



Journal of Statistical Software

March 2005, Volume 12, Book Review 1.

<http://www.jstatsoft.org/>

Reviewer: Thomas Lumley
University of Washington

An Introduction to Survival Analysis using Stata (Revised Edition)

Mario A. Cleves, William W. Gould, Roberto G. Gutierrez
Stata Press, College Station, TX, 2004.
ISBN 1-881228-84-3. 308 pp. \$52.00 (P).

The authors note in the Preface to this book that “[it] grew out of software, but nonetheless, it is not a manual”. *An Introduction to Survival Analysis using Stata* is certainly focused on the **Stata** implementation of survival analysis. It is published by the publishers of **Stata**, and written by the authors of the **Stata** survival analysis commands. There are many advantages in having data analysis methodology presented with concrete examples, by people who understand what computations are going on. There are also risks: the temptation to ignore or sidestep infelicities in implementation, and to overemphasize computational points that are of no importance to the typical user. Fortunately, the **Stata** implementation of survival analysis is well thought out, and a description of what **Stata** can do is a reasonable introduction to the field of survival analysis.

The book assumes some familiarity with **Stata** and with statistics. Epidemiologists with a background in logistic regression modeling would be one example of a suitable audience. It begins by describing hazard and cumulative hazard, truncation and censoring. This leads to a discussion of the possibility of different time scales for a survival analysis, and the manifold options for specifying survival data structures in the `-stset-` command. While covering all this material in one place saves space, it is perhaps a little confusing for someone without survival analysis experience — a course in survival analysis would usually start with simple right censoring and defer all the complications for a few weeks. After discussing one- and two-sample summaries and tests the book moves on to the Cox model, where it covers fitting, time-varying covariates, and shared frailty models. The final chapters cover similar topics for parametric proportional hazards and accelerated failure models.

The explanations of statistical methods and issues in the book are usually good. There are a few things I would quibble with, such as the characterization of marginal hazard ratios as ‘biased’ in the presence of correlation — an issue that some of these same authors have handled more carefully in discussions of longitudinal data analysis. Perhaps the most significant complaint is that the term ‘at risk’ is used for times at which a person might have had an event, but when such an event would not have been observed, and the term ‘under observation’ is used for times when an event would have been observed. In both epidemiology and the

mathematical parts of survival analysis the term ‘at risk’ is more often used to mean ‘able to have an event observed’. The authors’ terminology is arguably more logical, but as they note in their introduction “if we are going to communicate by the written word, we have to agree on what these terms mean.”

An obvious comparison of this book is to ?, whose first author wrote the very similar survival analysis implementation in S-PLUS and R. The book of ? is tied less tightly to the implementation, and covers rather more of the statistical theory. Perhaps a fair summary of the difference is that ? have written the sort of book that would be appreciated by people who use S, and Cleves *et al.* have written for the sort of audience that uses Stata.

References

Therneau TM, Grambsch PM (2000). *Modelling Survival Data: Extending the Cox Model*. Springer, New York.

Reviewer:

Thomas Lumley
University of Washington
Department of Biostatistics
Box 357232
Seattle, WA 98195-7232, United States of America
E-mail: tlumley@u.washington.edu