Computational Statistics and Data Analysis 121 (2018) 57–70

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Concordance correlation coefficients estimated by variance components for longitudinal normal and Poisson data

Miao-Yu Tsai [a,*], Chao-Chun Lin [b]

[a] Institute of Statistics and Information Science, College of Science, National Changhua University of Education, Changhua, Taiwan
[b] Department of Radiology, China Medical University Hospital, Taichung, Taiwan

ARTICLE INFO

Article history:
Received 9 May 2017
Received in revised form 9 September 2017
Accepted 2 December 2017
Available online 20 December 2017

Keywords:
Agreement
Concordance correlation coefficient
Generalized linear mixed-effects models
Longitudinal repeated measurements
Variance components

ABSTRACT

The concordance correlation coefficient (CCC) is widely used to assess agreement between two observers for continuous responses. Further, the CCC is extended for measuring agreement with discrete data. This paper proposes a variance components (VC) approach that allows dependency between repeated measurements over time to assess intra-agreement for each observer and inter- and total agreement among multiple observers simultaneously under extended three-way generalized linear mixed-effects models (GLMMs) for longitudinal normal and Poisson data. Furthermore, we propose a weight matrix to compare with existing weight matrices. Simulation studies are conducted to compare the performance of the VC, generalized estimating equations and U-statistics approaches with different weight matrices for repeated measurements from longitudinal normal and Poisson data. Two applications, of myopia twin and of corticospinal diffusion tensor tractography studies, are used for illustration. In conclusion, the VC approach with consideration of the correlation structure of longitudinal repeated measurements gives satisfactory results with small mean square errors and nominal 95% coverage rates for all sample sizes.

1. Introduction

In many clinical studies, measurements of interest can be assessed on dichotomous, polychotomous, ordinal, count or continuous scales by several methods (e.g., observers, raters, devices, technologies, etc.). Note that hereafter we refer to measurement methods as observers. Furthermore, a subject can be measured many times by each of several observers to produce repeated measurements over time for longitudinal data. Therefore, it is necessary to assess the agreement or reproducibility between replicated readings produced by a single observer or among multiple measurement methods. The concordance correlation coefficient (CCC) proposed by Lin (1989), is used to assess agreement between two observers on a continuous scale by measuring the variation of the linear relationship between each pair of data from a 45° line through the origin. Some extensions of CCC are used to measure agreement among more than two observers (Barnhart et al., 2002; Carrasco and Jover, 2003; King and Chinchilli, 2001) or to produce CCC estimates on discrete data (Carrasco and Jover, 2005; King and Chinchilli, 2001).

There are at least three frequentist research directions for the estimation of CCC. One direction is to estimate CCC via generalized estimating equations (GEE). Barnhart and Williamson (2001) proposed the GEE approach with three sets of estimating equations to estimate CCC between two observers for continuous data. Further, the estimation of indices for

* Correspondence to: Institute of Statistics and Information Science, National Changhua University of Education, No.1, Jin-De Road, Chang-Hua 500, Taiwan.

E-mail addresses: mytsai@cc.ncue.edu.tw (M.-Y. Tsai), d16021@mail.cmuh.org.tw (C.-C. Lin).

https://doi.org/10.1016/j.csda.2017.12.003
0167-9473/

assessing intra-, inter- and total agreement with replicated readings produced by multiple observers simultaneously via GEE has also been constructed (Barnhart et al., 2005). As an example of the second direction, King and Chinchilli (2001) proposed a generalization of CCC for continuous and categorical data by using U-statistics (US). Furthermore, a class of repeated measures CCC with an unstructured correlation structure of repeated measurements that can be expressed as a ratio of functions of U-statistics was developed by King et al. (2007a, b). In addition to GEE and U-statistics, Carrasco and Jover (2003) proposed a third means of estimating CCC for more than two observers through variance components (VC) under a two-way linear mixed model (LMM). Carrasco (2010) extended the estimation of CCC via the VC approach to generalized linear mixed-effects models (GLMMs) for count data. When repeated measurements are assessed from different observers over time, the repeated measures CCC for longitudinal repeated measurements through VC was also developed (Carrasco et al., 2009). However, the repeated measures CCC through VC proposed by Carrasco et al. (2009) can only assess total agreement among multiple observers for continuous responses. Moreover, it has been adapted to estimate the CCC only with a diagonal matrix of weights between different repeated measurements over time for longitudinal data. For count data, the two-way GLMM incorporating only the subject- and observer-specific random effects, together with the random subject–observer interaction effect proposed by Carrasco (2010) has been considered but is not appropriate to estimate CCC for longitudinal data.

