

## Book review of “Analyzing Categorical Data” by Jeffrey S. Simonoff

Jeroen K. Vermunt

Department of Methodology and Statistics

Tilburg University

Over the last decade, various textbooks on categorical data analysis have appeared (e.g., Agresti, 1996, 2002; Long, 1997; Powers and Xie, 2000; Vermunt, 1997). An important question therefore, is whether there is a need for another textbook on this topic. The answer to this question would be yes if the new book turns out to be clearly different from the existing textbooks in terms of level or focus, or if it does a better job in explaining the main categorical data concepts.

It is clear that the focus of Simonoff’s book is different from other books on categorical data analysis. After reviewing the normal linear model (Chapters 2 and 3), the author moves to distributions and regression models for count data (Chapters 4 and 5), which in my opinion forms the core of the book. The next three chapters (6, 7 and 8) deal with simple and more complicated log-linear models for frequency tables, as well as an interesting bivariate extension of the Poisson model. Chapters 9 and 10 discuss regression models for binary and polytomous dependent variables, respectively.

The order in which the various topics are discussed is very different from what is done in other textbooks. The more typical order is either to start with the Gaussian model and subsequently move to regression models for binary outcome variables, or to start with the analysis of two and three way tables and then move to logistic regression analysis. Also, the focus of the book differs from other books in that much less attention is paid to binary regression models, and much more to regression models for counts and to regression diagnostics.

As an introductory textbook, the book is comprehensive enough since all basic topics in categorical data analysis are discussed. On the other hand, because of its emphasis on count data, the book is somewhat unbalanced in terms of comprehensiveness and level. Very sophisticated models for count data are discussed, such as mixed models, zero-inflated models, models for truncated counts, and nonparametric regression for counts. The book is, therefore, much more advanced in *these* chapters than in the others. Nevertheless, some important topics in the analysis of count data are not covered at all. For example, the strong connection between zero-inflated and gamma mixture models and finite mixture count regression models is not discussed, nor the connection between the Poisson regression model for counts and regression models for rates (survival models), two of the main topics in Vermunt (1997).

Although Simonoff’s book is a good textbook – it contains lots of examples and exercises – depending on the focus of the course, I might prefer another text. For example, in a course on regression models for categorical data, I would prefer Long (1997), who also dedicates quite some attention to models for counts, but near the end of the book rather than in one of the early chapters. Better introductory textbooks in categorical data analysis are Agresti (1996) and Powers and Xie (2000). Long (1997), Agresti (1996), and Powers and Xie (2000) do a better job explaining how to interpret results from a categorical data analysis.

In addition, Agresti (2002) provides a much better reference work on categorical data analysis since it also contains important state-of-art issues such as transition models, marginal models, random effects models, and latent class models, which are missing from Simonoff's book. However, despite these limitations and omissions, I think Simonoff's book is a valuable addition to the literature because it discusses important models for counts that Agresti does not pay attention to, such as models for truncated counts, zero-inflated models, and models for bivariate counts.

## References

- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. New York: John Wiley and Sons.
- Agresti, A. (2002). *Categorical Data Analysis*. Hoboken, NJ: John Wiley and Sons.
- Powers, D.A., Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. San Diego: Academic Press.
- Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oakes, CA: Sage Publications.
- Vermunt, J.K. (1997). *Log-linear Models for Event Histories*. Thousand Oakes, CA: Sage Publications.