Discrete Multivariate Analysis

Yvonne M. M. Bishop, Stephen E. Fienberg, and Paul W. Holland
with the collaboration of
Richard J. Light and Frederick Mosteller

Discrete Multivariate Analysis
Theory and Practice

Springer

Yvonne M. Bishop
Washington, DC 20015-2956
Ymbishop@verizon.net

Stephen E. Fienberg
Department of statistics
Carnegie-Mellon University
Pittsburgh, PA 15213
Fienberg@stat.cmu.edu

Paul W. Holland
Educational Testing Service
Princeton, NJ 08541
pholland@ets.org

ISBN 978-0-387-72805-6

Library of Congress Control Number: 2007928365

Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springer.com

Preface

The analysis of discrete multivariate data, especially in the form of cross-classifications, has occupied a prominent place in the statistical literature since the days of Karl Pearson and Sir R. A. Fisher. Although Maurice Bartlett's pioneering paper on testing for absence of second-order interaction in $2 \times 2 \times 2$ tables was published in 1935, the widespread development and use of methods for the analysis of multidimensional cross-classified data had to await the general availability of high-speed computers. As a result, in the last ten years statistical journals, as well as those in the biological, social, and medical sciences, have devoted increasing space to papers dealing with the analysis of discrete multivariate data. Many statisticians have contributed to this progress, as a glance at the reference list will quickly reveal. We point, especially, to the sustained and outstanding contributions of Joseph Berkson, M. W. Birch, I. J. Good, Leo A. Goodman, James E. Grizzle, Marvin Kastenbaum, Gary G. Koch, Solomon Kullback, H. O. Lancaster, Nathan Mantel, and R. L. Plackett.

The one person most responsible for our interest in and continued work on the analysis of cross-classified data is Frederick Mosteller. It is not an overstatement to say that without his encouragement and support in all phases of our effort, this book would not exist. Our interest in the analysis of cross-classified data goes back to 1964 and the problems which arose during and after Mosteller's work on the National Halothane study. This work led directly to the doctoral dissertations of two of us (Bishop and Fienberg), as well as to a number of published papers. But Fred's contributions to this book are more than just encouragement; he has read and copiously commented on nearly every chapter, and while we take complete responsibility for the final manuscript, if it has any virtues they are likely to be due to him.

Richard Light enthusiastically participated in the planning of this book, and offered comments on several chapters. He prepared the earlier drafts of Chapter 11, Measures of Association and Agreement, and he made the major effort on the final version of this chapter.

We owe a great debt to many of our colleagues and students who have commented on parts of our manuscript, made valuable suggestions on aspects of our research, and generally stimulated our interest in the subject. Those to whom we are indebted include Raj Bahadur, Darrell Bock, Tar Chen, William Cochran, Joel Cohen, Arthur Dempster, O. Dudley Duncan, Hillel Einhorn, Robert Fay, John Gilbert, Anne Goldman, Shelby Haberman, David Hoaglin, Nathan Keyfitz, William Kruskal, Kinley Larntz, Siu-Kai Lee, Lincoln Moses, I. R. Savage, Thomas Schoener, Michael Sutherland, John Tukey, David Wallace, James Warram, Sanford Weisberg, Janet Wittes, and Jane Worcester.

For the production of the manuscript we are indebted to Holly Grano, Kathi Hirst, Carol Lambert, and Mary Jane Schleupner.

The National Science Foundation has provided substantial support for our research and writing under grant GS-32327X1 to Harvard University. We have also received extensive support from other research grants. These include: research grants CA-06516 from the National Cancer Institute and RR-05526 from the Division of Research Facilities and Resources, National Institutes of Health to the Children's Cancer Research Foundation; National Science Foundation research grants GP-16071, GS-1905, and a grant from the Statistics Branch, Office of Naval Research, to the Department of Statistics, University of Chicago, as well as a grant from the Alfred P. Sloan Foundation to the Department of Theoretical Biology, University of Chicago; National Science Foundation research grant GJ-1154X to the National Bureau of Economic Research, Inc., and a faculty research grant from the Social Science Research Council to Paul W. Holland.

Earlier versions of material in several chapters appeared in *The Annals of Statistics*, *Biometrics*, *Biometrika*, and *The Journal of the American Statistical Association.*

Brookline, Massachusetts Y.M.M.B.
New Brighton, Minnesota S.E.F.
Hingham, Massachusetts P.W.H.

February 1974

CONTENTS