This section will review those books whose content and level reflect the general editorial policy of Technometrics. Publishers should send books for review to Eric R. Ziegel, BP Naperville Complex, Mail Code C-7, 150 West Warrenville Road, Naperville, IL 60563-8460 (ziegeler@bp.com). The opinions expressed in this section are those of the reviewers. These opinions do not represent positions of the reviewer’s organization and may not reflect those of the editors or the sponsoring societies. Listed prices reflect information provided by the publisher and may not be current. The book purchase programs of the American Society for Quality can provide some of these books at reduced prices for members. For information, contact the American Society for Quality, 1-800-248-1946.

Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications
Wayne B. Nelson William Q. Meeker 263

Statistical Models and Methods for Lifetime Data
Jerald F. Lawless Gordon Johnston 264

The Statistical Analysis of Failure Time Data
John D. Kalbfleisch and Ross L. Prentice Laurence L. George 265

The Mahalanobis–Taguchi Strategy
Genichi Taguchi and Rajesh Jugulum Wei Jiang 266

The Elements of Statistical Learning
Trevor Hastie, Robert Tibshirani, and Jerome Friedman Eric R. Ziegel 267

Elements of Computational Statistics
James E. Gentle William J. Owen 268

Smoothing Spline ANOVA Models
Chong Gu Michael Frey 269

Growth Curve Models and Statistical Diagnostics
Jian-Xin Pan and Kai-Tai Fang Andrew M. Kuhn 270

Elements of Applied Stochastic Processes
U. Narayan Bhat and Gregory K. Miller Errol Caby 270

Testing Statistical Hypotheses of Equivalence
Stefan Wellek Nicole Lazar 271

Causation, Prediction, and Search
Peter Spirtes, Clark Glymour, and Richard Scheines Tom Burr 272

The Six Sigma Journey From Art to Science: A Business Novel
Larry Walters George S. Kalachkarian 273

Case Studies in Bayesian Statistics, Vol. VI
Constantine Gatsonis, Robert E. Kass, Alicia Carriquiry, Andrew Gelman, David Higdon, Donna K. Pauler, and Isabella Verdinelli (editors) Kenny Q. Ye 273

Introductory Statistics With R
Peter Dalgaard Thomas D. Sandry 274

Applied Statistics With Microsoft Excel®
Gerald Keller Thomas H. Short 275

Editor Reports on New Editions, Proceedings, Collections, and Other Books

Statistical Design and Analysis of Experiments
Robert L. Mason, Richard F. Gunst, and James L. Hess 275

Probability, Statistics, and Reliability for Engineers and Scientists
Bilal M. Ayyub and Richard H. McCuen 276

Principal Component Analysis
I. T. Jolliffe 276

Case Studies in Reliability and Maintenance
Wallace R. Blischke and D. N. Prabhakar Murthy (editors) 276

Data Mining: Concepts, Models, Methods, and Algorithms
Mehmed Kantardzic 277

Introduction to Survey Quality
Paul P. Biemer and Lars E. Lyberg 277

Models for Investors in Real World Markets
James R. Thompson, Edward E. Williams, and M. Chapman Findlay, III 277

Six Sigma Quality for Business & Manufacture
M. Joseph Gordon, Jr. 278

The Power of Ultimate Six Sigma™
Keki R. Bhote 278

Six Sigma and Beyond: Statistics and Probability
D. H. Stamatis 279
Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications, by Wayne B. Nelson, Philadelphia, PA: ASA-SIAM, 2003, ISBN 0-89871-522-9, 151 pp., $85.00.

Recurrence count data arise when an observational unit or group of units is monitored over time and the times of a particular event or class of events are recorded. Examples include counts of maintenance actions on repairable systems, transactions with customers, recurrences of a disease, and the birth of children. More generally, each event may have an associated “value” (e.g., the cost of a maintenance action or the size of an order). Questions of interest include the behavior of the recurrence rate (or cost accumulation rate) as a function of time (i.e., whether the rate is increasing, decreasing, or constant over time). The population mean cumulative number of events (or mean cumulative cost) per unit at a particular time (e.g., at the end of the warranty period) is also of primary interest. Such a cumulative function is called, generically, a mean cumulative function (MCF). The need to compare MCFs (e.g., for units manufactured in different manufacturing periods or individuals receiving different treatments for a disease) is also common.

