



Repeated measures proportional odds logistic regression analysis of ordinal score data in the statistical software package R

Nick R. Parsons*, Matthew L. Costa, Juul Achten, Nigel Stallard

Warwick Medical School, University of Warwick, UK

ARTICLE INFO

Article history:

Available online 9 August 2008

ABSTRACT

The widely used proportional odds model is developed for correlated repeated ordinal score data, using a modified version of the generalized estimating equation (GEE) method for model fitting for a range of working correlation models. The algorithm developed estimates the correlation parameter, by minimizing the generalized variance of the regression parameters at each step of the fitting algorithm. Methods for parameter estimation are described for the widely used uniform and first-order autoregressive correlation models, for data potentially recorded at irregularly spaced time intervals. A full implementation of the algorithm (*repolr*) in the R statistical software package, that both tests the assumption of proportional odds and accommodates missing data, is described and applied to a clinical trial of post-operative treatment, after rupture of the Achilles tendon and a study of patient pain response after hip joint resurfacing.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Ordinal score variables, that have a clear hierarchical ordering, recorded from the same patient or experimental unit over time are common in clinical research studies. For instance a pain or discomfort score (mild, moderate or severe) may be recorded at a number of occasions, corresponding to routine assessments or visits to a clinic, or may be used to establish a patient's state at entry to a clinical trial and again at the conclusion, with differences between assessments attributed to the treatment effect. Ordinal score scales are often used to quantify symptoms or conditions that are difficult or impossible to assess in any other way, and are thus commonly used for evaluating the effectiveness of many surgical procedures (e.g. hip replacement or tendon repair) at a number of repeat occasions post-operatively; these two applications provide motivating examples for this paper.

Ordinal scores are common in health-related research, and many approaches have been described for developing regression models for efficient analysis of these data (Lall et al., 2002). In particular, the modelling of repeated ordinal scores is a widely studied statistical problem and an active area of research; Agresti and Natarajan (2001) provide a comprehensive review of available models and methods. A number of parametric and nonparametric methods have been proposed for analysis of repeated ordinal responses; the two most widely used approaches to modelling differ in their formulation based on either population-averaged or subject specific effects (Diggle et al., 2002). The latter models represent subject (or cluster) effects by a random effects term in the model; see for example Coull and Agresti (2000). The most widely used approach for population-averaged (marginal) models for repeated ordinal responses (e.g. Clayton (1992), Lipsitz et al. (1994), and Kenward et al. (1994)) is the generalized estimating equation (GEE) method originally proposed by Liang and Zeger (1986) for the proportional odds model (McCullagh, 1980); the model formulation follows directly from considering an ordinal score to be a continuous (unobserved) variable that is divided into a small number of categories, in an attempt to provide an objective evaluation of a quantity that would otherwise be impossible to measure directly. GEE methods provide reliable parameter

* Corresponding author. Tel.: +44 2476968627; fax: +44 2476968617.

E-mail address: nick.parsons@warwick.ac.uk (N.R. Parsons).

estimates in most situations, however, a number of alternative GEE methods have been suggested that overcome some of the pitfalls in estimation of the correlation parameters that occur in many of the most widely used methods (Chaganty, 1997; Parsons et al., 2006).

Commercial software for model fitting using the GEE method is now widely available (Horton and Lipsitz, 1999), although implementation for ordinal regression models is often difficult, even for an experienced user. In practice in many clinical trials, information in the data is often not used efficiently, and ordinal scores are treated as if they were continuous variables for the purposes of analysis (Forrest and Andersen, 1986; Lavalley and Felson, 2002; Jakobsson, 2004); the main reason for this is the lack of available, simple to use, statistical tools. In order to address this perceived problem we have generalized and extended the algorithm originally suggested by Parsons et al. (2006), which they show to be particularly robust and reliable, for model parameter estimation and analysis of typical clinical trials data and written functions for routine implementation and distribution in the freely available statistical software R (R Development Core Team, 2007). We extend the methodology, which was previously only available for the very limited case of complete ordinal scores recorded at evenly spaced time intervals with an assumed first-order autoregressive correlation model, in a number of ways that will facilitate more widespread use for analysis of data from clinical trials. To achieve this, we develop correlation models that accommodate positively correlated repeated ordinal scores within subjects recorded at any desired spacing of time intervals and, as data from clinical follow-up are in general incomplete due to patient drop-out, we modify the algorithm to allow missing data in either the response or explanatory variables, and also incorporate a formal test of the proportional odds assumption into the algorithm. We introduce the proportional odds model in Section 2.1, describe an algorithm for parameter estimation in Section 2.3 and give examples of correlation models appropriate for clinical trial data in Section 2.4. Section 3 describes R functions that implement the GEE method, and other available methods in R for modelling ordinal score data, and Section 4 demonstrates model fitting for two example sets of data; first a surgical trial of post-operative treatment after rupture of the Achilles tendon and second, a study of patient pain response after hip joint resurfacing.

