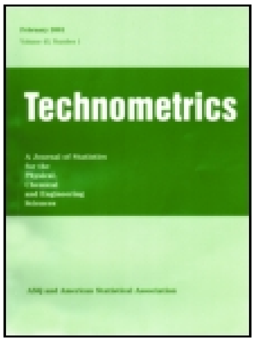

Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

Statistics With Applications in Biology and Geology

James T Wassell

To cite this article: James T Wassell (2004) Statistics With Applications in Biology and Geology, Technometrics, 46:1, 111-111, DOI: 10.1198/tech.2004.s739

To link to this article: https://doi.org/10.1198/tech.2004.s739

Published online: 01 Jan 2012.

Statistics With Applications in Biology and Geology, by Preben BLÆSILD and Jørgen GRANFELDT, Boca Raton, FL: Chapman & Hall/CRC, 2003, ISBN 1-58488-309-X, ix + 555 pp., $59.95 (softcover).

If teaching statistics by example is your goal, then this book will provide a wealth of interesting examples that demonstrate the methods of statistics and practical data analysis. The focus of the book is biology and geology, as the title indicates, with a greater emphasis on biology. There are about 70 separate datasets in the examples and exercises, providing a great deal of material for teaching. About one-third of the examples and exercises are based on problems in geology, including sediment composition, sound propagation in rock, metal content of geologic samples, sand sifting rates, pH of core samples, shape and weight of stones, magnetism in lava flows, gas diffusion in stones, mineral content in water, earthquakes, sediment transport, wind direction, and orientation of crystals. The biological exercises and examples encompass problems related to fisheries taxonomy, vertebrate physiology, invertebrate population density and size distribution, ecology, toxicology, both plant and animal genetics, agronomy, microbiology, and human studies involving asthma, twin studies, cancer risks, chronic disease, and physiology.

The authors provide the data on their website along with SAS programs used for analysis. The book devotes more pages to displaying SAS program code than I enjoy reading, and much of this is redundant with the website, but the duplication of material may be helpful for students. The first example of a SAS program includes a SAS macro and programming using PROC IML, a matrix programming language within SAS. This is much too advanced for a student's first encounter with SAS, and is a challenge to even experienced SAS programmers. However, the website is well organized and enables the user to easily locate the programs for the specific examples.

The book comprises 12 chapters, beginning with a discussion of statistical models and inference in Chapter 1. Chapter 2 emphasizes graphics including histograms and probability plots. Chapter 3 discusses normally distributed data procedures, including one- and two-sample tests, one-way analysis of variance, and simple regression. Chapter 4 presents two-way analysis of variance with interaction as an example that extends the simple linear models. Chapter 5 discusses the concepts of power, noncentral distributions, and sample size efficiency through experimental design. Chapter 6 introduces correlation and the bivariate normal distribution. Chapter 7 presents the multinomial distribution and statistical methods for categorical data, and Chapter 8 introduces the Poisson distribution for analysis of rate data. Chapter 9 introduces Poisson and logistic regression as part of a brief discussion of generalized linear models. Chapter 10 presents methods for analyzing directional data in two and three dimensions using circular normal distributions (von Mises and Fisher distributions). Although concepts of maximum likelihood are briefly mentioned in various parts of the book, Chapter 11, "The Likelihood Method," offers more details and a discussion of quadratic approximation. Chapter 12 introduces some nonparametric tests for one-sample, two-sample, and k-sample Kruskal–Wallis tests.

Unfortunately, my copy of the book had too many printing errors, especially in the mathematical notation, which would cause confusion for students seeing this material for the first time. Some of the discussion assumes a background that is too advanced for the intended audience of undergraduate students in biology and geology. Discussions of subspaces, projections about the geometry of linear models, and the distinctions between "affine subspace" and "linear subspace" in the section on generalized linear models seem to be a distraction from the theme of practical data analysis in biology and geology. The book's strength lies in the emphasis on practical examples and realistic research problems that include many interesting datasets. The analyses are presented in a way to facilitate teaching the principles of statistics and the importance of quantitative studies to students in the natural sciences.

