
Introducing Survival and Event History Analysis

by Melinda Mills

London: Sage Publications 2011

ISBN 978-1848601024

Paperback, C\$58.95, 279 pp.

Reviewed by Md Kamrul Islam

Department of Sociology, University of Alberta
mdkamrul@ualberta.ca

In this book, Melinda Mills aims to introduce survival and event history analysis by covering a wide range of topics to non-specialists and specialists. What makes the book special is the wide range of topics covered—ranging from non-parametric to parametric methods. The book is most instructive for illuminating the use of popular statistical packages, such as R and Stata, for survival and event history analysis. Whereas most books on statistical analysis focus on either theoretical aspects or statistical features, this book intends to bridge the two. Considering the increasing interest in survival and event history analysis in the social sciences, and the availability of sophisticated statistical packages such as Stata and R, the timing of the book is just right.

The book is divided into 11 chapters. Chapter 1 commences with a brief introduction of the survival and event history analysis. Mills clearly defines the key concepts and terminologies, and explains the importance of studying survival and event history analysis. Then, in Chapter 2, the author provides excellent guidance on how to use the statistical package R—one of the most widely used data analysis software in recent years. In this chapter, the author focuses on the advantages and disadvantages of using different statistical programs for survival and event history analysis, shows how to download instructions of the base program R, explains three different approaches of using R, and provides empirical examples of using R to analyse survival data. The chapter is very well written and easy to follow.

The first difficulty that researchers in general and lay researchers in particular face is preparing datasets for analysis, a topic that has received limited attention in most books. Mills addresses the components of rearranging and preparing datasets for survival and event history analysis in a highly readable manner in Chapter 3. Three types of data are mostly used in survival and event history analysis: single-episode data, multi-episode data, and subject-period (discrete-time) data. The particular focus on converting single-episode data into multiple-episode data and transforming date variables into numeric variables is very useful for new researchers.

The starting point of survival and event history analysis after rearranging and preparing the dataset is to obtain the Kaplan-Meier survival estimates (KM) and plot them to get a first idea about the effect of exposure variable of interest on the outcome variable. Chapter 4 clearly illustrates the computing of KM survival estimates, plotting the KM survival curves, and comparing two KM survival curves. This chapter would guide researchers in selecting the most appropriate test in examining differences between two groups. The limitation of the KM survival estimates is that they fail to take other covariates into account, which can be done through the application of Cox proportional hazards (PH) regression. Chapter 5 explains why the Cox PH model is so popular as compared to other parametric models. The procedures for estimating and interpreting the Cox model, with both fixed and time-varying covariates, are clearly outlined.

Chapter 6 is devoted to specification of five parametric models: exponential, piecewise constant exponential, Weibull, log-logistic, and log-normal models. The chapter also contains a specific focus on the relationship between probability density function, hazard function, and survival function. Mills explains the estimation and interpretation of each of the parametric models, along with the relative advantage and disadvantage of using

these models. The author is careful to point out that parametric models directly estimate the hazard rate; hence, wrongly specified parametric models can lead to incorrect interpretation. It is important to ensure that all procedures of model building and diagnostics have been followed properly. This point is reinforced in other chapters, as well.

Chapter 7 explores the various steps of proper model building and diagnostics: selecting covariates; following a decision path to choose an appropriate model; assessing the overall goodness of fit of the selected model; and testing overall model adequacy. In this connection, Mills discusses Schoenfeld residuals to test the proportional hazards assumption. Finally, the author shows how to deal with non-proportional hazards through the introduction of interaction terms or by stratifying the data. Since model building and diagnostics are central aspects of quantitative analysis, researchers will find this chapter most useful thanks to the clarity of presentation.

Frailty models are gaining increasing importance as an advanced topic in survival and event history analysis. Use of frailty models helps to avoid the risk of overestimating or underestimating the effect of exposure variable(s) on the outcome variable. Five types of frailty models are covered in Chapter 8: shared, unshared, nested, joint, and additive frailty. The chapter also covers frailty model estimation using Gamma and Gaussian distributions.

Another popular approach in survival and event analysis is the application of discrete-time models. This type of model is predominantly useful when researchers prefer to use time as discrete (i.e., piece-wise) instead of continuous. While most of the books on survival analysis usually focus on continuous time to an event, in Chapter 9, Mills also illuminates the specification of the hazard, survival, and cumulative probability density functions of three types of discrete-time models: logit, probit, and complementary log-log. This chapter also details the advantages and disadvantages of the discrete-time models.

In the two subsequent chapters, Mills introduces advanced topics in survival analysis: competing risk and multi-state models (Chapter 10), and sequence analysis (Chapter 11). Chapter 10 focuses on the estimation and interpretation of three techniques of competing risk models: latent, cumulative incidence curve, and Lunn-McNiel approaches. In addition, Mills explores the estimation and application of three multi-state models: Markov, semi-Markov, and extended Markov models. Chapter 11 explains how to describe and visualize sequence datasets, how to measure similarities and distances between sequences, and how to produce typologies of trajectories in sequence analysis. The sequence analysis should generate particular interest among sociologists, because of its ability to depict various transition and trajectories of social events.

This book has many strengths. First, it provides clear guidelines on how to interpret findings obtained with the various methods through the use of empirical examples. This makes the book especially compelling for interested readers. Second, in each chapter the author cites numerous references for those who are interested in pursuing more detailed explanations. The book is an important resource for further studies in survival and event history analysis. Third, the author demonstrates R programs thoroughly for all analyses, starting from KM estimates in Chapter 4 to sequence analysis in Chapter 11. Mills' book should serve to motivate further interest among social science researchers to learn the R statistical package. Finally, by including an extensive appendix on Stata for survival and event history analysis, the author has made the book highly accessible to a large audience. In future editions, the author should expand the discussion on individual-level frailty. The book is appropriate for both graduate students and seasoned researchers.