

Preface

The focus of this book is on models for contingency table analysis. It deals mainly with log-linear models and special models for ordinal data (log-linear or nonlinear). Special models for ordinal data are dealt with to a greater extent, covered in Chaps. 6, 7, and 9. These models, as for example, association models for two- and multi-way contingency tables or symmetry models for square tables, though very powerful and of great interpreting value, are not very popular in use. This is mainly because they are not readily provided as model options in standard statistical software. Though they can be fitted in R by some special packages, their application usually requires some expertise or experience on these models and R. Existing books on contingency tables or categorical data analysis either do not include association models in their contents or refer only to the simplest types of them. The few exceptions, although treating these models in more detail, remain in the methodological part without providing guidance for applying them in practice. Thus, such models can be used mainly by experts on the topic. This ascertainment was the motivation and the idea behind the concept of this book: to provide a reference that exploits these models, explains their features and interpretation aspects in detail, and simultaneously trains the reader to fit them in practice. Additionally, special issues not covered in other books, such as the adjustment of models to account for structural zeros, are addressed. The goal was to end up with a methodological book that makes all approaches, models, or graphs presented, easy reproducible.

The aim to familiarize readers with methods and models for the analysis of contingency tables and their use in practice is served by discussing the models' features and interconnections and by giving special emphasis on the examples' analysis, their interpretation, and their implementation in R. When needed, special R functions are provided (in the web companion of the book) that automatize the required procedures, simplifying thus their applicability. For example, functions are provided for deriving the midrank scores of a classification variable or computing the local odds ratios of a two-way contingency table (or other types of generalized odds ratios). Hence, all models, measures, and graphs discussed can easily be realized in R by handy functions. The examples are worked out in R, explaining the use of the functions, so that the reader is gradually trained and at the end in the

position to alter the functions and adjust them to special needs. The web companion of the book is to be found under

<http://cta.isw.rwth-aachen.de>

Framing this book on model-based analysis, the great body of nonparametric methods (especially for ordinal data) and smoothing methods for contingency tables is not considered. Emphasis is given primarily on models that treat the variables symmetrically rather than distinguishing between response and explanatory variables. Additionally, since the book deals only with contingency tables, regression-type models that involve also continuous explanatory variables are not addressed at all. Thus, logistic regression, though very important in categorical data analysis, is partially covered, only for the case of categorical explanatory variables (*logit models*). Furthermore, clustered categorical data and multivariate response models are not considered. Only bivariate response models are considered for data represented in square two-way contingency tables with commensurable classification variables, nominal or ordinal. For more than two occasions, we refer to other special reference sources.

The approach adopted is the asymptotic frequentist approach. A short reference on Bayesian analysis of contingency tables and on small sample inference is provided in the last chapter.

The readership target groups are (a) graduate students or researchers (in statistics or in psychometric, social, biomedical, and pedagogical sciences) and (b) practitioners (e.g., for social or consumers' surveys). The first five chapters (up to Sect.5.4) address both groups, though group (a) could go quickly through it, for filling gaps and building up gradually the R part. Expertise in R is not required. Regarding the following material (from Sect.5.5), group (b) would probably be more interested in the simpler models (as linear-by-linear, row, or column effect association models), easy to present and interpret, while group (a) also in more advanced topics (such as handling structural zeros, the multiplicative row-column association model, models for the marginal distributions, or the generalized odds ratios). An updated rich literature review on a bright aspect of topics is provided at the end of each chapter and is mainly addressed to group (a).

The scope is to simultaneously develop the theoretical and R programming skills required for analyzing contingency tables, as well as evaluating and interpreting the results. Parts of the manuscript are based on my notes for the graduate course on categorical data analysis, which I held for about 10 years in the Department of Statistics and Insurance Science of the University of Piraeus in Greece.

I would like to thank those who have been involved in this book project. Special thanks to Alan Agresti who influenced the most my view on categorical data. His *Categorical Data Analysis* book, in its first 1990 edition, is partly responsible for my involvement with categorical data. I also thank him for providing valuable comments on the manuscript, suggesting alterations and additions. I thank Panayiotis Bobotas for proofreading the complete draft, Anna Gottard for her critical comments and suggestions on graphical models, and Anestis Touloumis for his helpful comments on Chaps. 5–10. I appreciate all those who accompanied my scientific journey, in particular my former supervisor Takis Papaioannou.

Thanks also to Narayanaswamy Balakrishnan, editor in chief for the series “Statistics for Industry and Technology,” for his interest in this project, and to Mitch Moulton, editorial assistant at Springer, for his help facilitating it. This book would not have been possible without the generous support of my parents, Athina and Dimitris Kateris, to whom I express my deep gratitude. Finally, I thank my daughter Zenia for her patience during the preparation period and her interest in the development of the project.

Aachen, Germany
January 2014

Maria Kateri

Contingency Table Analysis

Methods and Implementation Using R

Kateri, M.

2014, XVII, 304 p. 21 illus., 8 illus. in color., Hardcover

ISBN: 978-0-8176-4810-7

A product of Birkhäuser Basel