

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

USGS Northern Prairie Wildlife Research Center

US Geological Survey

1991

Book Review: Modern Elementary Probability and Statistics

Douglas H. Johnson

U.S. Fish and Wildlife Service, Douglas_H_Johnson@usgs.gov

Follow this and additional works at: <https://digitalcommons.unl.edu/usgsnpwrc>



Part of the [Other International and Area Studies Commons](#)

Johnson, Douglas H., "Book Review: Modern Elementary Probability and Statistics" (1991). *USGS Northern Prairie Wildlife Research Center*. 239.

<https://digitalcommons.unl.edu/usgsnpwrc/239>

This Article is brought to you for free and open access by the US Geological Survey at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in USGS Northern Prairie Wildlife Research Center by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

tend to ignore warnings and use normal approximations without any thought.

Chapter 5 on continuous random variables sorely needs more on transformations and change of variables. It would be nice to see the discussion on page 203 expanded to include discussion of the Jacobian rather than just the simple example given. Chapter 6 contains a short but excellent section on reliability.

In Chapter 7 are two errors in this otherwise well-edited book. On page 319, the total in the table at the top should be 48, not 40. On page 326, Figure 7.6 is the wrong figure. Figures 7.2 and 7.3 are excellent graphics. Figure 7.2 is a bivariate bar chart, and it really helps one visualize a bivariate relationship. Figure 7.3 illustrates a star plot, which also effectively illustrates the same discrete bivariate relationship.

A disturbing definition of a statistic is given on page 329. Highlighted in a box is the following definition: "A statistic is a number calculated from the observations in a sample." A more rigorous definition should be used in a book that places theory prior to methods in its title. For example, "A function of one or more random variables which does not depend upon any unknown parameter is called a statistic" (Hogg and Craig 1978, p. 122). Moreover, it is usually emphasized that the statistic is a random variable and the distribution of the statistic may depend on unknown parameters.

Chapters 8 and 9 would have to be supplemented with other source material for a class in which the exponential family, completeness, and uniqueness are important concepts. A far too brief overview of the exponential family is presented, while the latter two topics are completely ignored. Chapter 8 provides a brief but good section on simulating sample distributions. Chapter 9 covers the determination of sample size. It is well written; however, it is based on the tightness of confidence intervals because the concept of power has not yet been introduced.

Chapter 10 on significance testing wisely uses P values (the observed level of significance) throughout the chapter. This chapter contains an excellent section on statistical versus practical significance. Nonparametric tests are introduced along with parametric and receive nice coverage here and in Chapter 11.

Chapters 11–13 provide reasonable coverage of their designated topics. As with the preceding chapters, the extensive use of solved and unsolved problems helps clear up any confusion that might arise in this well-written book.

Chapter 14 on analysis of variance presents multiple-comparisons methods (though limited to Bonferroni and Scheffé). This is good to see in an introductory textbook because in practice multiple-comparison procedures play a major role in many analyses.

Chapter 15 has an important section on the regression effect. The authors mention that in baseball the regression effect is termed the *law of competitive balance*. Chapter 16 provides respectable coverage of Bayesian methods for an introductory book.

There are 45 pages of detailed statistical tables in the first appendix. The second appendix provides answers to selected unsolved problems presented in the individual chapters. A minor error appears in the index—"Two-way ANOVA" appears before "Transformations."

Overall, this is a fine introductory statistics book. Although it will be too light in certain areas for those desiring a detailed theoretical treatment, it should provide a sound, nonintimidating introduction to mathematical statistics for many.

William V. Harper
Resource International
Columbus, Ohio

REFERENCE

- Hogg, R. V. and Craig, A. T. (1978). *Introduction to Mathematical Statistics*. New York: Macmillan.

Modern Elementary Probability and Statistics, by Edward J. Dudewicz, Pinyuen Chen, and Baldeo K. Tanuja, Columbus, OH: American Sciences Press, 1989, ix + 372 pp., \$34.50 (paperback).

This is an introductory textbook, "developed for, and class-tested in," first courses in elementary probability and/or statistics. The only prerequisite is high-school algebra. It has some nice features, but shortcomings in both substance and style detract from its value.

