ and $\mathbf{n}$ are compared by means of G^2 in order to assess whether the two coefficients are equal.

If the coefficient of interest is the mean μ, the population means for the two items are denoted by μ_1 and μ_2, and calculated by using Eq. (23.9). For testing Case 2, θ_1 and θ_2 in Eq. (23.14) should be replaced by μ_1 and μ_2, respectively. Comparing the observed and expected frequencies allowed us to reject the null hypothesis ($G^2 = 5.429$, $df = 1$, $p = 0.020$), and conclude that the means are significantly different from each other.

If the coefficient of interest is Cronbach's alpha, the population alphas for the two subscales are denoted by α_1 and α_2, and calculated using Eq. (23.12). For testing Case 2, θ_1 and θ_2 in Eq. (23.14) should be replaced by α_1 and α_2, respectively. Comparing the observed and expected frequencies allowed us to reject the null hypothesis ($G^2 = 8.939$, $df = 1$, $p = 0.003$), and conclude that the alphas are not equal.

For GEE estimation, constraint $\theta_1 = \theta_2$ must be cast into Eq. (23.2). One possibility is defining a regression model with only an intercept β_0, which can be interpreted as the value of the coefficient under the constraint that $\theta_1 = \theta_2$. Let $\mathbf{Z} = \mathbf{u}_2$, then $\theta_1 = \theta_2$ is equivalent to

$$\mathbf{f}(\mathbf{C}'\mathbf{m}) = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \mathbf{u}_2\beta_0.$$

If the vector of sample estimates of θ_1 and θ_2 is represented by $(\hat{\theta}_1, \hat{\theta}_2)'$, then the estimating equation [Eq. (23.7)] reduces to

$$\mathbf{u}_2'\mathbf{V}^{-1}\left(\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} - \mathbf{u}_2\beta_0\right) = \mathbf{0}. \tag{23.15}$$

For an arbitrary correlation matrix $\mathbf{V}$, Eq. (23.15) reduces to

$$\begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} - \mathbf{u}_2\beta_0 = \mathbf{0},$$

which is minimized for $\hat{\beta}_0 = \frac{\hat{\theta}_1+\hat{\theta}_2}{2}$. So the estimated values for θ_1 and θ_2 are then both equal to the mean of the two values. The hypothesis $\theta_1 - \theta_2 = 0$ can be tested by computing the standard errors by means of the sandwich estimator, computing the confidence interval, and then checking whether 0 is included in the interval.

Using GEE for testing the equality of the means of the two items General Attitude towards a Demonstrative Walkout and General Attitude towards a Collective Walkout, the analysis only estimates a mean value for both values and a standard error, model fit statistics are not available. The estimated mean value for the two means is equal to 2.214, which is obtained independent of the correlation structure. The standard error equals 0.037. To test whether the hypothesis of equal means could be rejected, a 95 % Wald confidence interval for the difference between the two means (denoted by $\Delta\mu$) was constructed using $\widehat{\Delta\mu} \pm 1{,}96 * se(\widehat{\Delta\mu})$. Zero was not included in the interval, so the means are significantly different. GEE was also used for testing the equality of the two alphas of the subscales Permissibility and Intention. The mean value for the two alphas equaled 0.859. The standard error equaled 0.013. A 95 % confidence interval for the difference between the two alphas was constructed in a way similar to the computation for the means. Zero was not included in the confidence interval, so the alphas are significantly different.

23.5.4 Case 3: $\theta = \beta_0 + \beta_1 X$

Here, the question was whether the Effectiveness of an action can explain the General Attitude towards that action. We used Effectiveness measured for a Strike (denoted by X_1) and a Work-to-rule (X_2) as the explanatory variables, and General Attitude measured for a Strike (Y_1) and a Work-to-rule (Y_2) as the outcome variables. Hence, we had $T = 2$ actions and $z + 1 = 5$ levels of the exploratory variable. In longitudinal research, one would consider T time points rather than actions. Estimating a regression model in which Cronbach's alpha is the dependent variable seemed artificial from a substantive point of view. Hence, we only investigated Case 3 for μ. However, there are other situations in which testing the effects of one or more (continuous) variables on the value of a particular coefficient is interesting. For instance, using the log-odds ratio as a measure of association, Bergsma et al. (2013) tested whether the association between two categorical variables remained stable over time.

