Quintero-Sarmiento, Adrián; **Cepeda-Cuervo, Edilberto**; **Núñez-Antón, Vicente**

Estimating infant mortality in Colombia: some overdispersion modelling approaches. (English) Zbl 1514.62260

J. Appl. Stat. 39, No. 5, 1011-1036 (2012).

Summary: It is common to fit generalized linear models with binomial and Poisson responses, where the data show a variability that is greater than the theoretical variability assumed by the model. This phenomenon, known as overdispersion, may spoil inferences about the model by considering significant parameters associated with variables that have no significant effect on the dependent variable. This paper explains some methods to detect overdispersion and presents and evaluates three well-known methodologies that have shown their usefulness in correcting this problem, using random mean models, quasi-likelihood methods and a double exponential family. In addition, it proposes some new Bayesian model extensions that have proved their usefulness in correcting the overdispersion problem. Finally, using the information provided by the National Demographic and Health Survey 2005, the departmental factors that have an influence on the mortality of children under 5 years and female postnatal period screening are determined. Based on the results, extensions that generalize some of the aforementioned models are also proposed, and their use is motivated by the data set under study. The results conclude that the proposed overdispersion models provide a better statistical fit of the data.

MSC:

62P10 Applications of statistics to biology and medical sciences; meta analysis
62J12 Generalized linear models (logistic models)

Keywords:

Bayesian approaches; envelope plot; generalized overdispersion models; infant mortality; overdispersion; postnatal period screening; random mean models

Software:

Stata; SPSS; PROC NLMIXED; xtnbreg

Full Text: DOI

References:

[1] Aitkin, M. 1999. A general maximum likelihood analysis of variance components in generalized linear models. \textit{Biometrics}, 55(1): 117-128. · Zbl 1059.62564 · doi:10.1111/j.0006-341X.1999.00117.x

[2] Atkinson, A. C. 1985. \textit{Plots, Transformations and Regression}, Oxford: Clarendon Press.

[3] Berk, R. A. and MacDonald, J. M. 2008. Overdispersion and Poisson regression. \textit{J. Quant. Criminol}, 24(3): 269-285.· doi:10.1007/s10940-008-9048-4

[4] Breslow, N. 1984. Extra-Poisson variation in log-linear models. \textit{Appl. Stat}, 31: 38-44.· doi:10.2307/2347661

[5] Breslow, N. E. 1990. Test of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. \textit{J. Amer. Statist. Assoc}, 85(410): 565-571.

[6] Breslow, N. E. and Clayton, D. G. 1993. Approximate inference in generalized linear mixed models. \textit{J. Amer. Statist. Assoc}, 88(421): 9-25. · Zbl 0775.62195

[7] Cepeda, E. and Achcar, J. 2010. Heteroscedastic nonlinear regression models. \textit{Comm. Statist. Simulation Comput}, 39: 405-419. · Zbl 1183.62045

[8] Cepeda, E. and Gamerman, D. 2001. Bayesian modeling of variance heterogeneity in normal regression models. \textit{Braz. J. Probab. Stat}, 14(1): 207-221. · Zbl 0983.62013

[9] Cepeda, E. and Gamerman, D. 2005. Bayesian methodology for modeling parameters in the two parameter exponential family. \textit{Estadística}, 57(168-169): 93-105.

[10] N.-T. Chou and D. Steenhard, \textit{A flexible count data regression model using SAS PROC NLMIXED}, SAS Global Forum 2009, Statistics and Data Analysis. Available at http://support.sas.com/resources/papers/proceedings09/250-2009.pdf

[11] Cox, D. R. 1983. Some remarks on overdispersion. \textit{Biometrika}, 70(1): 269-274. · Zbl 0511.62007 · doi:10.1093/biomet/70.1.269

[12] Crowder, M. J. 1978. Beta-binomial anova for proportions. \textit{Appl. Stat}, 27(4): 34-37.· doi:10.2307/2346223

[13] Crowder, M. J. 1985. Gaussian estimation for correlated binomial data. \textit{J. R. Stat. Soc. Ser. B}, 47(2): 229-237.

[14] Dean, C. D. 1992. Testing for overdispersion in Poisson and binomial regression models. \textit{J. Amer. Statist. Assoc}, 87(418): 451-457.

