Package 'MNB'

May 7, 2026

Type Package

Title Diagnostic Tools for a Multivariate Negative Binomial Regression Model

Version 1.2.0

Date 2025-03-05

Author Jalmar Carrasco [aut, cre],
Cristian Lobos [aut],
Lizandra Fabio [aut]

Maintainer Jalmar Carrasco <carrascojalmar@gmail.com>

Description Diagnostic tools as residual analysis, global, local and total-local influence for the multivariate model from the random intercept Poisson generalized log gamma model are available in this package. Including also, the estimation process by maximum likelihood method, for details see Fabio, L. C; Villegas, C. L.; Carrasco, J.M.F and de Castro, M. (2023) <doi:10.1080/03610926.2021.1939380> and Fábio, L. C.; Villegas, C.; Mamun, A. S. M. A. and Carrasco, J. M. F. (2025) <doi:10.28951/bjb.v43i1.728>.

License GPL (>= 2)

LazyData TRUE

Imports flexsurv, numDeriv, graphics, methods,

Suggests stats,

Depends R (>= 3.5.0)

NeedsCompilation no

RoxygenNote 7.3.2

Encoding UTF-8

Repository CRAN

Date/Publication 2025-03-06 00:00:16 UTC

Contents

`alzheimer` *Alzheimer data*

Description

The Alzheimer's data is presented in Hand and Taylor (1987) and Hand and Crowder (1996) to assess deterioration aspects of intellect, self-care and personality in senile patients with Alzheimer's disease. Two groups of patients were compared, one of which received a placebo and the other treatment with lecithin. In the data, each of the subjects, 26 in the placebo group and 22 in the lecithin group, were measured on five occasions (initially, 1st, 2nd, 4th and 6th). The measurements were the number of words that the patients could recalled from lists of words.

Usage

```
data(alzheimer)
```

Format

This data frame contains the following columns:

- Y: The number of words that the patients could recalled from lists of words.
- trt: Placebo ano lecithin groups.
- ind: Indicator on the ith patient.
- time: initially, 1st, 2nd, 4th and 6th visit.

References

- Hand, D. J. and Crowder, M. (1996). Practical Longitudinal Data Analysis. London: Chapman and Hall.
- Hand, D. J. and Taylor, C. C. (1987). Analysis of Variance and Repeated Measures. London: Chapman and Hall.

- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.
- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

Examples

```
data(alzheimer)
head(alzheimer)
```

`envelope.MNB` *Simulation envelope*

Description

Simulated envelopes in normal probability plots

Usage

```
envelope.MNB(star, formula, dataSet, n.r, nsim, plot = TRUE)
```

Arguments

`star`	Initial values for the parameters to be optimized over.
`formula`	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones).
`dataSet`	data
`n.r`	Indicator which residual type graphics. 1 - weighted, 2 - Standardized weighted, 3 - Pearson, 4 - Standardized Pearson, 5 - standardized deviance component residuals and 6 - randomized quantile residuals.
`nsim`	Number of Monte Carlo replicates.
`plot`	TRUE or FALSE. Indicates if a graph should be plotted.

Details

Atkinson (1985), suggests the use of simulated envelopes in normal probability plots to facilitate the goodness of fit.

Value

L, residuals and simulation envelopes in normal probability plots

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Atkinson A.C. (1985). Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis. Oxford University Press, New York.
- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.
- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

Examples

```
data(seizures)
head(seizures)

star <-list(phi=1, beta0=1, beta1=1, beta2=1, beta3=1)

envelope.MNB(formula=Y ~ trt + period + trt:period +
offset(weeks),star=star,nsim=21,n.r=6,
dataSet=seizures,plot=FALSE)

data(alzheimer)
head(alzheimer)

star <- list(phi=10,beta1=2, beta2=0.2)
envelope.MNB(formula=Y ~ trat, star=star, nsim=21, n.r=6,
dataSet = alzheimer,plot=FALSE)
```

`fit.MNB` *Maximum likelihood estimation*

Description

Estimate parameters by quasi-Newton algorithms.

Usage

```
fit.MNB(star, formula, dataSet, tab = TRUE)
```

Arguments

star	Initial values for the parameters to be optimized over.
formula	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones).
dataSet	data
tab	Logical. Print a summary of the coefficients, standard errors and p-value for class "MNB".

Details

Method "BFGS" is a quasi-Newton method, specifically that published simultaneously in 1970 by Broyden, Fletcher, Goldfarb and Shanno. This uses function values and gradients to build up a picture of the surface to be optimized.

Value

Returns a list of summary statistics of the fitted multivariate negative binomial model.

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Fabio, L., Paula, G. A., and de Castro, M. (2012). A Poisson mixed model with nonormal random effect distribution. Computational Statistics and Data Analysis, 56, 1499-1510.
- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.
- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

Examples

```
data(seizures)
head(seizures)

star <-list(phi=1, beta0=1, beta1=1, beta2=1, beta3=1)

mod1 <- fit.MNB(formula=Y ~ trt + period +
trt:period + offset(log(weeks)), star=star, dataSet=seizures)

mod1

seizures49 <- seizures[-c(241,242,243,244,245),]
```

```
mod2 <- fit.MNB(formula=Y ~ trt + period +
trt:period + offset(log(weeks)), star=star, dataSet=seizures49)

mod2
```

`global.MNB` *Global influence*

Description

It performers influence analysis by a global influence to evaluate the impact on the parameter estimates when we remove a particular observation.

