The binomTools package: Performing model diagnostics on binomial regression models

Merete K Hansen[1,*], Rune Haubo B Christensen[1]

1. Technical University of Denmark, Department of Informatics and Mathematical Modelling, Section for Mathematical Statistics, Richard Petersens Plads, Building 305, Room 122, DK-2800 Kgs. Lyngby, Denmark
*Contact author: mkh@imm.dtu.dk

Keywords: Binomial regression models, diagnostics

Binomial regression models are widely used for modelling observations with dichotomous outcome. Since diagnostics are important for validation of model adequacy, the accessibility of suitable diagnostic methods for binomial regression models is crucial. The **binomTools** package (Hansen and Christensen, 2011) provides a range of diagnostic methods for this class of models extending the basic tool set in base *R* and enabling a thorough examination of the fitted model.

Appropriate use of the deviance or Pearson's X^2 reported by `glm` for goodness-of-fit assessment requires that observations are suitably grouped, e.g., binary observations should be grouped to binomial form. Our `group` method for glm objects performs this otherwise cumbersome task. Additional goodness-of-fit tests for binary data available in **binomTools** include the Hosmer-Lemeshow test.

Residual analysis in binary and binomial models is complicated by the non-unique definition of residuals and the sparseness of appropriate graphical methods. `rstudent` from **stats** provide approximate deletion residuals, the so-called likelihood residuals (Williams, 1987); **binomTools** enhance this by providing the exact counterparts. The half-normal plot with simulated envelopes implemented in our `halfnorm` function is described by Collett (2003) and one of the most effective and sensitive residual plots for the notoriously difficult binary models. Further, the *coefficient of discrimination* (Tjur, 2008) and related figures are implemented for summarizing and describing predictive performance.

Our aim with **binomTools** is to provide a comprehensive collection of methods for residual analysis, model diagnostics and presentation of results from binomial regression models.

References:

Collett, D. (2003). *Modelling binary data* (Second edition). Chapman & Hall/CRC.

Hansen, M. K. and Christensen, R. H. B. (2011). binomTools – diagnostic tools for binomial regression models. 1.0.1. https://r-forge.r-project.org/projects/binomtools/.

Tjur, T. (2008). Coefficients of determination in logistic regression models – a new proposal: The coefficient of discrimination. *The American statistician 63*(4), 366-372.

Williams, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Applied Statistics 36*, 181-191.