This Week's Citation Classic®

CC/NUMBER 14
APRIL 3, 1989

Bishop Y M M, Fienberg S E & Holland P W (with the collaboration of Light R J & Mosteller F). *Discrete multivariate analysis: theory and practice.* Cambridge, MA: MIT Press, 1975. 557 p.

[Dept. Biostatistics, Harvard Sch. Public Health, Boston, MA; Dept. Applied Statistics, Univ. Minnesota, St. Paul, MN; and Natl. Bureau for Economic Research, Cambridge, MA]

This book is the first comprehensive treatise on the analysis of categorical data using loglinear and related statistical models, especially in the context of multivariate cross-classifications. It emphasizes the basic methodology underlying asymptotic theory and gives detailed examples illustrating the application of loglinear model techniques to data from a variety of substantive disciplines. [The *SSCI*® and *SCI*® indicate that this book has been cited in over 1,970 publications.]

Approaches for Statistical Analysis of Contingency Tables

Stephen E. Fienberg
Departments of Statistics and
Social and Decision Sciences
Carnegie Mellon University
Pittsburgh, PA 15213-3890

November 25, 1988

When I was a Harvard University graduate student in 1966, Frederick Mosteller, my adviser, asked if I'd work on "a little problem" on smoothing counts. He was in the midst of a major collaborative investigation of the anesthetic halothane, which has come to be known as the National Halothane Study. The little problem was one of many methodological issues in the investigation, and several statisticians were approaching it in different ways under the guidance of Lincoln Moses, Mosteller, and John Tukey. Yvonne M.M. Bishop was already working full-time on this project, and this effort led to her PhD dissertation on loglinear models for multidimensional contingency tables, completed in 1967. My problem on smoothing counts led to an extended collaboration with Paul W. Holland, then an assistant professor, and to a portion of my own PhD dissertation on contingency table analysis.

As I was completing my dissertation in the spring of 1968, Mosteller gathered together a group of graduate students and junior faculty one evening at his home in order to discuss the preparation of a book that would pull together all of the research that had been done in the department over the preceding years on the analysis of contingency tables. Over the next several years various members of the group wrote up papers and draft chapters for the book. In the end, Yvonne, Paul, and I became the principal authors, and, Richard J. Light helped with the chapter on measures of association and agreement. But Mosteller's hand can be seen in virtually every chapter. His suggestions and skilled editing were responsible for most of the book's special features, e.g., the extensive use of substantive examples and the index of data sets at the end of the book.

Few parts of the book resemble the research reports and other materials with which we began in 1968. The National Halothane Study and the methodology developed for it spawned considerable interest in the statistical community. Ultimately, we blended together this new literature with our own approaches, including several previously unpublished methods and applications, focusing wherever possible on what working statisticians should know about the subject. For example, the final chapter on asymptotic methods grew out of a desire to make a variety of specialized results accessible to applied statisticians and available in a single source.

The book was well-received from the outset and became known as "the jolly green giant" because of the color of its cover and its size. It stimulated the development of new graduate courses on categorical data analysis, and books and papers continue to appear that extend our ideas and approaches, e.g., graphical loglinear models,[1] generalized linear models,[2] capture-recapture models for population estimation,[3] and families of goodness-of-fit statistics.[4]

1. **Darroch J N, Lauritzen S L & Speed T P.** Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* 8:522-39, 1980. (Cited 15 times.)
2. **McCullagh P & Nelder J A.** *Generalized linear models.* London: Chapman and Hall, 1983. 261 p. (Cited 325 times.)
3. **Cowan C D & Malec D.** Capture-recapture models when both sources have clustered observations. *J. Amer. Statist. Assn.* 81:347-53, 1986.
4. **Read T R C & Cressie N A C.** *Goodness-of-fit statistics for discrete multivariate data.* New York: Springer-Verlag, 1988. 211 p.

1A-16