

Package ‘RAZIAD’

May 2, 2024

Title Regression Analysis of Zero-Inflated and Zero-Altered(Hurdle) Data

Version 0.0.1

Maintainer Niloufar Dousti Mousavi <niloufar.dousti@gmail.com>

Description It performs regression analysis for zero-inflated and zero-altered data, enabling modeling of sparse response variable distributions and assessing their association with covariates. It computes Maximum Likelihood Estimates (MLE) and conducts model selection using AIC and BIC criteria. Additionally, it calculates Fisher Information, confidence intervals, standard errors, and z-scores for all model parameters.

Encoding UTF-8

RoxygenNote 7.3.1

Imports Matrix, extraDistr, stats

Depends R (>= 4.3.1.0)

LazyData true

Suggests testthat (>= 3.0.0)

Config/testthat/edition 3

License MIT + file LICENSE

NeedsCompilation no

Author Niloufar Dousti Mousavi [aut, cre, cph],
Hani Aldirawi [aut, cph],
Donald Hedeker [aut, cph],
Jie Yang [aut, cph]

Repository CRAN

Date/Publication 2024-05-02 15:12:43 UTC

R topics documented:

DebTrivedi	2
fisher.reg	3
reg.model	8

Index	13
--------------	-----------

DebTrivedi

DebTrivedi Dataset

Description

The DebTrivedi dataset contains data on health care utilization and insurance coverage.

Usage

```
data("DebTrivedi")
```

Format

A data frame with 4406 observations on the following 19 variables.

ofp a numeric vector indicating physicians office visits

ofnp a numeric vector

opp a numeric vector

opnp a numeric vector

emer a numeric vector

hosp a numeric vector indicating number of hospital stays

health a factor with levels poor average excellent indicating self-perceived health status

numchron a numeric vector indicating number of chronic conditions

adldiff a factor with levels no yes

region a factor with levels midwest noreast other west

age a numeric vector

black a factor with levels no yes

gender a factor with levels female male

married a factor with levels no yes

school a numeric vector indicating number of years of education

faminc a numeric vector

employed a factor with levels no yes

privins a factor with levels no yes indicating private insurance indicator

medicaid a factor with levels no yes

Details

Deb and Trivedi (1997) conducted an analysis on data comprising 4406 individuals aged 66 and above, enrolled in Medicare, a public insurance program. These data were initially sourced from the US National Medical Expenditure Survey (NMES) for the years 1987/88. They are accessible through the data archive of the Journal of Applied Econometrics at <http://qed.econ.queensu.ca/jae/1997-v12.3/deb-trivedi/>. Additionally, the dataset was compiled for an R package associated with Kleiber and Zeileis (2008) and is also retrievable as `DebTrivedi.rda` from the Journal of Statistical Software, alongside Zeileis (2006). The primary aim is to develop models for medical care demand, encompassing physician/non-physician office and hospital outpatient visits, based on available patient covariates.

Source

<http://www.jstatsoft.org/v27/i08/paper>

References

Zeileis, A. and Kleiber, C. and Jackma, S. (2008). "Regression Models for Count Data in R". *JSS* 27, 8, 1–25.

Examples

```
data(DebTrivedi)
plot(table(DebTrivedi$ofp), main = "physicians office visit (N=4406)", ylab="frequency")
#the response variable "physicians office visits" is highly right skewed
```

fisher.reg

Fisher Information computation for calculating the confidence intervals in Zero-inflated and Zero-altered regression models

Description

Fisher Information computation for calculating the confidence intervals in Zero-inflated and Zero-altered regression models

Usage

```
fisher.reg(x, y, b0 = NULL, m, dist = "ZIP", link = "logit")
```

Arguments

<code>x</code>	a design matrix containing an intercept column (all ones) along with other available covariates for the response variable
<code>y</code>	a zero-inflated or zero-altered(Hurdle) count response variable, represented as an integer vector
<code>b0</code>	the initial parameters for the model, calculated as the product of the number of parameters in the specified models and the number of covariates