James T. WASSELL
National Institute for Occupational Safety and Health
Centers for Disease Control and Prevention

Survival Analysis (2nd ed.), by John P. KLEIN and Melvin L. MOESCHBERGER, New York: Springer-Verlag, 2003, ISBN 0-387-95399-X, xv + 536 pp., $89.95.

Comprising 13 chapters and 5 appendixes, with 97 illustrations and several exercises at the end of each chapter, this book is an excellent graduate-level text for a course in survival analysis. Students will definitely find the authors' systematic treatment of topics, clear discussions and derivations, and numerous detailed examples useful. This book is also a good reference source for practicing statisticians, biostatisticians, and public health professionals with a basic statistics and applied statistics background. Although the examples are biomedical in nature, most methods described in the book for time-to-event data are applicable to other fields, including engineering and economics, and the book should be useful for researchers in these disciplines. The authors use semiparametric and nonparametric methods extensively, and also discuss parametric models. The "Practical Notes" and "Theoretical Notes" provided in many sections are very attractive and give readers information and citations beyond the material in the text.

Chapter 1 introduces 19 datasets that are used to illustrate various aspects of survival analysis throughout the text. These datasets pertain to biomedical or public health examples. As indicated in the Preface, these are available at the authors' website, accessible through the Springer website (*http://www.springer-ny.com*) or the first author's website, (*http://www.biostat.mcw.edu/homepgs/klein/book.html*), which also provides the book's outline, errata, and some SAS macros.

Chapter 2 defines and illustrates basic parameters related to time-to-event data, including the survival function, the hazard function, the density function, cumulative distribution function, and the mean and median residual life at a given time. Table 2.2 summarizes these parameters for 11 common parametric families. A discussion of summary statistics for competing risk probabilities is also included; these quantities are estimated in Chapter 4.

Chapter 3 describes censoring (right, left, and interval) and a common approach for the construction of the corresponding likelihood function. This chapter also defines truncation (left and right) and discusses the use of a conditional distribution for constructing the likelihood function. The counting process approach for construction of the likelihood with truncated or censored survival data is briefly described in the last section.

Chapter 4 describes nonparametric estimation of the cumulative hazard function and survival function for right-censored data. The product-limit estimator and associated confidence intervals are described and illustrated, and a confidence band for the survival function is also discussed. The chapter ends with a section on estimation of the survival function for right-censored and left-truncated survival data. Chapter 5 presents estimation of the survival function under other schemes, including left censoring, double censoring, interval censoring, right-truncation, and grouped data.

Chapter 6 explores a few topics on estimation for univariate survival data. Kernel methods for estimating the hazard function are explained in Section 6.2. The next section summarizes estimation of excess mortality. Section 6.4 discusses Bayesian nonparametric estimation, incorporating MCMC methods.

Chapter 7 takes a detailed look at hypothesis testing for hazards, starting with the simplest single population case and going on to tests for two or more populations, tests for trend (via an ordered alternative), stratified tests (on levels of covariates), and Renyi-type tests (which are powerful in detecting crossing hazards). A new section on testing the equality of survival curves at a fixed time point is an attractive inclusion in the second edition.

Chapter 8 discusses the Cox proportional hazards model with fixed covariates (discrete or continuous). The partial likelihood is clearly explained and is illustrated with examples. Although this chapter generally treats the baseline hazard function as a nuisance parameter, the last section is devoted to estimation of the survival function via a variant of Breslow's estimator of the baseline cumulative hazard. Extensions to the proportional hazards model, such as inclusion of time-dependent covariates, stratified proportional hazards model, left-truncation, and a multistate model for handling time-varying effects, are discussed. Using several examples, Chapter 11 explains model assessment via Cox–Snell residuals, martingale residuals, deviance residuals, and influence diagnostics. Graphical methods for assessing the proportionality assumption are also presented.

Chapter 10 describes two additive hazards regression models. First, Aalen's nonparametric additive model for the conditional hazard given covariates is discussed in detail, allowing the regression coefficients to be functions of time.