Package ‘lbreg’

December 13, 2019

Type Package
Title Log-Binomial Regression with Constrained Optimization
Description Maximum likelihood estimation of log-binomial regression with special functionality when the MLE is on the boundary of the parameter space.
Version 1.3
Date 2019-12-13
Author Bernardo B. Andrade
Maintainer Bernardo Andrade <bbandrade@unb.br>
License GPL-2
Depends R (>= 3.2.0)
Imports MASS
NeedsCompilation no
Repository CRAN
Date/Publication 2019-12-13 14:20:06 UTC

R topics documented:

lbreg-package .................................................... 2
Birth ............................................................ 3
Caesarian ......................................................... 4
Death ............................................................. 5
Evans ............................................................. 6
Heart ............................................................... 7
HL_test ........................................................... 8
lbreg ............................................................... 9
PCS ................................................................. 11
predict.lbreg ..................................................... 12
relrisk ............................................................. 13

Index 15
Description

Maximum likelihood estimation of log-binomial regression with special functionality when the
MLE is on the boundary of the parameter space.

Package lbreg performs maximum likelihood estimation of Log-Binomial Regression. The main
functions are lbreg which provides a shortcut to constrOptim to estimate LBR coefficients and
relrisk which takes lbreg results to produce estimated relative risks and associated confidence
intervals and prediction. Results differ from glm when the MLE is on the boundary of the parameter
space as explained in the reference below (Andrade, Andrade (2018)).

Details

The DESCRIPTION file:

Package: lbreg
Type: Package
Title: Log-Binomial Regression with Constrained Optimization
Description: Maximum likelihood estimation of log-binomial regression with special functionality when the MLE is on the boundary of the parameter space.
Version: 1.3
Date: 2019-12-13
Author: Bernardo B. Andrade
Maintainer: Bernardo Andrade <bbandrade@unb.br>
License: GPL-2
Depends: R (>= 3.2.0)
Imports: MASS

Index of help topics:

Birth Birth Weight Data
Caesarian Caesarian Infection Dataset
Death Death Penalty Data
Evans Evans County dataset
HL_test Hosmer-Lemeshow Goodness of Fit Test
Heart Heart Dataset
PCS PCS Dataset
lbreg Log-Binomial regression
lbreg-package Log-Binomial Regression with Constrained Optimization
predict.lbreg Predict method for Log-Binomial regression.
relrisk Regression Adjusted Relative Risks
Birth Weight Data

Description
Data used by Wacholder (1986) to illustrate the use of log binomial regression for estimating adjusted relative risks of a low-birthweight baby.

Usage
```r
data("Birth")
```

Format
A data frame with 900 observations on the following 5 variables.

- `lowbw` low birth weight delivery (1=yes)
- `alc` mother’s alcohol drinking frequency (1=Light, 2=Moderate, 3=Heavy)
- `smo` mother smoked (1=no)
- `soc` mother’s social status (1=I and II (lower), 2=III (middle), 3=IV and V (upper))

Source
Stata’s online manual http://www.stata.com/manuals13/rbinreg.pdf

References
Caesarian

**Caesarian Infection Dataset**

**Description**

Adapted dataset from Fahrmeir et al (2013): grouped data on infections of 251 mothers after a C-section collected at the clinical center of the University of Munich.

**Usage**

data("Caesarian")

**Format**

A data frame with 7 rows and 5 variables.

- **n1**  Caesarians with infections.
- **n0**  Caesarians without infections.
- **NPLAN** = 1 if C-section was not planned.
- **RISK** = 1 if risk factors existed.
- **ANTIB** = 1 if antibiotics were administered as prophylaxis.

**Source**

http://www.uni-goettingen.de/de/551625.html

**References**


**Examples**

data(Caesarian)
Caesarian
# no observations for case (RISK=0, NPLAN=1, ANTIB=1)
y = Caesarian[,1:2]
cbind(Caesarian[,3:5], total=rowSums(y))
colSums(y)
**Death Penalty Data**

**Description**

See references.

**Usage**

```r
data("Death")
```

**Format**

A data frame with 147 observations on the following 6 variables.

- death  death = 1, life in prison = 0
- blackd  black defendant = 1
- whitvic  white victim = 1
- serious  a measure of crime seriousness
- culp  a measure of culpability
- serious2  another measure of crime seriousness

**Source**


**References**


**Examples**

```r
data(Death)
dim(Death)
names(Death)
```
Evans County dataset

Description

Data from cohort study in which white males in Evans County were followed for 7 years, with coronary heart disease as the outcome of interest.

