Journal of Statistical Software

April 2009, Volume 30, Book Review 4. *http://www.jstatsoft.org/*

Reviewer: Joseph M. Hilbe
Jet Propulsion Laboratory, California Institute of Technology

---

Survival Analysis for Epidemiologic and Medical Research

Steve Selvin
Cambridge University Press, Cambridge, UK, 2008.
ISBN 978-0-521-71937-7. 282 pp. USD 39.99 (P).
http://www.cambridge.org/9780521719377

Introductory comments

Sevin's book is an excellent overview of survival analysis which can be used as a classroom text or as ancillary reading material for practicing health researchers.

There are larger more comprehensive texts available on survival analysis, but Selvin presents the subject in a concise manner, with the appropriate formulae and a clear presentation on how to understand them. Moreover, and of considerable importance, Selvin intersperses well-taken examples throughout the discussion.

Unless already familiar with the subject, the typical reader will likely have to pay carfeful attention to the logic of Selvin's presentation. There is not a lot of repetition, which I very much appreciate. The examples are not contrived, but will likely engage the readers interest, and make the points being expressed in the discussion come alive.

Given that R is currently being used at many universities, Selvin adds a final section titled "Examples of R". It is not presented as an appendix, which does not appear in the text, but is such for all practical purposes. In this section Selvin provides a description and example code for four R functions: `survfit`, `survreg`, `survdiff`, and `coxph`. Subsequently, he gives the reader five fully worked-out examples using real data including estimation using exponential and Weibull regression, exponential and Weibull regression for a two-sample analysis, a proportional hazards model, residual analysis for a proportional hazards model, and simulation code for a weibull distribution with censored observations. I found this part of the book to be very well done, and essential reading for anyone wanting to use R for analyzing survival data.

The text is divided into nine chapters, for each of which I provide a brief description in this review. Selvin states as a key element of the text to present the subject emphasizing the statisical notions of bias, confounding, independence, and iteraction. He does this well. Note from glancing at the chapter titles below that the standard items found in most books on this subject are present.

Chapter 1: Rates and properties

Selvin discusses the nature and general formulae for survival function, its derivative, and how this leads to the instantaneous relative rate, or hazard ratio. The discussion progresses to understanding the concepts and formulae for average rate, mean time at risk. Other concepts are developed as well, including the biomial distribution, normal approximation, logistic transformation, homo- and heterogeneity, and survival probabilities. The relationships of these statistics are rather fully explained, and augmented with graphs and good examples. Confidence intevals and their calculation are also introduced in this initial chapter.

Chapter 2: Life tables

Selvin presents a step-by-step algorithm for constucting life tables. A worked out life table is developed from given data, and appropriate graphs developed. Smoothing, moving averages, and dealing with follow-up data are also examined. The formulae for all procedures are given and explained in detail

Chapter 3: Two especially useful estimation tools

This chapter is one of the shortest in the book, but important for understanding the estimation process. I have always added an estimation methods chapter to my books, and found that most readers appreciated it. Students generally want to pass this discussion when it is first addressed, but typically appreciate it later.

The author first discusses maximum-likelihood estimation, and presents four key properties of maximum-likelihood estimates that should be remembered when using this method. He then addresses likelihood in general, examining the likelihood ratio statistic and the reason for using the log-likelihood in estimation rather than the likelihood. A couple of very well presented examples caps off the first section. He concludes the chapter by giving an example of the Poisson distribution, demonstrating how to calculate its mean and variance functions, and how it differs from the binomial distribution.

Chapter 4: Product-limit estimation

This chapter is devoted to the Kaplan-Meier product-limit method of describing and graphing survival data. Selvin only mentions in passing that the product-limit method is also called the Kaplan-Meier estimate of the survival probability. I found this a bit unusual since most texts tend to relate the method to the names of its originators. Selvin mentions it once.

Aside from the above, which is merely an observation, not a criticism, Selvin uses real data to show step-by-step now the product-limit method works. Greenwood confidence intervals are also examined. Also discussed, among other relevant statistics, is the cumulative hazard function and its relationship to the mortality function.

Chapter 5: Exponential survival time probability distribution

Parametric survival models are introduced in this chapter. Selvin carefully explains the nature of e, and the exponential function. He follows this up with an exhaustive or thorough

discussion of the exponential survival model, together with its strengths and obvious limitations. Fully worked-out examples are presented. I thought that this was an excellent chapter. I would have liked to have seen the author demostrate how a Poisson regression model can be constructed so that it estimates an exponential model with right censoring, and how a gamma regression with a scale of one can be used as an exponential regression, but without censoring. These comparisons may be added by the instructor, and an exclusion of this discussion is not a problem with the book. Rather, I have found the relationships to be interesting to students, and gives a connection between survival models and generalized linear models.

Chapter 6: Weibull survival time probability distributions

Weibull regression is the standard extension of the exponential survival model. As done for the exponential model, the author gives the reader a solid overview of procedure. Graphs and their rationale and interpretation are presented.

Chapter 7: Analysis of two-sample survival data

Several two-sample analyses are preformed using real data. Each is fully examined. The first analysis is the log-rank test, the second an exponential proportional hazards model, and third a Weibull proportional hazards model. Selvin shows how to calculate the relevant statistics, e.g., the baseline hazard function, mean survial times, and so forth. The chapter is meant to be an introduction to the next chapter.

Chapter 8: General hazards model: Parametric

In this chapter Selvin disusses parametric modeling in general, together with goodness-of-fit statistics, residuals; e.g., Cox-Snell residuals and "modified" Cox-Snell residuals, and so forth. An example is given in which these statistics and residuals are used.

The text does not address the traditional parameteric survival models such as gamma, log-logistic, and log-normal. Many survival model texts do not address these models either, so Selvin is not alone. The fact is that these models are rarely used, and their ommission from the book is not a disservice to the reader. If an instructor wants to include these models into a survival models course, there are several very good sources to draw on. Again, I do not find this to be an essential shortcoming. However, it should be mentioned.

Chapter 9: General hazards model: Nonparametric

Most reseachers consider survival analysis to consist of the product-limit method and log-rank test, together with the Cox proportional-hazards model. The Cox model being thought of is a nonparametric method, which is the subject of this final chapter. Selvin gives a solid presentation of the model, its model fit statistics, and residual analysis.

Concluding comments

This is a nicely done presentation on survival models. The reader will learn a great deal

of practical knowledge on how to model and interpret survival data. The book uses many examples to help solidify the concepts in the readers mind, but does not show program code in the main text. I tend to prefer seeing code and immediately applying the methods I am learning about to my own example data. Actually, students in my classes want to duplicate the code given in a book to make sure that they understand the concepts being learned. They then apply the code to different data until they feel confident that they have mastered – relatively so – the material. For them, the important issue is being able to construct, fit, interpret, and evaluate models being learned in a course.

The section on R in Selvin's book helps meet the above student desire. However, it seems that more books are being written in the manner I discribe above. Selvin does an outstanding job in explaning the logic of survival models, and their associated statistics, but many instructors will likely find that they will need additional materials to give their students so that they gain more working experience with the models being discussed.

My comments should in no way take away from Selvin's text. It is an excellent presentation of the subject.

Reviewer:

Joseph Hilbe
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109, United States of America
E-mail: jhilbe@aol.com
URL: http://en.wikipedia.org/wiki/Joseph_Hilbe

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 30, Book Review 4 *Published:* 2009-04-27
April 2009