Usage

```
global.MNB(formula, star, dataSet, plot = TRUE)
```

Arguments

`formula`	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones).
`star`	Initial values for the parameters to be optimized over.
`dataSet`	data
`plot`	TRUE or FALSE. Indicates if a graph should be plotted.

Details

The function returns a list (L) with the generalized Cook distance, Likelihood displacement and index plot.

Value

L and graphics

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.

Examples

```
data(seizures)
head(seizures)

star <-list(phi=1, beta0=1, beta1=1, beta2=1, beta3=1)
global.MNB(formula=Y ~ trt + period +
trt:period + offset(log(weeks)),star=star,dataSet=seizures,plot=FALSE)
```

`local.MNB` *Local influence*

Description

It performes influence analysis by a local influence approach by Cook (1986). It is considering three perturbation schemes: Case weights, explanatory variable and dispersion parameter perturbation. Another procedure which considering is the total local curvature corresponding to the ith element approach by Lesaffre and Verbeke (1998).

Usage

```
local.MNB(star, formula, dataSet, schemes, cova, plot = TRUE)
```

Arguments

`star`	Initial values for the parameters to be optimized over.
`formula`	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones).
`dataSet`	data
`schemes`	Perturbation scheme. Possible values: "cases" for Case weights perturbation on ith subject or cluster, "cases.obs" for Case weights perturbation on jth measurement taken on the ith subject or cluster, "cova.pertu" for explanatory variable perturbation, "dispersion" for dispersion parameter perturbation
`cova`	Indicator which column from dataset (continuous covariate) must be perturbation.
`plot`	TRUE or FALSE. Indicates if a graph should be plotted.

Details

The function returns a list (L) with the eigenvector associated with the maximum curvature, the total local influence and the index plot.

Value

L and graphics

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Cook, R. D. (1986). Assessment of local influence (with discussion). Journal of the Royal Statistical Society B, 48, 133-169.
- Lesaffre E. and Verbeke G. (1998). Local influence in linear mixed models. Biometrics, 54, 570-582.
- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.

Examples

```
data(seizures)
head(seizures)

star <-list(phi=1, beta0=1, beta1=1, beta2=1, beta3=1)

local.MNB(formula=Y ~ trt + period + trt:period + offset(log(weeks)),star=star,dataSet=seizures,
schemes="weight",plot=FALSE)

local.MNB(formula=Y ~ trt + period + trt:period + offset(log(weeks)),star=star,dataSet=seizures,
schemes="weight.obs",plot=FALSE)

local.MNB(formula=Y ~ trt + period + trt:period + offset(log(weeks)),star=star,dataSet=seizures,
schemes="dispersion",plot=FALSE)
```

MNB *Diagnostic tools for a multivariate negative binomial model*

Description

Diagnostic tools as residual analysis, global, local and total-local influence for the multivariate model from the random intercept Poisson-GlG mode. Including also, the estimation process by maximum likelihood and generating multivariate negative binomial data.

MNB package functions

- rMNB
- fit.MNB
- qMNB
- re.MNB
- envelope.MNB
- global.MNB
- local.MNB

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.
- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

---

qMNB *Randomized quantile residual*

Description

randomized quantile residual is available to assess possible departures from the multivariate negative binomial model for fitting correlated data with overdispersion.

Usage

```
qMNB(par, formula, dataSet)
```

Arguments

par	the maximum likelihood estimates.
formula	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones).
dataSet	data

Details

The randomized quantile residual (Dunn and Smyth, 1996), which follow a standard normal distribution is used to assess departures from the multivariate negative binomial model.

Value

Randomized quantile Residuals

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5, 236-244.
- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.
- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

Examples

```
data(seizures)
head(seizures)

star <-list(phi=1, beta0=1, beta1=1, beta2=1, beta3=1)
mod <- fit.MNB(formula=Y ~ trt + period +
trt:period + offset(log(weeks)),star=star,dataSet=seizures,tab=FALSE)
par <- mod$par
names(par)<-c()

res.q <- qMNB(par=par,formula=Y ~ trt + period + trt:period +
offset(log(weeks)),dataSet=seizures)

plot(res.q,ylim=c(-3,4.5),ylab="Randomized quantile residual",
xlab="Index",pch=15,cex.lab = 1.5, cex = 0.6, bg = 5)
abline(h=c(-2,0,2),lty=3)
#identify(res.q)


data(alzheimer)
head(alzheimer)

star <- list(phi=10,beta1=2, beta2=0.2)
mod <- fit.MNB(formula = Y ~ trat, star = star, dataSet = alzheimer,tab=FALSE)

par<- mod$par
names(par) <- c()
re.q <- qMNB(par=par,formula = Y ~ trat, dataSet = alzheimer)
head(re.q)
```

re.MNB *Residual analysis*

Description

Weighted, standardized weighted, Pearson, standardized Pearson and standardized deviance component residuals are available to assess possible departures from the multivariate negative binomial model for fitting correlated data with overdispersion.