Usage

data("Evans")

Format

A data frame with 609 observations on the following 9 variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDH</td>
<td>outcome variable; 1 = coronary heart disease</td>
</tr>
<tr>
<td>CAT</td>
<td>1 = high, 0 = normal catecholamine level</td>
</tr>
<tr>
<td>AGE</td>
<td>age (in years)</td>
</tr>
<tr>
<td>CHL</td>
<td>cholesterol, mg/dl</td>
</tr>
<tr>
<td>SMK</td>
<td>1 = subject has ever smoked</td>
</tr>
<tr>
<td>ECG</td>
<td>1 = presence of electrocardiogram abnormality</td>
</tr>
<tr>
<td>DBP</td>
<td>diastolic blood pressure, mmHg</td>
</tr>
<tr>
<td>SBP</td>
<td>systolic blood pressure, mmHg</td>
</tr>
<tr>
<td>HPT</td>
<td>1 = SBP greater than or equal to 160 or DBP greater than or equal to 95</td>
</tr>
</tbody>
</table>

Source

http://web1.sph.emory.edu/dkleinb/logreg3.htm#data

References


Examples

data(Evans)
dim(Evans)
names(Evans)
Heart Dataset

Description
Heart attack data from the ASSENT-2 study.

Usage
data("Heart")

Format
A data frame with 16,949 observations on the following 5 variables.

Heart binary response; 1 = death
age categorized into <65, 65-75 or >75 years
severity Killip class I, II, or III/IV
region code for three USA regions
onset treatment delay categorized into <2, 2-4 or >4 hours

Source
http://biostatistics.oxfordjournals.org/content/13/1/179/suppl/DC1

References

Examples
data(Heart)
dim(Heart)
names(Heart)
HL_test  

**Hosmer-Lemeshow Goodness of Fit Test**

**Description**

The HL decile-of-risk test. Validity of the test assumes that the number of covariate patterns is close to the number of observations which is violated when many observations have the same covariate pattern and several ties will impact the required ordering and grouping (by deciles) of observations. This is less likely when there is at least one continuous covariate. Not valid for grouped data.

**Usage**

```r
HL_test(object, g = 10)
```

**Arguments**

- `object` object of class 'lbreg'.
- `g` number of groups

**Value**

A list with elements

- `X2` HL statistic
- `pvalue` p-value for the test from Chi Squared with df = g-2

**Author(s)**

Bernardo B. Andrade

**References**


**See Also**

`lbreg`

**Examples**

```r
require(lbreg)

# data preparation
data(PCS)
w <- PCS
w <- w[, -1]
w$race <- factor(w$race)
```
lbreg <- factor(w$dpros)
w$dcaps <- factor(w$dcaps)

fm <- lbreg(tumor ~ ., data=w)
HL_test(fm)

---

### lbreg

#### Log-Binomial regression

**Description**

Fitting a Log-Binomial Regression Model

**Usage**

```r
lbreg(formula, data, start.beta, tol=0.9999, delta=1, ...)
```

**Arguments**

- **formula**: an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted.
- **data**: an optional data frame containing the variables in the model. If not found in data, the variables are taken from environment(formula), typically the environment from which lbreg is called.
- **start.beta**: starting values for the parameters in the linear predictor. If missing, the default value explained in Andrade and Andrade (2018) is used according to the choice of `delta`.
- **tol**: defaults to 0.9999; threshold for declaring a probability on the boundary (p = 1).
- **delta**: defaults to 1. See reference below.
- **...**: not used.

**Details**

This function uses `constrOptim` with the BFGS method in order to perform maximum likelihood estimation of the log-binomial regression model as described in the reference below. When the MLE is the interior of the parameter space results should agree with `glm(..., family=binomial(link='log'))`. `lbreg` uses the adaptive logarithmic barrier algorithm rather than iteratively weighted least squares (`glm`).

**Value**

- **Active**: matrix of active constraints.
- **barrier.value**: same as in `constrOptim`.
- **coefficients**: named vector of estimated regression coefficients.
- **convergence**: same as in `constrOptim`. 
call the matched call.
cook.distance Cook’s distance.
data the data argument.
deviance residual deviance.
dev.resid deviance residuals.
fitted.values fitted probabilities.
formula the formula supplied.
hat.matrix hat matrix for GLMs (whose diagonal contains leverage values).
loglik maximized loglikelihood.
outer.iterations same as in constrOptim.
residuals Pearson residuals.
se standard errors of estimated coefficients.
start.beta starting values used by constrOptim.
vcov variance-covariance matrix of estimates.
vcov0 inverse of observed Fisher information; should be equal to vcov if there are no active constraints (Active = NULL).
X2 sum of squared residuals (variance-inflation estimate (dispersion) = X2/df).