In this paper, we use the definitions of agreement coefficients proposed by Barnhart et al. (2005) and King et al. (2007a) with specifying a weight matrix to obtain the indices of intra-, inter- and total agreement through VC from an extended three-way GLMM for longitudinal normal and Poisson data. In addition, we propose a weight matrix to compare with the general weight matrix provided by King et al. (2007a) and develop the estimation of simultaneously assessing intra- and inter-observer agreement, as well as total agreement with information allowing dependency between repeated measurements over time. The rest of this paper is organized as follows. Section 2 introduces the extended three-way GLMM for agreement data, together with the definitions of intra-agreement for each observer and inter- and total agreement among multiple observers for longitudinal normal and Poisson data. In addition, the proposed weight matrix and the estimation of these agreement coefficients under VC for longitudinal repeated measurements are provided in Section 2. Simulation studies in Section 3 are constructed to compare the performance of the VC, GEE and US approaches. Section 4 presents the applications of myopia monozygotic (MZ) twin and corticospinal diffusion tensor tractography (DTT) studies, and shows the results from the VC and GEE approaches. Final conclusions and discussions are given in Section 5.

2. Models and methods

2.1. Generalized linear mixed-effects models (GLMMs)

Here we consider GLMMs to incorporate the possible heterogeneity due to different subjects, and subject–observer and subject–time interactions. Let y_{ijkl} denote the lth observed reading assessed by the jth observer at the kth time for the ith subject, where $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$ and $l = 1, \ldots, L$. The total number of observations is $n = I \times J \times K \times L$. Let $\mathbf{Y}_i = (y_{i111}, \ldots, y_{i1KL}, \ldots, y_{iJ11}, \ldots, y_{iJKL})^t$ be a $JKL \times 1$ vector of responses for the ith subject. Conditional on the unobserved q-dimensional random-effects vector $\mathbf{b}_{ijk}$, y_{ijkl} are independently distributed from an exponential family with means $E(y_{ijkl}|\mathbf{b}_{ijk}) = \mu_{ijk}$ and variances $Var(y_{ijkl}|\mathbf{b}_{ijk}) = v_{ijk} = \phi v(\mu_{ijk})$. Here $v(\cdot)$ is a specified variance function and ϕ is a dispersion parameter that may not be known. Through the link function $g(\mu_{ijk}) = \eta_{ijk}$, the conditional mean associated with a linear predictor is given by

$$\eta_{ijk} = \mathbf{x}_{ijk}^t \boldsymbol{\beta} + \mathbf{z}_{ijk}^t \mathbf{b}_{ijk}, \tag{1}$$

where $g(\cdot)$ is a monotone differentiable link function, the vectors $\mathbf{x}_{ijk}(p \times 1)$ and $\mathbf{z}_{ijk}(q \times 1)$ are explanatory variables associated with the fixed effects $\boldsymbol{\beta}$ and the random effects $\mathbf{b}_{ijk}$, respectively. We extend the three-way LMM proposed by Carrasco et al. (2009) and Tsai (2017) to the GLMM with incorporating repeated measurements rated by an observer at a certain time for each subject. The extended three-way GLMM can be written as

$$\eta_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk}, \tag{2}$$

where μ is the overall mean, α_i is the subject-specific random effect assumed to be distributed as $\alpha_i \sim N(0, \sigma_\alpha^2)$, β_j is the observer-specific fixed effect, γ_k is the time-specific fixed effect, $\alpha\beta_{ij}$ is the random subject–observer interaction effect assumed to be distributed as $\alpha\beta_{ij} \sim N\left(0, \sigma_{\alpha\beta}^2\right)$, $\alpha\gamma_{ik}$ is the random subject–time interaction effect assumed to be distributed as $\alpha\gamma_{ik} \sim N\left(0, \sigma_{\alpha\gamma}^2\right)$, and $\beta\gamma_{jk}$ is the fixed observer–time interaction effect. The parameters α_i, $\alpha\beta_{ij}$ and $\alpha\gamma_{ik}$ are all assumed to be mutually independent. We assume that the fixed-effects vector is $\boldsymbol{\beta} = (\mu, \beta_1, \ldots \beta_J, \gamma_1, \ldots, \gamma_K, \beta\gamma_{11}, \ldots, \beta\gamma_{JK})^t$ and the random-effects vector $\mathbf{b}_{ijk} = (\alpha_i, \alpha\beta_{ij}, \alpha\gamma_{ik})^t$ follows a multivariate normal distribution, $\mathbf{b}_{ijk} \sim MVN(\mathbf{0}, \mathbf{G})$, where $\mathbf{G}$ is a diagonal matrix with elements σ_α^2, $\sigma_{\alpha\beta}^2$ and $\sigma_{\alpha\gamma}^2$ on the diagonal and zero otherwise.