Accepted Manuscript

Optimal scaling for survival analysis with ordinal data

S.J.W. Willems, M. Fiocco, J.J. Meulman

PII: S0167-9473(17)30103-2
DOI: http://dx.doi.org/10.1016/j.csda.2017.05.008
Reference: COMSTA 6468

To appear in: *Computational Statistics and Data Analysis*

Received date : 28 February 2016
Revised date : 10 May 2017
Accepted date : 20 May 2017

Please cite this article as: Willems, S.J.W., et al., Optimal scaling for survival analysis with ordinal data. *Computational Statistics and Data Analysis* (2017), http://dx.doi.org/10.1016/j.csda.2017.05.008

Optimal Scaling for Survival Analysis with Ordinal Data

S.J.W. Willems[1], M. Fiocco[1,2], and J.J. Meulman[1]

[1]Mathematical Institute, Leiden University, Leiden, The Netherlands
[2]Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

May 10, 2017

Abstract

Medical and psychological studies often involve the collection and analysis of categorical data with nominal or ordinal category levels. Nominal categories have no ordering property, e.g. gender, with the two unordered covariates male and female. Ordinal category levels, however, have an ordering, e.g. when subjects are classified according to their education level, often categorised as low, medium or high education. When analysing survival data, currently two methods can be chosen to include ordinal covariates in the Cox proportional hazard model. Dummy covariates can be used to indicate category memberships, as is usually done for nominal covariates. Estimated parameters for each category indicate the increase or decrease in risk of experiencing the event of interest compared to the reference category. Since these parameters are estimated independently from each other, the ordering property of the categories is lost in the process. To keep the ordinal property, integer values can be given to the covariate's categories (e.g. low = 0, medium = 1, high = 2), and the variable is included in the model as a numeric covariate. However, since the ordinal data are now interpreted as numeric data, the property of equal distances between consecutive categories is introduced. This assumption is too strict for this data type; distances between consecutive categories do not necessarily have to be equal. A method is described to include ordinal data in the Cox model. The method implements optimal scaling to find optimal quantifications for the ordinal category levels. These quantifications are chosen such that they preserve the categories' ordering, and do not force equal distances between consecutive category levels. A simulation study is carried out to compare the performance of optimal scaling with the performance of the two currently used methods described above. Results show that the optimal scaling method increases the model fit if ordinal covariates are included in the model.

[1]Niels Bohrweg 1, 2333 CA Leiden, The Netherlands, s.j.w.willems@math.leidenuniv.nl, +31 71 527 7091