

Package ‘lboxcox’

December 16, 2023

Type Package

Title Implementation of Logistic Box-Cox Regression

Version 1.2

Date 2023-12-12

Maintainer Li Xing <sfulxing@gmail.com>

Description Implements a logistic box-cox model. This model is fully described in Xing, L. et al. (2021) <[doi:10.1002/cjs.11587](https://doi.org/10.1002/cjs.11587)>.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

Suggests knitr, rmarkdown, testthat (>= 3.0.0)

VignetteBuilder knitr

Config/testthat/edition 3

Depends R (>= 3.50)

Imports survey, maxLik, doParallel, foreach, MASS, dplyr, R.utils

NeedsCompilation no

Author Li Xing [cre, aut],
Shiyu Xu [aut],
Jing Wang [aut],
Kohlton Booth [aut],
Xuekui Zhang [aut],
Igor Burstyn [aut],
Paul Gustafson [aut]

Repository CRAN

Date/Publication 2023-12-15 23:10:02 UTC

R topics documented:

box_cox_new	2
depress	3
lbc_maxlik	3
lbc_train	5
lboxcox_maxLik.predict	6
LogLikeFun	6
LogLikeFun_new	7
median_effect	7
ScoreFun	8
ScoreFun_new	8
svyglm_ms	9

Index	10
--------------	-----------

box_cox_new	<i>Box-Cox transform</i>
-------------	--------------------------

Description

This function processes the box cox transform on variables

Usage

```
box_cox_new(formula, mydata, ixx, lambda)
```

Arguments

formula	a formula of the form $y \sim x + z1 + z2$ where y is a binary response variable, x is a continuous predictor variable, and z1, z2, ... are covariates
mydata	dataset used in box cox transform
ixx	continuous predictor
lambda	lambda used in box cox transform

Value

data set after transform, contains transformed ixx

depress	<i>Depression dataset</i>
---------	---------------------------

Description

The depress data frame has 8,893 rows and 5 columns from the National Health and Nutrition Examination Survey (NHANES) 2009–2010.

Usage

```
depress
```

Format

Sample survey data

depression binary response variable indicating whether the participant has depression (=1) or not (=0)

mercury a numeric vector giving the log-transformed total blood mercury in micro-grams per litre

age 0 if participant is female and 1 if they are male

gender age of the participant

weight a numeric vector giving the sampling-weight.

Source

Xing, L., Zhang, X., Burstyn, I., & Gustafson, P. (2021). On logistic Box–Cox regression for flexibly estimating the shape and strength of exposure-disease relationships. *Canadian Journal of Statistics*, 49(3), 808-825.

lbc_maxlik	<i>Train a Logistic Box-Cox model using MaxLik</i>
------------	--

Description

Train the given formula using a Logistic Box-Cox model.

Usage

```
lbc_maxlik(  
  formula,  
  weight_column_name,  
  data,  
  init = NULL,  
  svy_lambda_vector = seq(0, 2, length = 4),  
  init_lambda_vector = seq(0, 2, length = 100),
```

```

    num_cores = 1,
    seed,
    iterlim,
    timelim
  )

```

Arguments

<code>formula</code>	a formula of the form $y \sim x + z1 + z2$ where y is a binary response variable, x is a continuous predictor variable, and $z1, z2, \dots$ are covariates
<code>weight_column_name</code>	the name of the column in 'data' containing the survey weights.
<code>data</code>	dataframe containing the dataset to train on
<code>init</code>	initial estimates for the coefficients. If NULL the svyglm model will be used
<code>svy_lambda_vector</code>	values of lambda used in training svyglm model. Best model is used for initial coefficient estimates. If <code>init</code> is not NULL this parameter is ignored.
<code>init_lambda_vector</code>	values of lambda used in finding the optimal lambda with the best loglikelihood
<code>num_cores</code>	the number of cores used when finding the best svyglm model. If <code>init</code> is not NULL this parameter is ignored.
<code>seed</code>	set seed for MaxLik function
<code>iterlim</code>	Maximum number of iterations of MaxLik
<code>timelim</code>	Maximum iteration time of MaxLik

Value

object of class 'maxLik' from the 'maxLik' package. Contains the coefficient estimates that maximizes likelihood among other statistics.

Note

This is reliant on the following work:

Henningsen, A., Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. Computational Statistics, 26(3), 443-458.

Microsoft Corporation, Weston, S. (2020). foreach: Provides Foreach Looping Construct. R package version 1.5.1.

Microsoft Corporation, Weston, S. (2020). doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.16.

lbc_train	<i>Train a Logistic Box-Cox model</i>
-----------	---------------------------------------

Description

Train the given formula using a Logistic Box-Cox model.

