

Kwantitatieve Methoden

Book Review Section

Book Review Editor: Alex J. Koning

Book review 72B35

SIMONOFF, JEFFREY S. (2003).

Analyzing Categorical Data.

Springer Texts in Statistics. Springer-Verlag, Berlin.

ISBN 0-387-00749-0, XV, 496 p. 64 illus., Hardcover, EUR 84,95.

This textbook provides a comprehensive introduction to the analysis of categorical data. This type of data arises often in many fields, including biometrics, business administration, economics, insurance, management, marketing, medicine, psychology, sociology, sports and so on. Nonetheless, in statistical courses, categorical data usually get less attention than the examination of continuous data. And when the analysis of categorical data is on the schedule, then it is usually restricted to elementary principles such as the analysis of simple two-by-two tables.

The book has a wide coverage of topics, including count regression models, such as Poisson, negative binomial, zero-inflated and zero-truncated models; loglinear models for two-dimensional and multidimensional contingency tables, including for square tables and tables with ordered categories; and regression models for two-category (binary) and multiple-category target variables, such as logistic and proportional odds models.

The ten chapters of the book are divided into three parts. Chapter 1 introduces the special nature of categorical data. Chapters 2 and 3 review fundamental Gaussian-based data-analytic methods and concepts like least squares estimation, interval estimation, hypothesis testing, checking assumptions, regression diagnostics, model fitting and model selection. The second part of the book, chapters 4 through 8, examines the modelling of count data, using regression techniques. Chapter 4 focuses on the most important distributions for categorical data, the binomial, Poisson and multinomial. Also the formulation of tests of goodness-of-fit, over- and underdispersion and robust estimation is addressed. Chapter 5 is devoted to regression modelling. Specifically, attention is paid how the results for the generalized linear model can be applied to specific models for count data. In chapter 6, this modelling is extended to two-dimensional contingency tables. Chapter 7 reviews model fitting for tables with more structure, such as tables with ordered categories, square tables, and data that arise in the form of matched pairs. Chapter 8 addresses the analysis and presentation of the multidimensional contingency tables. Chapters 9 and 10 constitute the third part of the book.

Chapter 9 discusses regression analysis of binary data using logistic regression and its competitors. Chapter 10 subsequently extends these models to allow for multiple category response data.

All methods in the book are illustrated with detailed analyses of real data examples, many from recent subject area journal articles. Also a large number of exercises and references are provided for each chapter. Data sets and computer code, for the greater part in Splus, are available at the web.

The author has succeeded in writing a useful and readable textbook combining most of general theory and practice of count data. The author claims that the book is suitable for a very heterogeneous group of students, including undergraduate statistics majors, M.B.A. students, M.S and Ph.D. (statistics) students. Although the material is presented very clearly and illustrated by many examples, the textbook is probably not so easily accessible. Students who do not have a firm knowledge of statistics and mathematics or those who are not familiar with mathematical notation will have trouble with the presented material. A strong knowledge of basic statistical principles and mathematical notation is still a prerequisite to become familiar with the presented material.

dr ir M. H. J. de Bruijne