Usage

```
re.MNB(star, formula, dataSet)
```

Arguments

star	Initial values for the parameters to be optimized over.
formula	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones).
dataSet	data

Details

Similarly to GLMs theory (Agresti, 2015; Faraway, 2016), weighted and the standardized weighted residuals are deduced trough Fisher scoring iterative process. Based in the Pearson residual, Fabio (2017) suggest the standardized Pearson residuals for the multivariate model from the random intercept Poisson-GLG model. In addition, it is available the standardized deviance component residual for the ith subject (Fabio et al., 2012).

Value

Residuals

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

References

- Agresti, A. (2015). Foundations of Linear and Generalized Linear Models. Wiley.
- Faraway, F. (2016). Extending the Linear Model with R: Generalized Linear, Mixed Effects and nonparametric regression models. Taylor & Francis, New York.
- Fabio, L., Paula, G. A., and de Castro, M. (2012). A Poisson mixed model with nonormal random effect distribution. Computational Statistics and Data Analysis, 56, 1499-1510.
- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.

- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

Examples

```
data(seizures)
head(seizures)

star <-list(phi=1, beta0=1, beta1=1, beta2=1, beta3=1)

r <- re.MNB(formula=Y ~ trt + period + trt:period +
offset(weeks),star=star,dataSet=seizures)

plot(r$ij.Sweighted.residual,cex.axis = 1.2, cex.lab = 1.2,
pch = 15,cex = 0.6, bg = 5,ylab="weighted.residual")

abline(h=c(-3,0,3),lwd = 2, lty = 2)

data(alzheimer)
head(alzheimer)

star <- list(phi=10,beta1=2, beta2=0.2)
r <- re.MNB(formula = Y ~ trat,star=star,dataSet=alzheimer)
names(r)
```

---

rMNB *Generating Multivariate Negative Binomial Data*

Description

It simulates a multivariate response variable, Y_{ij}, that is jth measurement taken on the ith subject or cluster, i = 1,...,n and j= 1,...,mi.

Usage

```
rMNB(n, mi, formula, p.fix)
```

Arguments

n	Length of the sample.
mi	replicates on the ith subject or cluster.
formula	The structure matrix of covariates of dimension n x p (in models that include an intercept x should contain a column of ones)
p.fix	Vector of theoretical regression parameters of length p.

Value

Generated response (Y_{ij})

Author(s)

Jalmar M F Carrasco <carrascojalmar@gmail.com>, Cristian M Villegas Lobos <master.villegas@gmail.com> and Lizandra C Fabio <lizandrafabio@gmail.com>

Examples

```
n <- 100
mi <- 3
x1 <- rep(rnorm(n,0,1),each=mi)
x2 <- rep(c(0,1),each=150)
p.fix <- c(10,2.0,0.5,1)

#generating a sample
sample.ex <- rMNB(n=n,mi=mi,formula=~x1+x2, p.fix=p.fix)
head(sample.ex)
```

---

seizures *Seizures data*

Description

The data set described in Diggle et.al (2013) refers to an experiment in which 59 epileptic patients were randomly assigned to one of two treatment groups: treatment (progabide drug) and placebo groups. The number of seizures experienced by each patient during the baseline period (week eight) and the four consecutive periods (every two weeks) was recorded. The main goal of this application is to analyze the drug effect with respect to the placebo. Two dummies covariates are considered in this study; Group which assumes values equal to 1 if the patient belongs to treatment group and 0 otherwise, and Period which assumes values equal to 1 if the number of seizures are recorded during the treatment and 0 if are measured in the baseline period. It is taking into account the Time covariate which represents the number of weeks required for the counting of seizures in each patient of the placebo and treatment groups.

Usage

```
data(seizures)
```

Format

This data frame contains the following columns:

- Y: The number epileptic seizure.
- trt: Treatment: binary indicators for the prograbide and placebo groups.
- period: binary indicator for the baseline period.

- week: number od weeks
- ind: Indicator on the ith patient.

References

- Diggle, P. J., Liang, K. Y., and Zeger, S. L. (2013). Analysis of Longitudinal Data. Oxford University Press, N.Y., 2 edition.
- Fabio, L. C., Villegas, C., Carrasco, J. M. F., and de Castro, M. (2023). Diagnostic tools for a multivariate negative binomial model for fitting correlated data with overdispersion. Communications in Statistics - Theory and Methods, 52, 1833–1853.
- Fabio, L. C., Villegas, C., Mamun, A. S., and Carrasco, J. M. F. (2025). Residual analysis for discrete correlated data in the multivariate approach. Brazilian Journal of Biometrics, 43, e43728.

Examples

```
data(seizures)
head(seizures)
```

Index