Error and the Growth of Experimental Knowledge.

Deborah G. MAYO. Chicago: University of Chicago Press, 1996. xvi + 493 pp. \$29.95 (P); \$74 (H).

The aim of this book, written by a philosopher of science, is to analyze how practicing scientists formulate hypotheses, test them, and draw conclusions from their results. It is thus concerned with the use of statistics as a tool for progress in science.

The author discusses two principal methodologies proposed by statisticians for this purpose: the Bayesian approach and that of Neyman–Pearson. After a critical examination, she concludes that the former is unsatisfactory as a scientific methodology, and instead takes the error probabilities (of the first and second kind) of Neyman and Pearson as her starting point. However, she follows this theory only part way, accepting the central role of alternatives to the hypothesis and the calculation of error probabilities, but rejecting the behavioristic interpretation of the Neyman–Pearson theory. In fact, despite her disagreement with Fisher on the relevance of alternatives (and although the book does not include a systematic examination of his views), Mayo's outlook seems close to Fisher's. This is evident, for example, in her views on probability, the central role in scientific progress of experimentation under varying conditions, and the possibility of inductive inference.

Mayo's central idea of progress through "learning from error" appears to place her close to Popper's falsification philosophy. However, she contrasts Popper's purely negative approach with her own position that subjecting an experimental hypothesis to sufficiently severe testing can ultimately lead to its confirmation; that is, to reliance on it for further applications. How to achieve this aim by considering the primary scientific questions, the design of suitable experiments, and the analysis of the data is not fully explicated—partly because it is not possible to give cut-and-dried rules for such a strategy, and partly because the strategy's implementation is perhaps a task for statistics (and science) rather than for philosophy. Relevant techniques such as model selection, multiple comparisons, meta-analysis, and cross-validation are beyond the scope of this book, which is intended to be "nontechnical and open to readers without background in statistics and probability."

What the author provides instead (and what constitutes one of the most valuable aspects of the book) is a detailed account of a number of scientific investigations. Prominent examples are the experimental work by several laboratories of neutral currents (neutrino events without muons), Perrin's study of Brownian motion, and the eclipse experiments to test Einstein's theory of gravitation. These analyses illustrate and clarify the extended process through which scientists reach their conclusions.

In her central examination of the Neyman–Pearson theory, Mayo arrives at an interesting historical insight: that the behavioristic interpretation of this theory and Neyman's insistent denial of the possibility of "inductive inference" was not part of Pearson's original conception and that Pearson later distanced himself from it. Pearson's statistical philosophy (discussed in a chapter titled "Why Pearson Rejected the Neyman–Pearson [Behavioristic] Philosophy and a Note on Objectivity in Statistics") seems to be her own starting point. He is the hero of Mayo's story, and the frontispiece accordingly is a sketch by her of young Egon Pearson recalling "how certain early ideas came into my head as I sat on a gate overlooking an experimental blackcurrent plot."

Error and the Growth of Experimental Knowledge is an attractive and informative book that successfully knits together ideas from the philosophy of science, statistical concepts, and accounts of scientific practice. It provides a new framework for ideas in both statistics and the philosophy of science, and should be of interest to any statistician concerned with the foundations of our discipline.

Erich L. LEHMANN
University of California
and *Educational Testing Service*

Nonlinear Models for Repeated Measurement Data.

Marie DAVIDIAN and David M. GILTINAN. London: Chapman and Hall, 1995. xx + 359 pp. \$54.95.

The use of longitudinal designs to study the mechanisms of pharmacokinetics, development, and other biological and biomedical applications has increased over the past 20 years. This proliferation is due in part to the increasing sophistication of researchers, coupled with increasing demands from medical journals regarding the defensibility of the statistical methods used in an analysis. Another reason is the quest to produce more efficient designs, reducing the cost and the number of subjects needed. Equally important is that for many hypotheses, longitudinal designs are the best way to address the question, and for some hypotheses they are the only way. Collection of longitudinal data typically involves repeated measurements on subjects at different times or under changing experimental conditions, and there may be considerable variation among subjects in the number and timing of observations.

Methods for longitudinal data of a linear form, often called linear mixed-effects models, are well developed and include an array of estimation and inferential tools long available for fixed-effects models. The same cannot be said about nonlinear mixed-effects models. Although much progress has been made (see Yuh et al. 1994 for a recent overview), the field remains limited with respect to the availability of methods and the accessibility of software for implementation. In addition, uncertainty prevails among practicing biostatisticians with regard to model selection and computational approaches. In particular, there has been no single source to which to turn for an authoritative review of nonlinear models for repeated-measures data or to begin the study thereof.

Nonlinear Models for Repeated Measurement Data is the first complete book in an important and rapidly developing field. Davidian and Giltinan provide an up-to-date and extensive account of the state of the art in nonlinear mixed-effects models. The text is well organized and substantial, and provides sufficient background in nonlinear regression and linear mixed-effects methodologies so as to be self-contained. The authors clearly describe and contrast the currently available inferential approaches and pay particular attention to computational procedures and their advantages and shortcomings. They pay somewhat less attention to ancillary tools such as hypothesis testing and confidence interval estimation, but this is partly due to the paucity of methods available.

Chapter 1 introduces four motivating examples, along with a brief discussion of model specification. Chapter 2 reviews nonlinear regression models for independent data, beginning with the classical assumptions and moving rapidly to inferential procedures. Methods of maximum likelihood (ML), restricted maximum likelihood (REML), and pseudolikelihood (PL) are considered. Ordinary least squares (OLS) and generalized least squares (GLS) estimation of parameters are described clearly and succinctly, followed by a discussion of variance function estimation, confidence interval construction, and hypothesis testing. Computational aspects are reviewed, Gauss–Newton (Fisher scoring) and Newton–Raphson schemes are discussed specifically, and the importance of model parameterization is stressed. Two examples are presented for which the OLS and GLS fits are compared. The chapter closes with a clear discussion of related approaches, generalized linear models (GLM's) and generalized estimating equations (GEE's). GLM's and GEE's are important in providing a framework for regression analysis when the assumptions of normality and homoscedasticity do not hold, but when the variance is a known function of the mean response. Finally, a brief discussion and bibliographic notes are included. Subsequent chapters follow a similar outline.

Chapter 3 covers familiar hierarchical linear models. The linear mixed-effects model of Laird and Ware (1982) and a Bayesian specification as detailed by Searle, Casella, and McCulloch (1992, chap. 9) are introduced. Approaches to inference are discussed, emphasizing the interplay between estimation of fixed and random effects in each, followed by computational procedures such as the EM algorithm, Newton–Raphson, and the scoring method. The chapter closes with notes about software implementation in the context of two examples, limitations of the hierarchical linear model, and bibliographic notes. The work of Jones (1993) is mentioned anecdotally as a state-space approach to the analysis of hierarchical linear models, overlooking the fact that Jones extends his approach to nonlinear repeated measures in his chapter 7 and then to multiple responses in his chapter 8.

Although it largely is a much-needed review of hierarchical nonlinear models, Chapter 4 also heralds the beginning of the substance of the text. A general hierarchical nonlinear model along the lines of Racine-Poon (1985), Lindstrom and Bates (1990), and Vonesh and Carter (1992), as well as the author's own work, is described. The issues of intraindividual and interindividual variation are discussed, as are extensions for time-dependent covariates and multiple responses. The chapter closes with discussions of parametric, nonparametric, semiparametric, and Bayesian model specifications and an ample discussion of caveats.

Chapter 5 introduces inferential approaches based on estimating parameters for each individual where estimates of the covariance parameters are obtained by pooling, and focuses largely on GLS approaches. As one might imagine, the approach of this chapter requires sufficient data to describe each individual's response. Detail is provided on the estimation of population parameters via the standard two-stage method (Steimer, Mallet, Golmard, and Boisvieux 1984) or a global two-stage (GTS) method utilizing either a Newton-Raphson or an EM approach. Computational issues and software options are discussed, one example is worked through, and several important topics are discussed. The most important of these—