This is Wayne Nelson’s third book-length contribution to the statistical/engineering literature. Like its predecessors, this book provides a timely, self-contained treatment of an important area of application. All of the material is carefully motivated by actual applications. The examples consist of an interesting mixture of applications from engineering, medical science, and other areas. Despite the fact that some of the material is technically difficult, the writing style is crystal clear.

This book will appeal to statisticians, engineers, biomedical professionals, and other scientists who face the task of extracting and presenting information from recurrence data. Each chapter concludes with a collection of exercises; thus the book could also serve as a supplementary text for a course in biostatistics, reliability data analysis, reliability engineering, or a more general course on event time data analysis. The book will also be an important reference for those wanting to do research on the development of methods for recurrence data analysis.

Recurrence data have been around for a long time, as demonstrated by Nelson’s examples. However, such data have received little focus in practice, because there has not been a flexible, robust, consistent set of methods to analyze them (at least not in commercially available software). The use of the methods described in this book will escalate in the near future both because of this book itself and the fact that interest in such methods has sparked development of commercial software to do the analyses. The major statistical packages (e.g., Minitab 2000, JMP 2002, SAS 1999, S-PLUS 2002, and the SPLIDA 2003 add-on to S-PLUS) all have recently added or are planning to add capabilities to do some or all of the analyses contained in this book. Nelson surveys the capabilities and provides examples of the outputs for most of these packages.

This book comprises the following chapters:

1. Recurrent Events Data and Applications
2. Population Model, MCF, and Basic Concepts
3. MCF Estimates for Exact Age Data
4. MCF Confidence Limits for Exact Age Data
5. MCF Estimate and Limits for Interval Age Data
6. Analysis of a Mix of Events
7. Comparison of Samples
8. Survey of Related Topics.

Each chapter begins with a stated purpose and overview and concludes with problems on analyzing actual data and extending the methods.

Chapter 1 introduces recurrence data with six applications, including transmissions in automobiles, bladder tumors in patients, and births of children to statisticians (with an interesting comparison of the differences between male and female statisticians). Each application has a description of the motivating problem, the structure of the data, information sought, and a graphical (event plot) display of the data. This chapter also introduces and explains some of the important basic concepts and practical issues that need to be considered when attacking a real problem. These include such issues as definition of time zero, choice of an appropriate time scale (e.g., whether time is defined as the amount of calendar time or the amount of use that a units has seen), and how to view types of events. These discussions allow us to benefit from Nelson’s vast experience in working with the sticky, but important, definitional issues that arise with real problems involving data collection, analysis, and interpretation.

Chapter 2 describes the population stochastic model and its MCF, which contains most of the information sought from recurrence data. The MCF is the population mean cumulative number or cost of events, across all of the units in a population, as a function of time. This flexible nonparametric model requires no assumptions about the form of the MCF or about homogeneity or independence of population units. Although the MCF increases in most applications, this is not a requirement of the model (e.g., the response at any given point in time could be negative, as in the return of merchandise). This flexibility alone greatly increases the range of potential applications. Moreover, this model is more realistic than the well-known nonhomogeneous Poisson process (NHPP) model. There is also useful discussion that contrasts this recurrence model with the more commonly used life distribution model for survival data. Often there has been confusion between the hazard rate function of a life distribution and the recurrence rate function of a counting process. Although both have been referred to as “failure rate,” they are not the same, and thus many avoid using the term “failure rate” in favor of less ambiguous terms.

Chapter 3 develops the basic method for estimating the MCF in the presence of multiple censoring, which arises when different units are under observation for different amounts of time. This chapter presents MCF plots produced by the various statistical packages. Plots of the sample MCF provide insight into the behavior of a recurrence process and answers to most of the questions of interest.

Computing an estimate of the MCF would be trivial if all units were under observation for the same length of time, but such situations are rare in practice, due to staggered entry or “time” scales that depend on use, such as hours of service for a jet engine or miles driven for an automobile. The MCF estimation method also allows for left censoring and gaps in the observation of units. The MCF estimator presented in this chapter (first presented in Nelson 1988) is reminiscent of Nelson’s cumulative-hazard-based estimate of the fraction failing as a function of time for population nonrepairable units when the failure-time data are multiply censored (Nelson 1969). That is, estimates of the incremental change in the MCF can be obtained directly (based on the known number of units that are “at risk” for an event), even in the presence of censoring. These then are accumulated into an estimate of the MCF.

