



Regression models for grouped survival data: Estimation and sensitivity analysis

Elizabeth M. Hashimoto^a, Edwin M.M. Ortega^{a,*}, Gilberto A. Paula^b, Mauricio L. Barreto^c

^a Departamento de Ciências Exatas, Universidade de São Paulo, Brazil

^b Departamento de Estatística, Universidade de São Paulo, Brazil

^c Instituto de Saúde Coletiva, Universidade Federal da Bahia, Brazil

ARTICLE INFO

Article history:

Received 26 January 2010

Received in revised form 9 July 2010

Accepted 8 August 2010

Available online 14 August 2010

Keywords:

Censored data

Grouped survival data

Link function

Regression model

Sensitivity analysis

ABSTRACT

In this study, regression models are evaluated for grouped survival data when the effect of censoring time is considered in the model and the regression structure is modeled through four link functions. The methodology for grouped survival data is based on life tables, and the times are grouped in k intervals so that ties are eliminated. Thus, the data modeling is performed by considering the discrete models of lifetime regression. The model parameters are estimated by using the maximum likelihood and jackknife methods. To detect influential observations in the proposed models, diagnostic measures based on case deletion, which are denominated global influence, and influence measures based on small perturbations in the data or in the model, referred to as local influence, are used. In addition to those measures, the local influence and the total influential estimate are also employed. Various simulation studies are performed and compared to the performance of the four link functions of the regression models for grouped survival data for different parameter settings, sample sizes and numbers of intervals. Finally, a data set is analyzed by using the proposed regression models.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Grouped survival data are obtained from studies in which all sample units are evaluated at the same moment, which implies the occurrence of an excessive number of tied lifetimes. Hence, to eliminate the presence of ties, the lifetimes are grouped into intervals so that only the information concerning the intervals in which the individuals failed or were censored is available. According to Aranda-Ordaz (1983), the presence of many tied times is a particularly problematic point in model fitting, since the verification of the model's assumptions is based on continuous data. The first ideas concerning the treatment of grouped survival data arose in the 1980s from articles by Cox (1972) and Kalbfleisch and Prentice (1973), which presented semi-parametric models for analyzing this type of data. Some more recent applications on grouped survival data can be found in the literature. For instance, Hertz-Piccioto and Rockhill (1997) studied the efficiency of partial likelihood approximations in the presence of ties; Lam and Ip (2003) reported group-based modeling; and Yu et al. (2004) proposed models with a cure fraction for grouped survival data under the parametric approach in the absence of covariates. According to Heitjan (1991), inferential analyses have rarely been performed, particularly because of the difficulty of the computational calculation due to the fact that the response variable is changed into intervals. Therefore, this paper presents a modified regression model for grouped data in which the regression structure is modeled by using four link functions (logit, complementary log–log, log–log and probit). Also, the likelihood function is modified so as to include failing individuals, individuals at risk and censored individuals.

* Corresponding address: Departamento de Ciências Exatas, USP, Av. Pádua Dias 11, Caixa Postal 9, 13418-900 Piracicaba, São Paulo, Brazil. Tel.: +55 11 19 34294127; fax: +55 11 19 34294468.

E-mail address: edwin@esalq.usp.br (E.M.M. Ortega).

We consider a frequentist analysis of a regression model for grouped survival data. The inferential part was carried out using the asymptotic distribution of the maximum likelihood estimators, which, with a small sample, can present some difficulties. As an alternative to classic analysis, we explore the use of the jackknife estimator for regression models. In this case, it is not necessary to use the asymptotic distribution of the maximum likelihood estimators.

After modeling, it is important to check the assumptions in the model as well as to conduct a sensitivity study to detect influential or outlying observations that can distort the results. Numerous approaches have been proposed in the literature to detect influential or outlying observations. An efficient way to detect influential observations was proposed by Cook (1986). He suggested that more confidence can be placed in a model which is relatively stable under small modifications. The best-known perturbation schemes are based on case deletion, introduced by Cook (1977), in which the effect of completely removing cases from the analysis is studied. This reasoning forms the basis for our global influence introduced in Section 3.1, and in performing such an analysis, it will be possible to determine which subjects might be influential for the analysis (see, for example, Cook and Weisberg, 1982; Xie and Wei, 2007).