The regression model is $\mathbf{f}(\mathbf{C}'\mathbf{m}) = \mathbf{Z}\boldsymbol{\beta}$ [Eq. (23.2)], where $\mathbf{f}(\mathbf{C}'\mathbf{m})$ is the $T(z+1) \times 1$ vector of conditional means:

$$\begin{pmatrix} E(Y_1|X_1=0) \\ E(Y_2|X_2=0) \\ E(Y_1|X_1=1) \\ E(Y_2|X_2=1) \\ E(Y_1|X_1=2) \\ E(Y_2|X_2=2) \\ E(Y_1|X_1=3) \\ E(Y_2|X_2=3) \\ E(Y_1|X_1=4) \\ E(Y_2|X_2=4) \end{pmatrix}.$$

Matrix $\mathbf{Z}$ is a $T(z+1) \times 2$ design matrix:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 3 \\ 1 & 3 \\ 1 & 4 \\ 1 & 4 \end{pmatrix}.$$

The first column is a column of ones, and the second column contains the levels of X_1 and X_2. Vector $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ contains the intercept and the regression parameter. Vector $\mathbf{m}$ refers to the joint distribution of (X_1, X_2, Y_1, Y_2).

For ML estimation, first $\mathbf{C}'$ and $\mathbf{f}$ should be determined. In our example, premultiplying $\mathbf{n}$ by the $(T(z+1)^2 \times L)$ marginal matrix

$$\mathbf{C}' = \begin{pmatrix} \mathbf{I}_{z+1} \otimes \mathbf{u}'_{z+1} \otimes \mathbf{I}_{z+1} \otimes \mathbf{u}'_{z+1} \\ \mathbf{u}'_{z+1} \otimes \mathbf{I}_{z+1} \otimes \mathbf{u}'_{z+1} \otimes \mathbf{I}_{z+1} \end{pmatrix}$$

produces the bivariate marginal frequencies of (X_1, Y_1) and (X_2, Y_2). Function $\mathbf{f}$ consists of two design matrices: $\mathbf{A}_1$ and $\mathbf{A}_2$. Let $\mathbf{r}_{z+1}$ be a $(z+1) \times 1$ vector containing scores $0, 1, \ldots, z$; then $\mathbf{A}_1$ is a $2T(z+1) \times T(z+1)^2$ matrix

$$\mathbf{A}_1 = \mathbf{I}_T \otimes \begin{pmatrix} \mathbf{I}_{z+1} \otimes \mathbf{r}'_{z+1} \\ \mathbf{I}_{z+1} \otimes \mathbf{u}'_{z+1} \end{pmatrix}$$

and

$$\mathbf{A}_2 = \mathbf{I}_T \otimes \left(\mathbf{I}_{(z+1)} \; -\mathbf{I}_{(z+1)} \cdot \right)$$

Hence,

$$\mathbf{f}(\mathbf{C}'\mathbf{m}) = \exp\left(\mathbf{A}_2 \log\left(\mathbf{A}_1 \mathbf{C}'\mathbf{m}\right)\right).$$

Second, $\mathbf{B}$, the orthogonal complement of $\mathbf{Z}$, should be determined such that $\mathbf{B}'\mathbf{Z} = \mathbf{0}$. Third, the expected categorical marginal model $\mathbf{B}'\mathbf{f}(\mathbf{C}'\mathbf{m}) = \mathbf{0}$ is estimated, producing estimates for vector $\mathbf{m}$. Using this method for maximizing the likelihood includes the constraints, such that the expected frequencies in vector $\widehat{\mathbf{m}}$ sum to N (Agresti 2013, p. 460). Fourth, the estimates $\widehat{\mathbf{m}}$ are plugged into model $\mathbf{f}(\mathbf{C}'\mathbf{m}) = \mathbf{Z}\boldsymbol{\beta}$, producing $\mathbf{f}(\mathbf{C}'\widehat{\mathbf{m}})$. Fifth, parameters $\boldsymbol{\beta}$ are obtained by solving

$$\hat{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{f}(\mathbf{C}'\widehat{\mathbf{m}}).$$

Finally, the standard errors of $\hat{\boldsymbol{\beta}}$ are computed using the delta method (for more details, see, for instance, Bergsma et al. 2009, pp. 71–73), so that each individual parameter in $\boldsymbol{\beta}$ can be tested for significance.

