Package 'overdisp'

May 9, 2026

Type Package

Title Overdispersion in Count Data Multiple Regression Analysis

Version 0.1.2

Maintainer Rafael Freitas Souza <fsrafael@usp.br>

Description Detection of overdispersion in count data for multiple regression analysis. Log-linear count data regression is one of the most popular techniques for predictive modeling where there is a non-negative discrete quantitative dependent variable. In order to ensure the inferences from the use of count data models are appropriate, researchers may choose between the estimation of a Poisson model and a negative binomial model, and the correct decision for prediction from a count data estimation is directly linked to the existence of overdispersion of the dependent variable, conditional to the explanatory variables. Based on the studies of Cameron and Trivedi (1990) <doi:10.1016/0304-4076(90)90014-K> and Cameron and Trivedi (2013, ISBN:978-1107667273), the overdisp() command is a contribution to researchers, providing a fast and secure solution for the detection of overdispersion in count data. Another advantage is that the installation of other packages is unnecessary, since the command runs in the basic R language.

License GPL (>= 2)

Encoding UTF-8

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Author Rafael Freitas Souza [cre],
Hamilton Luiz Correa [ctb],
A. Colin Cameron [aut],
Pravin Trivedi [aut]

Repository CRAN

Date/Publication 2023-07-04 15:00:02 UTC

Contents

overdisp 2

Index 4

`overdisp` *Overdispersion in Count Data Multiple Regression Analysis*

Description

Log-linear count data regression is one of the most popular techniques for predictive modeling where there is a non-negative discrete quantitative dependent variable. In order to ensure the inferences from the use of count data models are appropriate, researchers may choose between the estimation of a Poisson model and a negative binomial model, and the correct decision for prediction from a count data estimation is directly linked to the existence of overdispersion of the dependent variable, conditional to the explanatory variables. Based on the studies of Cameron and Trivedi (1990, 2013), the `overdisp()` command is a contribution to researchers, providing a fast and secure solution for the detection of overdispersion in count data. Another advantage is that the installation of other packages is unnecessary, since the command runs in the basic R language.

Usage

```
overdisp(x, dependent.position = NULL, predictor.position = NULL, sig = NULL)
```

Arguments

`x`	The user's dataset.
`dependent.position`	A single number that declares the position of the dependent variable in the user dataset.
`predictor.position`	A number, or a set of numbers, that declares the position of explanatory variables in the dataset.
`sig`	Optional. If the argument value is not provided, the 5% level of statistical significance will be used.

Details

The test for detecting overdispersion of count data proposed by Cameron and Trivedi (1990) is based on following equation, where H_0 is the equidispersion given by $Var(Y|X) = E(Y|X)$ as follows:

$$Var(Y|X) = E(Y|X) + \Phi[E(Y|X)]^2$$

which is similar to the variance function of the negative binomial model indicated by: $Var(Y_i) = u + \Phi u^2$, where $\Phi = 1/\Psi$ and $u_i = exp(\alpha + \beta_1 X_{1i} + \beta_2 X_{2i}) + ... + \beta_k X_{ki}$. For the test in highlighted expression, the significance of parameter Φ must be verified, in which $H_1 : \Phi > 0$ e $H_0 : \Phi = 0$.

For the detection of overdispersion in the count data, at a certain level of significance, Cameron and Trivedi (1990) postulated that a Poisson model should be estimated a priori. According to the

authors, after this, an auxiliary ordinary least squares (OLS) model should be estimated without the intercept, whose dependent variable $Y*$, given by expression $[(Y_i - \lambda_i)^2 - Y_i]/\lambda_i$, should be calculated using the fitted values of λ from the initially established Poisson model.

The auxiliary model given by $Y*_i = \beta\lambda_i$ should use λ as the sole predictor variable. After the estimation of the auxiliary model, Cameron and Trivedi (1990) recommend checking the p value from the Student's t-test for the predictor variable λ. In the cases where $P > |t| > sig$, equidispersion at a pre-established significance level is indicated; when $P > |t| <= sig$, overdispersion at a pre-established significance level is indicated.

Value

A list with class "htest" containing the following components:

`statistic`	the value of the Lambda t test score.
`p.value`	the p-value for the test.
`method`	the character string "Overdispersion Test - Cameron & Trivedi (1990)".
`data.name`	a character string giving the name(s) of the data.
`alternative`	the character string "overdispersion if lambda p-value is less than or equal to the stipulated significance level".

Author(s)

Rafael de Freitas Souza and Hamilton Luiz Correa.

References

Cameron, A. C. & Trivedi, P. K. (1990). Regression-Based Tests for Overdispersion in the Poisson Model. Journal of Econometrics, 46(3), 347-364. <doi:10.1016/0304-4076(90)90014-K>.

Cameron, A. C. & Trivedi, P. K. (2013). Regression Analysis of Count Data. 2nd ed. Cambridge: Cambridge University Press. ISBN:978-1107667273.

See also:

Blackburn, M. L. (2015). The Relative Performance of Poisson and Negative Binomial Regression Estimators. Oxford Bulletin of Economics ans Statistics, 77(4), 605-616. <doi:10.1111/obes.12074>.

Payne, E. H., Hardin, J. W., Egede, L. E., Ramakrishnan, V., Selassie, A. & Gebregziabher, M. (2015). Approaches for Dealing with Various Sources of Overdispersion in Modeling Count Data: Scale adjustment versus modeling. Statistical Methods in Medical Research, 26(4), 1802-1823. <doi:10.1177/0962280215588569>.

Smith, D & Faddy, M. (2016). Mean and Variance Modeling of Under and Overdispersed Count Data. Journal of Statistical Software, 69(6), 1-23. <doi:10.18637/jss.v069.i06>.

Examples

```
overdisp(warpbreaks, dependent.position = 1, predictor.position = 2:3)
```

Index

∗ **count-data**
overdisp, 2
∗ **overdipersion**
overdisp, 2

overdisp, 2