Stockholm University

Mathematical Statistics
Stockholm University
Master Thesis **2016:4**
http://www.math.su.se

Handling Ties in the Rank Ordered Logit Model Applied in Epidemiological Settings

Angeliki Maraki*

September 2016

Abstract

In epidemiological studies it is important to adjust for potential confounders when assessing the relationship between an explanatory variable and the outcome, in order to not obtain false, or miss true, statistically significant association between these two factors. Matching is a way to adjust for confounding having as a main advantage that it is not needed for the relationship between the outcome and the confounders to be specified, that way mis-specifications of the model can be avoided. In this Thesis, we consider continuous or ordinal outcomes and match away confounding using the Rank Ordered (RO) logit model by stratifying the cohort based on the confounders and ranking the outcomes within each stratum. When the underlying model is linear in parameters and the error terms have an Extreme Value Type I distribution the resulting likelihood is equivalent to the likelihood for the stratified Cox proportional hazards model. Consequently, the RO-logit model can be used for matching away possibly complex relationships between confounders and exposure/outcome by fitting stratified Cox-regressions. One challenge with the RO-logit model is ties. Similar to Cox-regression in survival analysis, the estimator assumes no ties in the outcome, but since the estimator has the same form as a stratified Cox-regression it is reasonable to assume that methodology for handling ties in survival analysis can be adopted in the RO-logit model. In this thesis we will investigate this by evaluating four methods for handling ties in Cox-regression, namely the Efron, Breslow, Discrete and Adding methods, by simulating scenarios with different degree of ties and different exposure and error distributions. We conclude that all four methods perform equivalently well, with the Adding method having a small advantage over the others. However, for some methods we found some bias. Moreover, we applied the RO-logit model on a data set from Maria Ungdom health clinic. The data set contains information about clients of the clinic, their family history, a selection of single nucleotide polymorphisms (SNP) and alcohol or drug abuse score of the individuals. The aim of the analysis was to identify significant relationships between some SNPs and alcohol or drug abuse score of the individuals participating in this study when the family history is adjusted for by matching on it. Significant associations between some of the SNPs and the alcohol and drug abuse score were detected. Nevertheless, some methods of handling ties in the model were biased and they are advised to be applied with caution.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: aggmara@gmail.com. Supervisor: Tom Britton.