2. Methods

2.1. Proportional odds model

Suppose in a clinical research study, that ordinal scores are recorded at each of T time points from N subjects (patients) using a scale with K categories. Let Y_{it} represent the score on the i th subject at the t th time, where the individual scores are made on an ordered categorical scale, where integers from 1 to K represent the score categories and K represents the optimum score category (for example, 1 = mild, 2 = moderate and 3 = severe). Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iT})^T$ be a vector of ordinal scores for the i th subject at T time points and let \mathbf{x}_{it} be a matrix of observed explanatory variables (i.e. treatments, baseline covariates etc). The probability that an ordinal score Y_{it} falls within a particular score category can be related to the measured explanatory variables \mathbf{x}_{it} by a proportional odds model based on cumulative logits (McCullagh and Nelder, 1989, Chapter 5).

$$h(\mu_{itk}) = \beta_{0k} + \mathbf{x}_{it}^T \boldsymbol{\beta}. \quad (1)$$

Here, $\mu_{itk} = P(Y_{it} \leq k)$ is the probability that score Y_{it} is less than or equal to k , the parameters β_{0k} for $k = 1, \dots, K - 1$ are so-called cut-points ($-\infty < \beta_{01} < \dots < \beta_{0(K-1)} < \infty$), $\boldsymbol{\beta}$ is a vector of model regression parameters and $h(\cdot)$ is the logit function. The cumulative link model (Agresti, 2002) used in Eq. (1) effectively transforms the ordinal scale to a continuous scale based on the linear predictor $\mathbf{x}_{it}^T \boldsymbol{\beta}$, which proves to be a convenient model for relating the ordinal scores to the measured variates; the cut points β_{0k} define the class boundaries on the continuous scale of measurement defined by the linear predictor. The model is known as the proportional odds model, or proportional odds logistic regression model, because the ratio of the odds of the event $Y_{it} \leq k$ for any pair of sets of explanatory variables is independent of the choice of score category k . The applicability and intuitive assumptions behind the proportional odds model are the reasons for its popularity and widespread use. The assumption of proportional odds can be tested using the efficient score test suggested by Stiger et al. (1999). The cumulative logit model naturally results from an underlying latent variable with a logistic distribution, latent variables with assumed normal or extreme value distributions similarly lead to the probit or complementary log–log links, which provide practical alternatives to the proportional odds model (McCullagh and Nelder, 1989).

2.2. GEE

The most widely used method for parameter estimation for the proportional odds model when modelling correlated data, and the one described here, is the generalized estimating equation (GEE) approach (Hardin and Hilbe, 2002), first proposed by Liang and Zeger (1986). The method avoids strong assumptions about the distribution and the dependence structure of the repeated measures that would typically be required for a full likelihood model, by solving multivariate analogues of the quasi-likelihood estimating equations (Wedderburn, 1974). Resulting estimates are not maximum likelihood, but are asymptotically Normal and consistent (Liang and Zeger, 1986). Various options for implementation of GEE have been described elsewhere, but the most widely used method, and the one we adopt here, models the effects of the explanatory variables $\boldsymbol{\beta}$ and the correlation between repeated observations on an individual subject $\boldsymbol{\alpha}$, separately. The correlation

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات