# *Preface*

Researchers in many fields, such as biology, medicine, social and behavioral sciences, law, and economics regularly encounter data presented in contingency tables. The history of analysis of contingency tables goes back to the highly influential article by Karl Pearson in 1900. Since then, an almost inexhaustible amount of research has been done, and a vast number of different methods for the analysis of contingency tables is described in the literature. This is even the case for seemingly simple contingency tables such as the ubiquitous $2 \times 2$ table.

For the last fifteen years, we have taught courses and done research on statistical methods for the analysis of contingency tables. During this work, the need for an *up-to-date book* on this topic became increasingly clear to us. Beyond describing how to derive and calculate *effect size estimates*, *confidence intervals*, and *hypothesis tests* for contingency tables, we have emphasized *evaluations* and comparisons of methods, illustrations of the methods on *real-life data*, and *recommendations for practical use*. Large parts of the material in this book have until now been available only in specialized journals, and some parts of it have not been previously published at all.

This book should be accessible to anyone who has taken a basis course in statistics, but as a textbook in contingency tables, it aims primarily at the master's or PhD level.

This book covers contingency tables with unordered and ordered categories, with paired or unpaired data. It covers one- and two-dimensional tables, as well as three-dimensional tables in the form of stratified two-dimensional tables. Tables of higher dimensions are outside the scope of this book. *Logistic regression* is studied in several chapters; however, loglinear models are only briefly considered. *Meta-analysis* is covered in depth as part of Chapter 10 on stratified $2 \times 2$ tables. Although we have aimed at giving a comprehensive presentation of the analysis of contingency tables, several topics are not covered. Multivariate analyses, such as factor analysis of categorical data, latent class analysis, correspondence analysis, and item response theory are, for instance, not covered.

Chapter 1 deals with various introductory material, including a general introduction to categorical data and contingency tables in Sections 1.1–1.3. Sections 1.5–1.8 and Section 1.11 cover general theory that may be skipped by readers who use this book as a *reference book* or as a *guide for practical applications*.

Chapters 2–11 are organized by table size and type. Each chapter deals

with one particular type of table, characterized by size (the number of rows and the number of columns), whether the variables are ordinal or nominal, whether the tables are stratified or not, and whether the observations are paired or unpaired. The start of each chapter describes relevant study and sampling designs, followed by real-life data examples and relevant statistical models. The main part of each chapter contains descriptions of the most important and interesting statistical methods. The methods are divided into confidence intervals for relevant effect measures—such as the difference between probabilities, the ratio of probabilities, and the odds ratio—and hypothesis tests. Within each section, the methods are illustrated with data examples, and their properties are evaluated according to criteria described in Section 1.4. Each chapter ends with recommendations for which methods to use and when to use them. These chapters are to a large extent *self-standing*, such that a reader will find all that he or she needs to analyze a particular table type in the corresponding chapter, without having to read the preceding chapters.

To ease guidance for the reader, we also provide an appendix with a *list of the 250 methods* described in these chapters, with recommendations marked, and with page references to where the methods are defined. Also included in the appendix is a list of examples with page references to where the examples are introduced and analyzed.

Chapter 12 deals with *sample size and power calculations*, and includes descriptions of the superiority, equivalence, and non-inferiority study designs.

The last chapter in this book, Chapter 13, contains miscellaneous topics: *diagnostic accuracy, inter-rater agreement, missing data, structural zeros, categorization of continuous variable, and ecological inference.*

At the book's website `http://contingencytables.com`, you will find *Matlab* and *R-code* for almost all the methods described in this book. The website also includes a list of errata and contact information for the authors. Please contact us if you discover any errors or have comments or questions about the book.

<div align="right">

Morten W. Fagerland, Stian Lydersen, Petter Laake
Oslo/Trondheim, June 2017

</div>