The book consists of three parts. Part I develops the basic probability rules and describes some distributions. Part II briefly overviews major statistical practices, such as point and interval estimation, hypothesis testing, and the like. Some topics are touched on only lightly—for example, nine pages for analysis of variance and design, four pages for statistical selection and decision. Worse, little help is given readers who want further references to the material. Part III includes chapters on statistical programming in SAS, Minitab, and BMDP, as well as basic statistical tables. Material on the statistical packages is sometimes presented in too trivial detail, such as "Never put a period after END" in the discussion of BMDP. Moreover, it advises BMDP users to run programs with one error at a time to find out what kind of message gets spewed out. Is that state-of-the-art error detection? There are also hints in Part III on succeeding in elementary probability and statistics that are appropriate for college work in general. The "Selected Answers to Problems" section makes one wonder about the answers that were not selected.

Some matters of substance that troubled me include the use of grouped data to estimate variance, without mention of the bias associated with that procedure and the need for Sheppard's correction (p. 14). Most striking is the statement on page 15 that the central limit theorem applies to *observations*, not just means: "about 95% of the observations are between $\bar{x} - 2s$ and $\bar{x} + 2s$," which "can be shown mathematically if n is large." On page 183, the authors conclude that a tested difference is real, ignoring the possibility that a Type I error might have occurred. Residuals are defined as $\hat{Y} - Y$, the negative of their customary definition, which could cause some confusion.

There are some favorable aspects to the material and its presentation. The authors introduce percentile curves early, on page 17, and make effective use of them as descriptive and comparative tools. The book also advocates the use of the Hsu solution (Scheffé 1970) to the Behrens-Fisher problem of comparing two independent means. As a table is introduced in the text, the relevant portion is shown and annotated so that readers can readily see how to use the table.

The best feature of the book is its choice of examples, which are excellent and should appeal especially to college students. Probability is described not solely in terms of coin tossing, die throwing, and drawing balls from urns but in terms of blood alcohol levels, overbooking on airlines, and rates of indecent-exposure cases. Many examples are drawn from the front page—circumstantial evidence used in an assault trial, why predictions from polls may be wrong, and employment discrimination.

The worst features of the book involve style. It appears to have been printed photographically from typewritten text. A desktop publisher would have been far preferable (and a spell checker could have been used to advantage). About half of the material is double-spaced, and the rest single-spaced, giving an uneven appearance. The double spacing wastes much of the page, which is particularly annoying because the (photographically reduced) print is small. Had the text been printed, the material would be much more readable. Often words or phrases have been physically replaced—literally cut and pasted—leaving inappropriate gaps between words (or sometimes no spacing at all), a mixture of fonts, or jagged lines of type. Underlining sometimes is continuous,

sometimes is intermittent, and sometimes (with subscripted material) resembles a step function. The paper is a bit thin, allowing print on one side to interfere with reading the other side. Page numbers are buried between two lines of dashes and on the inside of a page, making them difficult to find quickly. The authors use "data" as a singular noun, a usage favored by some computer types and other untrustworthy guardians of the English language, but one I think we should resist.

With some revision, and true printing, this book could have considerable appeal to students learning introductory probability and statistics.

Douglas H. Johnson
U.S. Fish and Wildlife Service

REFERENCE

Scheffé, H. (1970), "Practical Solutions of the Behrens-Fisher Problem," *Journal of the American Statistical Association*, 65, 1501-1508.

Genstat 5: A Second Course, by Pete Digby, Nick Galwey, and Peter Lane, New York: Oxford University Press, 1989, xi + 233 pp., \$29.95 (paperback).

This book is a guide to the advanced features of Release 5 of the Genstat package distributed by the Numerical Algorithms Group of Oxford, England. It assumes familiarity with the entry-level guide, *Genstat 5: An Introduction*, already reviewed in this journal (Driscoll 1989). As was the case with that review, comments given here pertain to the book rather than the computer software, which has been reviewed elsewhere (Bratchell 1987).

Chapters 1, 2, and 3 of the book discuss Genstat's facilities for advanced regression analysis, including regression on grouped data, log-linear analysis of contingency tables, and nonlinear regression. Chapter 4 (an expansion of the treatment of designed experiments in the first book) treats nested and crossed factors, orthogonal polynomials and contrasts, and the analysis of covariance.

Chapters 5, 6, and 7 present techniques for multivariate analysis: principal components and linear discriminant analysis, principal coordinate analysis and Procrustes rotation, and cluster analysis. Chapters 8 and 9 consider structured programming in Genstat and the creation of new commands via Genstat procedures (i.e., macros). Chapters 10 and 11 present built-in commands and procedures for time series methods: correlograms, ARIMA models, intervention analysis, cyclical structure, and spectra.

The format of this second book is identical to that of the first, so it has all benefits (and the one small drawback) noted in the earlier review: clear presentations, textbook style, exercises with solutions, good cross-referencing, and a thorough index. Many users of Genstat will want a copy of this book on their desks; most should have one on their shelves.

Michael F. Driscoll
Arizona State University

REFERENCES

Bratchell, N. (1987), "Review of Genstat," *Chemometrics and Intelligent Laboratory Systems*, 1, 300-301.
Driscoll, M. F. (1989), Review of *Genstat 5: An Introduction*, by Peter Lane, Nick Galwey, and Norman Alvey, *Technometrics*, 31, 485-486.

Systems Analysis Using Simulation and Markov Models, by John R. Clymer, Englewood Cliffs, NJ: Prentice-Hall, 1990, xvi + 352 pp.

This book is a textbook for the advanced undergraduate or first-year graduate-level student. The author suggests that an understanding of operations research and the ability to program in Pascal or FORTRAN are the only prerequisites. It is assumed that readers will have completed the standard undergraduate mathematics courses for engineers. Some familiarity with matrix manipulations and statistical methods would be helpful, as well. The book should be valuable to practicing engineers who work in systems analysis but are not familiar with operational evaluation modeling. Although the author develops procedures gently through simple examples, the book gives readers the ability to apply these techniques to larger systems in any field. The frontiers of research are described by recent work with large-scale systems conducted by the author, his colleagues, and students. Those readers who augment the book with probability theory, stochastic processes, and the computer courses suggested by the author will be prepared to do valuable research in this field.

The methods in the book are operational, focusing on what the system does, rather than structural, which emphasizes how the system works. This method is given the name operation evaluation modeling (O_pEM). The technique is based on parallel processes, directed graphs, simulation, Markov processes, reliability processes, queuing theory, sensitivity analysis, and cybernetics. The first chapter presents an excellent overview of the procedure. Directed graphs are shown to aid in the communication between the programmer and the user. Involvement of the user in the design assures that the model will satisfy research objectives. The importance of understanding the system combined with the use of sensitivity analysis to optimize system performance and cost is stressed.

The second and the third chapters, "Simulation Based on the Concept of Parallel Processes" and "Applications of Operational Evaluation Modeling," show how the system works. Chapters 4-7 give detailed presentations of the components of Markov modeling, systems analysis using Markov processes, reliability processes, and queuing theory. Chapter 8 discusses in some detail a Pascal simulation programming system.

The author states in the book that Pascal and FORTRAN computer-programming languages are chosen to provide readers with portable, flexible, and inexpensive access to modeling methods that can be readily understood and investigated. An added advantage of this approach is that many applications are available in Pascal or FORTRAN programs. A software disk containing FORTRAN and Pascal programs of Markov modeling experiments is provided in the back of the book to help the students with some experiments.

The last chapter on cybernetics is intended to inform practicing engineers and to lead student readers to this new frontier. In the first part of the chapter, cybernetic theory, as discussed by Ashby (1956, 1960), is discussed from the O_pEM point of view. The second part presents an application of expert systems and fuzzy set theory to airport management and discusses proposed research.

Students or practicing engineers will be stimulated to consider many other examples or applications as they progress through this book. The tools are presented to enable ambitious researchers to capitalize on these methods of systems analysis.

R. B. Deal, Jr.
University of Oklahoma Health Sciences Center