Chapter 4 builds on the technical presentation in Chapter 3, developing expressions for the variance of the MCF estimator and an estimator for this variance. A variance estimate is then used to construct approximate confidence intervals for the population MCF. Such intervals should generally be presented in the plot of the sample MCF and are available in the statistical packages mentioned earlier.

Nelson compares his unbiased variance estimator with a “naïve” estimator based on the assumption of independent increments, a property of the commonly used parametric NHPP model (but not required in Nelson’s setup). The assumption of independent increments is, however, often inappropriate in applications. Nelson notes that in the usual case where population increments are positively correlated, use of the naïve variance estimator yields confidence intervals for the MCF that are too narrow. Such positive correlation arises in, for example, the common situation when samples are from a nonhomogeneous population. In Nelson’s approach, robustness is achieved by estimating the variances and covariances of the increments with simple moment statistics. These moment estimators are then combined with the usual formulas for the variance of a sum of dependent random variables. Lawless and Nadeau (1995) provided a similar robust estimator for the MCF, which, although biased, is guaranteed to be positive. Nelson’s unbiased estimator has a positive, albeit small, probability of being negative.

Chapter 5 extends the methods in Chapters 3 and 4 to provide MCF estimates and confidence intervals for interval age data. Such data are common and arise when data have been summarized into counts (or costs) of events that have occurred in given time intervals. For example, reporting the number of events in each month of service is often used to summarize product warranty data. When intervals are small, and when event histories are available for all units, the exact methods in Chapters 3 and 4 can be used directly. When, for example, unit
histories are grouped to give monthly counts of events, the information needed to compute Nelson’s variance estimate for the sample MCF, described in Chapter 4, has been lost. Instead, Nelson describes and illustrates how to compute a simple “naive” variance estimate and confidence intervals that would be correct under the assumption of an underlying nonhomogeneous Poisson process model.

Chapter 6 describes methods for analyzing recurrence data when events can be divided into categories. Examples include different failure modes for a system and gender of a child born. The basic underlying model is that each category has its own population MCF. Under weak assumptions, the MCFs for individual event types can be added to give the MCF for a specified set of event types. Knowing the MCF for each category would allow one to answer questions about the MCF for all types of events combined, for a particular type of event, and for any chosen subset of the event types. Each of these questions is illustrated with data on failures of traction motors for subway cars. For example, a reliability analyst can estimate the MCF for a system under the assumption that one or more failure modes can be eliminated.

As Nelson points out, the MCF estimators presented in this chapter do not require independence of the underlying stochastic processes that generate the different kinds of events. This is in contrast to the life data competing-risk model where the assumption of independence is critical for making inferences about the effect of removing a failure mode (see, e.g., chap. 5 of Nelson 1982). For the MCF model, there is, however, the tacit assumption that eliminating a failure mode will not affect the MCF functions of the other failure modes.

Although the methods based on the nonparametric recurrence stochastic process model are versatile and require minimal assumptions, in more complicated situations they cannot be applied without careful thought. Consider, for example, a repairable system that has a replaceable unit with two failure modes (A and B), both of which are caused by a common mechanism (e.g., corrosion). Due to the common cause, the times to failure for the two components are highly correlated. When the component fails from either A or B, it is replaced, censoring the other mode. If this censoring of the other mode is naively overlooked, and an engineering change is made to eliminate one of the failure modes (say A), then only the symptom has been fixed and there will be a corresponding increase in the MCF for mode B. Then looking at the past data for the occurrence of mode B to estimate the MCF of mode B alone would be misleading.

Chapter 7 presents methods for comparing sample MCFs to see whether they differ statistically. The methods are based on an estimate of the difference between two MCFs of processes to be compared and confidence intervals computed for this difference. The methods are illustrated by comparing treatments for recurrent bladder tumors and replacements of two different batches of locomotive breaking grids. Both pointwise and simultaneous (over time) comparisons are described. There is also brief discussion of multiple comparison methods that would be needed if more than two MCFs are to be compared.

Chapter 8 provides a very useful survey of other topics closely related to the methods presented in the main part of the book. Topics described include Poisson process and nonhomogeneous Poisson process models, renewal process models, models with covariates, and other models.

As with Wayne Nelson’s other books, this book contains a valuable collection of interesting actual applications and corresponding data that we can expect to be used in future publications by other people doing research in this area. An Excel workbook that contains all of the data in the book is available at http://www.siam.org/books/sa10.

In summary, this is an important and interesting book from which most statisticians and many others who analyze data will derive benefit. As the easy-to-use tools described here become more commonly known, I predict that the methods will be much more widely used.