The regression model describes the linear relation between the means that are calculated for each dependent variable given the response to the corresponding independent variable (i.e, the means for Y_1 given the different scores on X_1, and the means for Y_2 given the different scores on X_2). Table 23.2 provides the estimates for the parameters in the regression model.

The categorical marginal model also tests whether the regression model that assumes a linear relation between the means fits the data. The results of the analysis showed that the linear regression model does not fit the data, with $G^2 = 173.071$, $df = 8$ and $p < 0.000$, which implies that the means cannot be fitted onto a single straight line; thus, there is not a strictly common linear relation between the conditional means of Y_1 and Y_2 given the scores on X_1 and X_2. However, the regression coefficient is significant, meaning that the scores on X_1 and X_2 have a significant effect on the mean scores of Y_1 and Y_2.

Also, GEE was used to test whether the items Effectiveness of a Strike and Effectiveness of a Work-to-Rule predicted the mean response to General Attitude towards a Strike and General Attitude towards a Work-to-Rule. Table 23.3 shows the

Table 23.2 Parameter estimates using ML estimation

Parameter	Estimate	Standard error
β_0	1.003	0.063
β_1	0.471	0.032

Table 23.3 Parameter estimates using GEE estimation

Parameter	Estimate	Standard error
β_0	0.921	0.056
β_1	0.522	0.027

GEE estimates of the parameters in the regression model, as defined by Eq. (23.2). The regression coefficient is significantly different from zero, which indicates that the scores on X_1 and X_2 have a significant effect on the mean scores of Y_1 and Y_2. For the regression problems, alternative model fit statistics exist for GEE (e.g., Lipsitz and Fitzmaurice 2009, pp. 62–64; Molenberghs and Verbeke 2005, pp. 160–161) but these statistics were unavailable in the R package `geepack`, so the model fit could not be investigated.

23.6 Discussion

For this study, we explored to what extent the two estimation methods are appropriate for investigating and testing three types of research questions. The two estimation methods, ML and GEE, both have advantages and disadvantages. ML estimation is based on the likelihood function, so that model fit statistics can be obtained, models can be compared, and inferences about individual parameters can be made. In contrast to ML estimation, GEE does not assume a specific probability model for the data, but only assumes a mean-variance relationship for the response variable, making it impossible to obtain likelihood based model fit statistics. Furthermore, GEE replaces the often complex dependence structure by a simpler working correlation matrix. Therefore, GEE is more straightforward to compute than ML methods. For a large number of items, in contrast to GEE, using ML estimation becomes problematic, since it uses each cell of the contingency table for computation of the estimates (Bergsma et al. 2013; Van der Ark et al. 2013). However, ML estimation is asymptotically efficient (e.g., Agresti 2013), whereas GEE is not when the working correlation structure is not correctly specified.

By means of the three cases, we showed that ML estimation has to be preferred when one is more interested in testing hypotheses and assessing the fit of the marginal model. Both methods are appropriate when one investigates the effect of the independent factors in regression models. For Case 1, GEE could not be used. This is in line with Skrondal and Rabe-Hesketh (2004, p. 200) who stated that GEE has limitations with respect to hypothesis testing and assessing model adequacy. An alternative to solve some of the limitations would be to estimate the standard error of the saturated model, and then use a Wald-based confidence interval to assess whether the value c is included in the confidence interval (Lipsitz and Fitzmaurice 2009, p. 55). Furthermore, since standard goodness of fit statistics are unavailable for GEE, Lipsitz and Fitzmaurice (2009, pp. 62–64) suggested some alternative model fit diagnostics. For Case 2, ML was easier to apply than GEE, and for ML model fit statistics could be obtained right away. For Case 3, we found that GEE was easier to apply than ML from a computational perspective.

ML estimation uses all item-score patterns that are possible for a set of items, so all elements in vector $\mathbf{n}$ are used. ML estimation becomes problematic for large numbers of items (e.g., Agresti 2013, p. 462) because the number of elements in vector $\mathbf{n}$ and the size of the design matrices increase rapidly (Bergsma et al. 2013;

Van der Ark et al. 2013). For instance, for a set of ten items ($J = 10$) each with five answer categories ($z+1 = 5$), the number of elements in vector $\mathbf{n}$ is equal to $(z+1)^J = 5^{10} = 9{,}765{,}625$. An alternative is using MEL estimation (Owen 2001). MEL uses only the observed item-score patterns, so that the zero-frequencies in vector $\mathbf{n}$ can be ignored. MEL uses much less memory space than ML estimation, and as a result it also is computationally less complex. Therefore, computation time is much shorter, and MEL can be used for large numbers of variables. However, for large sparse contingency tables the empty set problem and the zero likelihood problem can occur when using MEL estimation (for details, see Van der Ark et al. 2013; also see Bergsma et al. 2012), which causes MEL to break down. Van der Ark et al. (2013) proposed MAEL estimation as a solution for the problems with MEL. MAEL uses all observed item-score patterns, plus a few well-chosen unobserved item-score patterns, the choice of which depends on different factors; see Van der Ark et al. (2013) for more details.