On the other hand, when using case deletion, all information from a single subject is deleted at once. Therefore, it is hard to tell whether that subject has any influence on a specific aspect of the model. A solution for the earlier problem can be found in a quite different paradigm, by taking a local influence approach where one again investigates how the results of an analysis change under small perturbations in the model and where these perturbations can be specific interpretations. Also, some authors have investigated the assessment of local influence in survival analysis models. For instance, Pettitt and Bin Daud (1989) investigated local influence in proportional hazard regression models; Escobar and Meeker (1992) adapted local influence methods to regression analysis with censoring; Ortega et al. (2003) considered the problem of assessing local influence in generalized log-gamma regression models with censored observations; Ortega et al. (2006) derived curvature calculations under various perturbation schemes in exponentiated Weibull regression models with censored data; Carrasco et al. (2008) investigated local influence in log-Weibull modified regression models with censored data; Silva et al. (2008) adapted global and local influence methods in log-Burr XII regression models with censored data; Ortega et al. (2009) investigated local influence in generalized log-gamma regression models with a cure fraction; and Hashimoto et al. (2010) derived the appropriate matrices for assessing local influences in the log-exponentiated Weibull regression model for interval-censored data. We have developed a similar methodology to detect influential subjects in regression models for grouped survival data.

The article is organized as follows. In Section 2, we describe the regression model for grouped survival data using one of the four link functions as well as the methods of estimation by maximum likelihood and jackknife employed in this study. In Section 3, we discuss the measurements used for analyzing sensitivity. In Section 4, we perform various simulation studies, and compare them to the performance of the four link functions, for different parameter settings, sample sizes and numbers of intervals. In Section 5, we present an application of the proposed method on a real data set. Finally, we present our main conclusions in Section 6.

2. Regression models for grouped survival data

T is defined as a random variable representing an individual's failure time and C as another random variable that is independent of T and represents the censoring time associated with this individual. The time observed for such individuals is $t = \min(T, C)$, and δ is a censoring indicator associated with the individual, with $\delta = 1$ if $T \leq C$ and $\delta = 0$ if $T > C$. If the data set observed, in addition to censoring, has an excessive number of ties, the observed times $t = \min(T, C)$ are grouped into k intervals to eliminate the ties and allow the intervals to be considered as discrete lifetimes so that discrete survival analysis techniques can be used. However, Chalita (1997) points out that the large number of intervals causes estimation bias. Consider $(T_i : 1 \leq i \leq n)$ to be a random sample of size n and let $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ be a p -dimensional vector of explanatory variables associated with the i th individual. Also consider that the lifetimes T_i are grouped into k mutually exclusive intervals, $I_j = [a_{j-1}, a_j)$, $j = 1, \dots, k$, where a_j are constructed using the methodology of life tables with $0 = a_0 < a_1 < \dots < a_k = \infty$, (Lawless, 2003). The probability of failure of the i th individual, conditional on the explanatory variable \mathbf{x}_i and survival at the beginning of the j th interval, is given by

$$p_j(\mathbf{x}_i) = P(T_i < a_j | T_i \geq a_{j-1}, \mathbf{x}_i). \quad (1)$$

In terms of the survival function $S(\cdot)$, expression (1) is written as

$$p_j(\mathbf{x}_i) = P(a_{j-1} \leq T_i < a_j | T_i \geq a_{j-1}, \mathbf{x}_i) = 1 - \frac{S(a_j | \mathbf{x}_i)}{S(a_{j-1} | \mathbf{x}_i)}. \quad (2)$$

The regression structure represented through probability $p_j(\mathbf{x}_i)$ can be modeled by using different link functions. In this study, in addition using the logit and complementary log–log links, which are usual in grouped survival data analysis, we propose use of the log–log and probit links, as specified below.

- Logit link:

$$p_j(\mathbf{x}_i) = \exp(\eta_{ij}) / [1 + \exp(\eta_{ij})]. \quad (3)$$

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