

Publication

Visualizing count data regressions using rootograms

JournalArticle (Originalarbeit in einer wissenschaftlichen Zeitschrift)

ID 3499374

Author(s) Kleiber, Christian; Zeileis, Achim

Author(s) at UniBasel [Kleiber, Christian](#) ;

Year 2016

Title Visualizing count data regressions using rootograms

Journal The American Statistician

Volume 70

Number 3

Pages / Article-Number 296-303

Keywords rootogram, visualization, goodness of fit, count data, Poisson regression, negative binomial regression, hurdle model, finite mixture

The rootogram is a graphical tool associated with the work of J. W. Tukey that was originally used for assessing goodness of fit of univariate distributions. Here we extend the rootogram to regression models and show that this is particularly useful for diagnosing and treating issues such as overdispersion and/or excess zeros in count data models. We also introduce a weighted version of the rootogram that can be applied out of sample or to (weighted) subsets of the data, e.g., in finite mixture models. An empirical illustration revisiting a well-known data set from ethology is included, for which a negative binomial hurdle model is employed. Supplementary materials providing two further illustrations are available online: the first, using data from public health, employs a two-component finite mixture of negative binomial models, the second, using data from finance, involves underdispersion. An `proglang{R}` implementation of our tools is available in the `proglang{R}` package `pkg{countreg}`. It also contains the data and replication code. The rootogram is a graphical tool associated with the work of J. W. Tukey that was originally used for assessing goodness of fit of univariate distributions. Here we extend the rootogram to regression models and show that this is particularly useful for diagnosing and treating issues such as overdispersion and/or excess zeros in count data models. We also introduce a weighted version of the rootogram that can be applied out of sample or to (weighted) subsets of the data, e.g., in finite mixture models. An empirical illustration revisiting a well-known data set from ethology is included, for which a negative binomial hurdle model is employed. Supplementary materials providing two further illustrations are available online: the first, using data from public health, employs a two-component finite mixture of negative binomial models, the second, using data from finance, involves underdispersion. An R implementation of our tools is available in the R package `countreg`. It also contains the data and replication code.

Publisher Taylor & Francis

ISSN/ISBN 0003-1305

URL <http://dx.doi.org/10.1080/00031305.2016.1173590>

edoc-URL <http://edoc.unibas.ch/43816/>

Full Text on edoc Available;

Digital Object Identifier DOI 10.1080/00031305.2016.1173590

ISI-Number WOS:000381650800011

Document type (ISI) Article