For marginal models, GEE and the likelihood methods require further research. We only illustrated the use of both estimation methods by means of three simple cases for two different coefficients. Many more cases and situations can be investigated. The research can be extended to more complex models and to other coefficients. Furthermore, the cases also can be investigated for MEL and MAEL estimation, which can be compared to GEE estimation in order to investigate which method yields more efficient estimates.

Acknowledgements The authors would like to thank Klaas Sijtsma for commenting on earlier versions of the manuscript.

References

Agresti A (2013) Categorical data analysis. Wiley, Hoboken

Aitchison J, Silvey SD (1958) Maximum likelihood estimation of parameters subject to restraints. Ann Math Stat 29:813–828

Bergsma WP (1997) Marginal models for categorical data. Tilburg University Press, Tilburg

Bergsma WP, Van der Ark LA (2013) cmm: An R-package for categorical marginal models (Version 0.7) [computer software]. http://cran.r-project.org/web/packages/cmm/index.html. Accessed 24 Feb 2013

Bergsma WP, Croon MA, Hagenaars JA (2009) Marginal models: for dependent, clustered, and longitudinal categorical data. Springer, New York

Bergsma WP, Croon MA, Van der Ark LA (2012) The empty set and zero likelihood problems in maximum empirical likelihood estimation. Electron J Stat 6:2356–2361

Bergsma WP, Croon MA, Hagenaars JA (2013) Advancements in marginal modelling for categorical data. Sociol Methodol 43:1–41

Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16: 297–334

Kesteloot H, Geboers J, Joossens JV (1989) On the within-population relationship between nutrition and serum lipids, the birnh study. Eur Heart J 10:196–202

Kritzer HM (1977) Analyzing measures of association derived from contingency tables. Sociol Methods Res 5:35–50

Kuijpers RE, Van der Ark LA, Croon MA (2013a) Testing hypotheses involving Cronbach's alpha using marginal models. Br J Math Stat Psychol 66:503–520. doi: 10.1111/bmsp.12010
Kuijpers RE, Van der Ark LA, Croon MA (2013b) Standard errors and confidence intervals for scalability coefficients in Mokken scale analysis using marginal models. Sociol Methodol 43:42–69
Lang JB (2004) Multinomial-Poisson homogeneous models for contingency tables. Annals of Statistics, 32:340–383.
Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. Biometrika 73:13–22
Lipsitz S, Fitzmaurice G (2009) Generalized estimating equations for longitudinal data analysis. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) Longitudinal data analysis. Chapman & Hall/CRC, Boca Raton, pp 43–78
Molenberghs G, Verbeke G (2005) Models for discrete longitudinal data. Springer, New York
Nunnally JC (1978) Psychometric theory. McGraw-Hill, New York
Owen AB (2001) Empirical likelihood. Chapman & Hall/CRC, London
Pawitan Y (2001) In all likelihood: statistical modelling and inference using likelihood. Clarendon Press, Oxford
Skrondal A, Rabe-Hesketh S (2004) Generalized latent variable modeling: multilevel, longitudinal, and structural equation models. Chapman & Hall/CRC, Boca Raton
Van der Ark LA, Croon MA, Sijtsma K (2008) Mokken scale analysis for dichotomous items using marginal models. Psychometrika 73:183–208
Van der Ark LA, Bergsma WP, Croon MA (2013) Augmented empirical likelihood estimation of categorical marginal models for large sparse contingency tables (under review)
Van der Veen G (1992) Principes in praktijk: CNV-leden over collectieve acties [Principles into practice. Labour union members on means of political pressure]. J.H. Kok, Kampen
Yan J, Højsgaard S, Halekoh U (2012) geepack: Generalized Estimating Equation package. (Version 1.1-6) [computer software]. http://cran.r-project.org/web/packages/geepack/index.html. Accessed 13 Dec 2013