<code>m</code>	M set in trigamma free approach only needed for ZIBNB, BNBH, ZINB, and NBH
<code>dist</code>	can be specified as follows: "ZIP" for "zero-inflated Poisson", "ZINB" for "zero-inflated negative binomial", "ZINB-r" for "zero-inflated negative binomial with fixed r", "ZIBNB" for "zero-inflated beta negative binomial", "ZIBB" for "zero-inflated beta binomial", "ZIBB-n" for "zero-inflated beta binomial with fixed n", "ZIBB-ab" for "zero-inflated beta binomial with fixed alpha and beta", "PH" for "zero-altered(hurdle) Poisson", "NBH" for "zero-altered(Hurdle) negative binomial", "NBH-r" for "zero-altered (Hurdle) negative binomial with fixed r", "BNBH" for " zero-altered (Hurdle) beta negative binomial", "BBH" for "zero-altered (Hurdle) beta binomial", "BBH-n" for "zero-altered(Hurdle) beta binomial with fixed n", and "BBH-ab" for "zero-altered(Hurdle) beta binomial with fixed alpha and beta".
<code>link</code>	can be set to one of four different options: "logit" for the logistic link function, "probit" for the probit link function, "loglog" for the log-log link function, and "cloglog" for the complementary log-log link function

Value

If `dist = ZIP`, the following values are returned:

- `FisherInformation`: Fisher Information matrix for all the parameters and covariates in the model.
- `ConfidenceIntervals`: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If `dist = ZINB`, the following values are returned:

- `FisherInformation`: Fisher Information matrix for all the parameters and covariates in the model.
- `ConfidenceIntervals`: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If `dist = ZINB-r`, the following values are returned:

- `FisherInformation`: Fisher Information matrix for all the parameters and covariates in the model.
- `ConfidenceIntervals`: Contains the following information:
 - Lower and upper bounds of the confidence interval.

- Estimated parameters.
- Estimation/length ratio.
- Standard error.
- Z-score.

If dist = ZIBNB, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = ZIBB, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = ZIBB-n, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = ZIBB-ab, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.

- Standard error.
- Z-score.

If dist = PH, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = NBH, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = NBH-r, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = BNBH, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = BBH, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = BBH-n, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

If dist = BBH-ab, the following values are returned:

- FisherInformation: Fisher Information matrix for all the parameters and covariates in the model.
- ConfidenceIntervals: Contains the following information:
 - Lower and upper bounds of the confidence interval.
 - Estimated parameters.
 - Estimation/length ratio.
 - Standard error.
 - Z-score.

Examples

```

intercept<- rep(1,4406)
dt = DebTrivedi[, c(6:8, 13, 15, 18)]
dt = cbind(intercept, dt)
dt$gender.male <- ifelse(dt$gender == 'male', 1, 0)
dt$gender.female <- ifelse(dt$gender == 'female', 1, 0)
dt$health.poor <- ifelse(dt$health == 'poor', 1, 0)
dt$health.average <- ifelse(dt$health == 'average', 1, 0)
dt$health.excellent <- ifelse(dt$health == 'excellent', 1, 0)
dt$privins.yes <- ifelse(dt$privins == 'yes', 1, 0)
dt$privins.no <- ifelse(dt$privins == 'no', 1, 0)
y = DebTrivedi[,1]
x = data.matrix(dt[, c(1, 2, 4, 6, 8, 10, 12, 13)])
np = dim(x)[2]

```

```
b0 = c (0.1,0.10,0.1,0.1,0.1,0.1,0.1,0.1,
0.2,0.20,0.20,0.20,0.20,0.20,0.20,0.20)
fisher.reg(x, y, b0=b0, m=1e3, dist="ZIP", link="probit")
```