Author(s)
Bernardo B. Andrade

References

See Also

 glm (family=binomial(link='log')). relrisk

Examples

require(lbreg)

# data preparation
data(PCs)  # ungrouped data
w <- PCS
w <- w[,,-1]
w$race <- factor(w$race)
w$dpros <- factor(w$dpros)
w$dcaps <- factor(w$dcaps)

# log-binomial regression
fm <- lbreg(tumor ~ ., data=w)
```r
fm
coef(fm)
summary(fm)

# grouped data
require(lbreg)
data(Caesarian)
m1 <- lbreg( cbind(n1, n0) ~ RISK + NPLAN + ANTIB, data=Caesarian)
summary(m1)

# dispersion estimate based on deviance residuals
sum(m1$dev.res^2)
# dispersion estimate based on Pearson residuals (reported in the summary above)
sum(m1$residuals^2)/(8-4)
predict(m1, newdata=data.frame(RISK=0, NPLAN=1, ANTIB=1))

# m0 <- glm( cbind(n1, n0) ~ RISK + NPLAN + ANTIB, data=Dat, family=binomial(link='log'))
# summary(m0)
```

---

**PCS Dataset**

**Description**

Prostate Cancer Study

**Usage**

```r
data("PCS")
```

**Format**

A data frame with 380 observations on the following 9 variables.

- **id** Identification Code; 1 - 380
- **tumor** Tumor Penetration of Prostatic Capsule, 0 = No Penetration
- **age** in years
- **race** Race; 1 = White, 2 = Black
- **dpros** Results of the Digital Rectal Exam, 4 levels
- **dcaps** Detection of Capsular Involvement in Rectal Exam; 1 = No, 2 = Yes
- **psa** antigen mg/ml
- **vol** Tumor Volume Obtained from Ultrasound, cm3
- **gleason** Total Gleason Score; 0 - 10
**predict.lbreg**

Predict method for Log-Binomial regression.

**Description**

Predicted values based on 'lbreg' object.

**Usage**

```r
## S3 method for class 'lbreg'
predict(object, newdata, ...)
```

**Arguments**

- `object`: Object of class inheriting from "lbreg"
- `newdata`: a data frame with covariate values with which to predict. If omitted, the fitted probabilities are returned.
- `...`: not used

**Details**

If newdata is omitted the predictions are simply the fitted values stored in the object supplied.

**Value**

- `Active`: active restrictions (taking newdata into account).
- `coef.pred`: regression coefficients re-estimated to satisfy possibly new restrictions imposed by newdata. See reference below.
- `convergence`: same as in the object supplied.
- `se.pred`: estimated standard errors of predictions.
- `tol`: same as in the object supplied.
- `ypred`: predicted probabilities for newdata.

**Source**

https://www.umass.edu/statdata/statdata/data/pros.txt

**References**


**Examples**

```r
data(PCS)
## View(PCS)
## str(PCS) ; plot(PCS) ...
```
relrisk

Author(s)

Bernardo B. Andrade

References


Examples

```r
require(lbreg)

# data preparation
data(PCS)
w <- PCS
w <- w[,,-1]
w$race <- factor(w$race)
w$dpros <- factor(w$dpros)
w$dcaps <- factor(w$dcaps)

# log-binomial regression
fm <- lbreg(tumor ~ ., data=w)
novo <- data.frame(age=c(41, 32), race=c(1,2), dpros=c(2,4),
dcaps=c(1,1), psa=c(7.24,3.25), vol=c(4.3,5.6),
 gleason=c(2,8))
predict(fm, newdata=novo)
```

relrisk | Regression Adjusted Relative Risks

Description

This function calculates the relative risks RR adjusted for covariates (acting on a previous log-binomial regression fit) and confidence intervals (by default 95 percent) for the estimated RR. The confidence interval is calculated from the log(RR) and backtransformed.

Usage

```r
relrisk(object, alpha = 0.05, dispersion = FALSE)
```

Arguments

- **object**: object of class 'lbreg'.
- **alpha**: 1 - desired confidence level.
- **dispersion**: logical. TRUE if standard errors should be adjusted for dispersion estimate based on Pearson residuals.
Value

value table with estimated relative risks, lower and upper bounds of confidence intervals.

Author(s)

Bernardo B. Andrade

References


See Also

lbreg

Examples

require(lbreg)

# ungrouped data
# data preparation
data(PCS)
w <- PCS
w <- w[, -1]
w$race <- factor(w$race)
w$dplos <- factor(w$dplos)
w$dcaps <- factor(w$dcaps)

# log-binomial regression
fm <- lbreg(tumor ~ ., data=w)

# relative risks
relikr(fm)
relikr(fm, alpha=.10)

# grouped data
require(lbreg)
data(Caesarian)
m1 <- lbreg( cbind(n1, n0) ~ RISK + NPLAN + ANTIB, data=Caesarian)
relikr(m1)
relikr(m1, dispersion=TRUE)
Index

*Topic datasets
  Birth, 3
  Caesarian, 4
  Death, 5
  Evans, 6
  Heart, 7
  PCS, 11

Birth, 3
Caesarian, 4
constrOptim, 2, 9
Death, 5
Evans, 6
glm, 2, 10
Heart, 7
HL_test, 8
lbreg, 2, 8, 9, 14
lbreg-package, 2
PCS, 11
predict.lbreg, 12
relrisk, 2, 10, 13