[15] Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. \textit{J. R. Stat. Soc. Ser. B}, 39(1): 1-38. · Zbl 0364.62022

[16] Efron, B. 1986. Double exponential families and their use in generalized linear regression. \textit{J. Amer. Statist. Assoc}, 81(395): 709-721. · Zbl 0611.62072

[17] Gordon, K. S. 1989. Generalized linear models with varying dispersion. \textit{J. R. Stat. Soc. Ser. B}, 51(1): 47-60.

[18] Guimaraes, P. 2005. A simple approach to fit the beta binomial model. \textit{STATA J}, 5(3): 385-394.

[19] Hinde, J. and Demétrio, C. G.B. 1998. Overdispersion: Models and estimation. \textit{Comput. Stat. Data Anal}, 27: 151-170. · Zbl 1042.62578 · doi:10.1016/S0167-9473(98)00007-3

[20] IBM Software Business Analytics. 2010. \textit{Make Smarter Decisions with Your Nested Data. Using Generalized Linear Mixed Models with Continuous and Categorical Targets}, \(IBM^© SPSS^©\) Statistics. Available at ftp://public.dhe.ibm.com/common/ssi/ecm

[21] Johnson, C. C., Ownby, D. R., Alford, S. H., Havstad, S. L., Williams, L. K., Zoratti, E. M., Peterson, E. L. and Joseph, C. L.M. 2005. Antibiotic exposure in early infancy and risk for childhood atopy. \textit{J. Allergy Clin. Immunol}, 115(6): 1218-1224.· doi:10.1016/j.jaci.2005.04.020

[22] Lawless, J. F. 1987. Negative binomial regression model. \textit{Canad. J. Statist}, 15(3): 209-225. · Zbl 0632.62060 · doi:10.2307/3314912

[23] Lindsey, J. K. 1999. On the use of corrections for overdispersion. \textit{Appl. Stat}, 48(4): 553-561.· doi:10.1111/1467-9876.00171

[24] Margolin, B. H., Kaplan, N. and Zeiger, E. 1981. Statistical analysis of the Ames Salmonella/ mocrosome test. \textit{Proc. Natl. Acad. Sci}, 76: 3779-3783.· doi:10.1073/pnas.78.6.3779

[25] Mathew, J. L. 2004. Effect of maternal antibiotics on breast feeding infants. \textit{Postgrad. Med. J.}, 80: 196-200.· doi:10.1136/pgmj.2003.011973

[26] McCullagh, P. and Nelder, J. A. 1989. \textit{Generalized Linear Models}, 2, London: Chapman and Hall. · Zbl 0588.62104 · doi:10.1007/978-1-4899-3242-6

[27] Nelder, J. A. and Pregibon, D. 1987. An extended quasi-likelihood function. \textit{Biometrika}, 74(2): 221-232. · Zbl 0621.62078 · doi:10.1093/biomet/74.2.221

[28] Nelder, J. and Wedderburn, R. W.M. 1972. Generalized linear models. \textit{J. R. Stat. Soc. Ser. A}, 135(3): 370-384.· doi:10.2307/2344614

[29] Paula, A. P. 2004. \textit{Modelos de Regressao con Apoio Computacional}, 1, Sao Paulo: Universidade de Sao Paulo.

[30] Rabe-Hesketh, S. and Skrondal, A. 2005. \textit{Multilevel and Longitudinal Modeling Using Stata}, 2, College Station, TX: Stata Press Book. · Zbl 1274.62031

[31] Wedderburn, R. W.M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. \textit{Biometrika}, 61(3): 439-447. · Zbl 0292.62050

[32] Williams, D. A. 1975. The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. \textit{Biometrics}, 31(4): 949-952. · Zbl 0333.62069 · doi:10.2307/2529820

[33] Williams, D. A. 1982. Extra-binomial variation in logistic linear models. \textit{Appl. Stat}, 31(2): 144-188. · Zbl 0488.62055 · doi:10.2307/2347977

[34] Wolfinger, R. D. \textit{Fitting nonlinear mixed models with the new NLMIXED procedure}. Proceedings of the Twenty-Fourth Annual SAS Users Group Conference. Cary, NC: SAS Institute Inc.