

Book Review

Fleiss, J, Levin, B, Paik, MC 2003: *Statistical methods for rates and proportions* (third edition). Hoboken, New Jersey: Wiley Series in Probability and Statistics, John Wiley & Sons, Inc. 704 pages, \$99.95 (hardback) ISBN 0471526290.

The second edition of this book was published in 1980 (the first edition in 1972). Joseph Fleiss, Bruce Levin and Myunghee Cho Paik have taken up the challenge to enhance a classic book with the latest methods available for discrete data. As Drs Levin and Paik note, this task was trying not just because of the wealth of methods that have been introduced with the boom in computing technology, but because of the illness and eventual loss of the primary author, Levin and Paik have retained all of the original chapters from the second edition, and introduced six new ones (one new chapter was formerly a section in the second edition). The first and second editions were aimed at “research workers and students” and assumed a mathematical knowledge of basic algebra (and the ability to take logarithms and extract square roots). This edition has introduced several sections that require more mathematical sophistication, including matrix algebra and mathematical statistics. These sections are identified with an asterisk. Additionally, the methods described in these new sections require more than a pocket calculator to apply. The book is still a “must have” reference for the researchers working with discrete data. Furthermore, the new edition includes many recent advances and expanded problem sets to make it a valuable textbook in a biostatistics or statistics program.

The following paragraphs highlight some of the new sections/chapters, but are not an exhaustive list. Chapter 2 expands a section from the second edition and covers inference about a single proportion. Exact inference and confidence intervals are included, as well a section using the delta method to estimate standard errors. Another very nice addition is a section on likelihood based methods utilizing the likelihood ratio test statistics and the development of confidence intervals based on the inversion of these P -values. Sample size calculations are covered in Chapter 4, with new sections that provide insight into how tables for equal sample sizes can be used for unequal sample sizes. This section also contains a very nice discussion of detectable effect sizes and interpolation. The chapter on randomization has been upgraded to include randomly permuted blocks,

while the section of variation on simple randomization has been rewritten.

A new chapter (Chapter 11) covers logistic regression. This chapter develops logistic regression historically, from bioassay models through the “analog of analysis of variance methods suitable for contingency table analysis”. The coverage is brought together through the use of likelihood ratio test statistics. There is a nice linkage between log-linear analysis, multinomial response models and discriminant analysis. In terms of modeling, both additive and interactive explanatory variables are characterized quite well. Finally, Chapter 11 includes an extensive section on polytomous logistic regression.

Poisson regression is detailed in Chapter 12. This is developed from a distribution approach (as compared to generalized linear models or log-linear models approach) with examples using Poisson processes and morbidity and mortality rates. Overdispersion and quasi-likelihood are presented, along with details on testing for overdispersion. Chapter 14 is a new chapter on regression analyses for matched samples. Section 14.1 of this chapter presents a very interesting overview of the problems in modeling data from matched sample designs. Both prospective and retrospective study designs are discussed, as well as the incidental parameter issues that are often overlooked. The authors explain, through examples, how misleading coefficients and biased log odds ratios can result from incorrectly modeling matched retrospective designs as though they were prospectively sampled. The chapter ends with a section on conditional logistic regression techniques.

The last two new chapters, 15 and 16, cover the latest techniques for correlated binary data and missing data, respectively. Chapter 15 begins with an overview of intraclass correlation for inference about one and two proportions in the presence of clustering, and progresses to using Mantel-Haenszel odds ratios for $2 \times 2 \times 2$ tables with clustered data. A more sophisticated section follows covering generalized estimating equations for correlated logistic regression with caveats on when estimators and test statistics may not be reliable. There are plenty of examples, including some sample SAS code (SAS Institute Inc., Cary, North Carolina). Chapter 16 provides a basic background into three mechanisms for missing data (MCAR, MAR and NI). The authors do a nice job relating these mechanisms to the analysis framework of binary outcomes, as

well as provide examples that are typical for 2×2 table analyses. Although the authors do not mark the introductory sections as “advanced”, these first few sections are fairly deep and quite comprehensive. Sections 16.5–16.7, covering logistic regression modeling under the various mechanisms of missingness, are extremely challenging and complete. There are plenty of references, examples and details on methods of modeling with missing data, including SAS code. The problem set for this chapter is, in my opinion, one of the best in the book.

One of the unique features of this book is the way that it develops problems, and solutions, from the very simple to the very complex. I think this is primarily due to Drs Levin and Paik’s desire to keep true to the previous editions, which they do very well. The problems for each chapter have been updated, and while some of them are quite challenging, they are presented in a stepwise manner with relevant hints that make them

practical for homework. In keeping with Dr Fleiss’s earlier editions, most of the material is presented in a frequentist manner; however, empirical Bayes methods are discussed in many chapters. The substantial amount of literature and techniques that are presented are impressive in their number, and in their arrangement. I would recommend this book as a valuable addition to any practicing statisticians’ bookshelf. The review of methods and literature alone make it handy resource. The presentation of the material, the examples, and the chapter problems make it an excellent textbook for use in a first course on discrete data in a statistics or biostatistics program and even a course in analyses in a public health program.

*Bonnie LaFleur,
Department of Biostatistics,
Vanderbilt University Medical Center,
Nashville, TN, USA*