William Q. Meeker
Iowa State University

REFERENCES

Lawless, J. F. and Nadeau, C. (1995), “Some Simple Robust Methods for the Analysis of Recurrent Events,” Technometrics, 37, 158–168.

Minitab (2000), Minitab User’s Guide 2: Data Analysis and Quality Tools, Release 13, State College, PA: Minitab.

Nelson, W. (1969), “Hazard Plotting for Incomplete Failure Data,” Journal of Quality Technology, 1, 27–52.

(1988), “Graphical Analysis of System Repair Data,” Journal of Quality Technology, 20, 24–35.

SAS Institute, Inc. (1999), SAS/QC User’s Guide for Release 8, Cary, NC: SAS.

(2002), JMP User’s Guide, Version 3, Cary, NC: SAS.

SPLIDA (2003), “S-PLUS Life Data Analysis,” available at www.public.iastate.edu/~splida.

Insightful, Inc. (2002), S-PLUS User’s Manual, Version 6.1, Seattle, WA: Insightful.

Statistical Models and Methods for Lifetime Data (2nd ed.), by Jerald F. Lawless, Hoboken, NJ: Wiley, 2003, ISBN 0-471-37215-3, xx + 630 pp., $89.95.

Statistical modeling and analysis of data on time to occurrence of an event of interest is of great importance in product reliability, a subject likely of interest to many Technometrics readers. This book is not concerned exclusively with reliability data analysis, although it gives many examples from product reliability. Rather, it strives to provide a “comprehensive account of models and methods for lifetime data” for use in a wide variety of fields. The first edition of this book is one of a few texts that I use consistently as a reference for lifetime data modeling. This second edition is expanded and updated with recent research and will be a valuable reference for anyone concerned with lifetime data analysis. The book could also serve as a text in a graduate-level course in statistical modeling of lifetime data. It is clearly written, concise, and contains numerous references to material beyond its scope.

Most chapters contain several examples from both the engineering and biomedical sciences. The datasets for the examples are not available at a single website; rather, website references are given for the larger datasets, and smaller datasets are listed in the text. Brief computational notes at the end of each chapter describe software that can be used to perform the analyses described in the chapter. Most of the examples were prepared using S-PLUS software, but the programs are not available to readers. It would be useful to make the programs and data for the examples available from a single website, as done by, for example, Meeker and Escobar (1998). By doing so, Lawless would do readers a service by showing them not only how to use the software to perform routine tasks, such as fitting the parametric regression models of Chapter 6, but also how to perform more advanced and less-obvious analyses, such as the multivariate modeling of Chapter 11. Another useful feature would be to show readers how to perform some of the analyses using other popular software, such as SAS and STATA.

Chapter 1, “Basic Concepts and Models,” introduces several examples used in later chapters. These examples introduce the concept of censored data. Basic concepts for continuous-lifetime data models, such as the hazard and the survivor functions, are introduced, and statistical distributions commonly used to model lifetime data are defined. Less commonly used distributions, such as those with piecewise constant or polynomial hazard functions, are also briefly discussed. Mixture models and regression models are introduced, as are models with multiple modes of failure.

Chapter 2, “Observation Schemes, Censoring, and Likelihood,” discusses likelihood-based methodology as a unifying principle for inference for right-censored, interval-censored, and truncated discrete- and continuous-lifetime data. Counting process notation is introduced in sufficient detail to formulate a model for general censoring processes, but the reader need not have detailed knowledge of counting process theory to understand most of this chapter or later chapters.

Chapter 3, “Some Nonparametric and Graphical Procedures,” presents basic nonparametric estimates of survival probabilities and quantiles. The Kaplan–Meier estimator of the survivor function and the Nelson–Aalen estimator of the cumulative hazard function are analyzed in detail. Estimation of the survivor function when the data are truncated or interval censored are discussed. Construction and interpretation of P–P plots, probability plots, and cumulative hazard plots are presented. Estimation of survival probabilities by life table methods is discussed in some detail.

Chapter 4, “Inference Procedures for Parametric Models,” presents likelihood-based inference for certain parametric lifetime models. The exponential distribution is discussed extensively. The gamma, inverse Gaussian, and models with polynomial hazard functions are discussed more briefly. Inference for interval-censored or truncated data is discussed in general terms. Mixture models and models with threshold parameters are discussed. Finally, the computation and interpretation of prediction intervals are presented.