reg.model

Regression Analysis for Zero-altered or Zero-inflated Data

Description

Regression Analysis for Zero-altered or Zero-inflated Data

Usage

```
reg.model(x, y, b0 = NULL, dist = "ZIP", link)
```

Arguments

x	a design matrix containing an intercept column (all ones) along with other available covariates available and selected for the response variable
y	a zero-inflated or zero-altered(Hurdle) count response variable, represented as an integer vector
b0	the initial parameters for the model, calculated as the product of the number of parameters in the specified models and the number of covariates. For simplicity one may put the MLE and intercept of the parameters and set the rest of covariates to zero or change them.
dist	can be specified as follows: "ZIP" for "zero-inflated Poisson", "ZINB" for "zero-inflated negative binomial", "ZINB-r" for "zero-inflated negative binomial with fixed r", "ZIBNB" for "zero-inflated beta negative binomial", "ZIBB" for "zero-inflated beta binomial", "ZIBB-n" for "zero-inflated beta binomial with fixed n", "ZIBB-ab" for "zero-inflated beta binomial with fixed alpha and beta", "PH" for "zero-altered(hurdle) Poisson", "NBH" for "zero-altered(Hurdle) negative binomial", "NBH-r" for "zero-altered (Hurdle) negative binomial with fixed r", "BNBH" for " zero-altered (Hurdle) beta negative binomial", "BBH" for "zero-altered (Hurdle) beta binomial", "BBH-n" for "zero-altered(Hurdle) beta binomial with fixed n", and "BBH-ab" for "zero-altered(Hurdle) beta binomial with fixed alpha and beta".
link	can be set to one of four different options: "logit" for the logistic link function, "probit" for the probit link function, "loglog" for the log-log link function, and "cloglog" for the complementary log-log link function

Value

A list containing AIC, BIC, the corresponding value of log likelihood, and the maximum likelihood estimate (MLE) of the unknown parameters in the model. If dist = ZIP, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log.likelihood: The value of log likelihood with maximum likelihood estimate plugged-in.
- Estimated.Parameters: The maximum likelihood estimate of Γ and β for intercept and covariates included in the design matrix.

If dist = ZINB, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (Γ) , (β_1) , and (β_2) for intercept and covariates included in the design matrix.

If dist = ZINB-r, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (β_1) for intercept only and (Γ) , and (β_2) for intercept and covariates included in the design matrix.

If dist = ZIBNB, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (Γ) , (β_1) , (β_2) , and (β_3) for intercept and covariates included in the design matrix.

If dist = ZIBB, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.

- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (Γ) , (β_1) , (β_2) , and (β_3) for intercept and covariates included in the design matrix.

If dist = ZIBB-n, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (β_1) for intercept only and (Γ) , (β_2) , and (β_3) for intercept and covariates included in the design matrix.

If dist = ZIBB-ab, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (β_2) and (β_3) for intercept only and (Γ) and (β_1) for intercept and covariates included in the design matrix.

If dist = PH, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated.Parameters: The maximum likelihood estimate of Γ and β for intercept and covariates included in the design matrix.

If dist = NBH, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (Γ) , (β_1) , and (β_2) for intercept and covariates included in the design matrix.

If dist = NBH-r, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (β_1) for intercept only and (Γ) , and (β_2) for intercept and covariates included in the design matrix.

If dist = BNBH, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (Γ) , (β_1) , (β_2) , and (β_3) for intercept and covariates included in the design matrix.

If dist = BBH, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (Γ) , (β_1) , (β_2) , and (β_3) for intercept and covariates included in the design matrix.

If dist = BBH-n, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.
- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (β_1) for intercept only and (Γ) , (β_2) , and (β_3) for intercept and covariates included in the design matrix.

If dist = BBH-ab, the following values are returned:

- AIC: Akaike Information Criterion, a measure of the model's goodness of fit adjusted for the number of parameters.

- BIC: Bayesian Information Criterion, a criterion for model selection among a finite set of models.
- log-likelihood: The value of log-likelihood corresponding to the maximum likelihood estimate.
- Estimated Parameters: Maximum likelihood estimates of (β_2) and (β_3) for intercept only and (Γ) and (β_1) , for intercept and covariates included in the design matrix.

Examples

```
intercept<- rep(1,4406)
dt = DebTrivedi[, c(6:8, 13, 15, 18)]
dt = cbind(intercept, dt)
dt$gender.male <- ifelse(dt$gender == 'male', 1, 0)
dt$gender.female <- ifelse(dt$gender == 'female', 1, 0)
dt$health.poor <- ifelse(dt$health == 'poor', 1, 0)
dt$health.average <- ifelse(dt$health == 'average', 1, 0)
dt$health.excellent <- ifelse(dt$health == 'excellent', 1, 0)
dt$privins.yes <- ifelse(dt$privins == 'yes', 1, 0)
dt$privins.no <- ifelse(dt$privins == 'no', 1, 0)
y = DebTrivedi[,1]
x = data.matrix(dt[, c(1, 2, 4, 6, 8, 10, 12, 13)])
np = dim(x)[2]
b0 = c(rep(0.3, np), rep(0.1, np))
reg.model(x, y, b0=b0, dist="PH", link="logit")
```

Index

* datasets

DebTrivedi, 2

DebTrivedi, 2

fisher.reg, 3

reg.model, 8