Hilbe, Joseph M.

Negative binomial regression. (English) Zbl 1131.62068

Cambridge: Cambridge University Press (ISBN 978-0-521-85772-7/hbk). xii, 251 p. (2007).

This book is devoted to the negative binomial model and its many variations. Every model currently offered in commercial statistical software packages is discussed in detail – how each is derived, how each resolves a distributional problem, and numerous examples of their application. Many have never before been thoroughly examined in a text on count response models: the canonical negative binomial; the NB-P model, where the negative binomial exponent is itself parameterized; and negative binomial mixed models. As the models address violations of the distributional assumptions of the basic Poisson model, identifying and handling overdispersion is a unifying theme. For practising researchers and statisticians who need to update their knowledge of Poisson and negative binomial models, the book provides a comprehensive overview of estimation methods and algorithms used to model counts, as well as specific guidelines on modeling strategy and how each model can be analyzed to access goodness-of-fit.

The book is organized as follows: Chapter 1 provides a brief overview of count response regression models. Chapter 2 examines the two major methods of parameter estimation relevant to modeling Poisson and negative binomial data. Chapter 3 is devoted to the deviation of the Poisson log-likelihood and estimation equations. Chapter 4 details the difference in real versus apparent overdispersion. Criteria are specified which can be used to distinguish real from apparent overdispersion. Simulated examples are constructed that show how apparent overdispersion can be eliminated. Chapter 5 gives the definition of the negative binomial probability distribution function and derivation of the various statistics required to model the canonical and traditional form of the distribution. In addition, it also gives the derivation of the Poisson-gamma mixture parameterization that is used in maximum likelihood algorithms.

Chapter 6 discusses the development and interpretation of the NB-2 model. Examples are provided that demonstrate how the negative binomial is used to accommodate overdispersed Poisson data. Chapter 7 addresses alternative parameterizations of the negative binomial. Chapter 8 deals with a common problem faced by researchers handling real data, some of which exclude a zero count and some of which have an excessive number of zeros. Chapter 9 discusses truncated and censored data and how they are modeled using appropriately adjusted Poisson and negative binomial models. Chapter 10 addresses the subject of negative binomial panel models. These models are used when he data are either clustered or when they are in the form of longitudinal panels.

Five appendices are provided to give some functions, an algorithm and some data sets (all data sets and additional codes can be found on a companion website). It is noted that each chapter ends with a summary section and many chapters have exercises. The Stata statistical package is used throughtout the text to display example model output and the LIMEDP software is also used to display some outputs.

I think that this book is well written and can serve as an excellent reference book for applied statisticians who would use negative binomial regression models in their work. It can also serve as a textbook for a special topic course in negative binomial regression modeling for undergraduate students or graduate students.

Reviewer: Yuehua Wu (Toronto)

MSC:

62J99 Linear inference, regression
62-02 Research exposition (monographs, survey articles) pertaining to statistics
62J12 Generalized linear models (logistic models)
62-01 Introductory exposition (textbooks, tutorial papers, etc.) pertaining to statistics

Keywords:

Anscombe residual; autoregressive correlation; Bernoulli trials; beta distribution; negative binomial model;

BIC; binomial; bootstrap; canonical; censored Poisson; censored negative binomial; censored econometric; censored survival; cluster; complementary loglog; conditional fixed effects negative binomial model; conditional fixed effects Poisson model; correlation structure; cumulant; deviance residual; dispersion; endogenous stratification; equidispersion; exchangeable correlation; exponential family; Fisher scoring; gamma model; generalized linear models; generalized estimating equation; geometric regression; goodness-of-fit; gradient; grouped logit; Hessian matrix; heterogeneity parameter; hierarchical; hurdle model; incidental parameter problem; incident rate ratio; incomplete beta; independent correlation; information matrix; iteratively re-weighted least squares; jackknife; La Grange multiplier; likelihood ratio test; LIMDEP; longitudinal; Marquardt; maximum likelihood; m-dependent; multilevel negative binomial; NB-1; NB-2; NB-C; NB-P; NB-H; heterogeneous negative binomial regression; overdispersion; left truncated; right truncated; negative exponential; Newton-Raphson algorithm; nonstationary correlation; outliers; Pearson residual; Pearson chi square; Poisson-gamma mixture; probit; generalized Poisson regression; quasi-deviance; quasi-likelihood; random coefficients; random effects; random intercept; rate parameter; residual analysis; robust variance estimator; sample selection; scale; score test; Stata; stationary correlation; unconditional fixed effect; underdispersion; unstructured correlation; zero-inflated count model; zero-inflated negative binomial; zero-truncated model; count response model

Software:

Stata

Full Text: DOI