Usage

```
lbc_train(
  formula,
  weight_column_name,
  data,
  init = NULL,
  svy_lambda_vector = seq(0, 2, length = 100),
  num_cores = 1
)
```

Arguments

formula	a formula of the form $y \sim x + z1 + z2$ where y is a binary response variable, x is a continuous predictor variable, and $z1, z2, \dots$ are covariates
weight_column_name	the name of the column in 'data' containing the survey weights.
data	dataframe containing the dataset to train on
init	initial estimates for the coefficients. If NULL the svyglm model will be used
svy_lambda_vector	values of lambda used in training svyglm model. Best model is used for initial coefficient estimates. If init is not NULL this parameter is ignored.
num_cores	the number of cores used when finding the best svyglm model. If init is not NULL this parameter is ignored.

Value

object of class 'maxLik' from the 'maxLik' package. Contains the coefficient estimates that maximizes likelihood among other statistics.

Note

This is reliant on the following work:

Henningsen, A., Toomet, O. (2011). maxLik: A package for maximum likelihood estimation in R. Computational Statistics, 26(3), 443-458.

Microsoft Corporation, Weston, S. (2020). foreach: Provides Foreach Looping Construct. R package version 1.5.1.

Microsoft Corporation, Weston, S. (2020). doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.16.

```
lboxcox_maxLik.predict
```

Lboxcox MaxLik Prediction Function

Description

Give the predicted p value of given LBC MaxLik model

Usage

```
lboxcox_maxLik.predict(myMaxLikfit, newdata, formula)
```

Arguments

myMaxLikfit	Fitted model using lboxcox_maxLik model
newdata	Given data for prediction
formula	a formula of the form $y \sim x + z1 + z2$ where y is a binary response variable, x is a continuous predictor variable, and z1, z2, ... are covariates

Value

p value

Note

This is reliant on the following work:

```
LogLikeFun
```

Log Likelihood of Logistic Box-Cox

Description

This function gives the log likelihood of the Box-Cox model. Main purpose is to be an input to the maxLik function.

Usage

```
LogLikeFun(bb, ixx, iyy, iw, iZZ)
```

Arguments

bb	current values for the intercept and slope coefficients
ixx	continuous predictor
iyy	binary outcome
iw	sample weight
iZZ	covariates to be incorporated in the model

Value

the log likelihood estimate for the coefficients in 'bb'

LogLikeFun_new	<i>New Log Likelihood of Logistic Box-Cox</i>
----------------	---

Description

This function gives the log likelihood of the Box-Cox model. Main purpose is to be an input to the maxLik function.

Usage

```
LogLikeFun_new(bb, ixx, iyy, iw, iZZ)
```

Arguments

bb	current values for the intercept and slope coefficients
ixx	continuous predictor
iyy	binary outcome
iw	sample weight
iZZ	covariates to be incorporated in the model

Value

the log likelihood estimate for the coefficients in 'bb'

median_effect	<i>Calculates the "slope" of the Logistic Box-Cox model</i>
---------------	---

Description

Calculates a number that represents the overall gradient measurement between the predictor and log-odds of the risk

Calculates a number that represents the overall gradient measurement between the predictor and log-odds of the risk

Usage

```
median_effect(formula, weight_column_name, data, trained_model)
```

```
median_effect(formula, weight_column_name, data, trained_model)
```

Arguments

formula	the formula used to train the logistic box-cox model
weight_column_name	the name of the column in 'data' containing the survey weights
data	dataframe containing the dataset to train on
trained_model	the already trained model. The output of 'lbc_train'

ScoreFun

*Log Likelihood Gradient of Logistic Box-Cox***Description**

This function gives the gradient of the log likelihood of the Box-Cox model. Main purpose is to be an input to the maxLik function.

Usage

```
ScoreFun(bb, ixx, iyy, iw, iZZ)
```

Arguments

bb	initial values for the intercept and slope coefficients
ixx	continuous predictor
iyy	binary outcome
iw	sample weight
iZZ	covariates to be incorporated in the model

Value

the gradient of the log likelihood estimate for the coefficients in 'bb'

ScoreFun_new

*New Log Likelihood Gradient of Logistic Box-Cox***Description**

This function gives the gradient of the log likelihood of the Box-Cox model. Main purpose is to be an input to the maxLik function.

Usage

```
ScoreFun_new(init, ixx, iyy, iw, iZZ)
```


Arguments

init	initial values for the intercept and slope coefficients
ixx	continuous predictor
iyy	binary outcome
iw	sample weight
izz	covariates to be incorporated in the model

Value

the gradient of the log likelihood estimate for the coefficients in 'bb'

svyglm_ms	<i>Svyglm of MaxLik_ms</i>
-----------	----------------------------

Description

This function gives the initial value list used in MaxLik_ms function

Usage

```
svyglm_ms(
  formula,
  data,
  lambda_vector = seq(0, 2, length = 100),
  weight_column_name = NULL,
  num_cores = 1
)
```

Arguments

formula	formula used in model
data	dataframe containing the dataset to train on
lambda_vector	values of lambda used in training svyglm model.
weight_column_name	the name of the column in 'data' containing the survey weights.
num_cores	the number of cores used when finding the best svyglm model.

Value

initial value list used in MaxLik_ms function

Index

* datasets

depress, [3](#)

box_cox_new, [2](#)

depress, [3](#)

lbc_maxlik, [3](#)

lbc_train, [5](#)

lboxcox_maxLik.predict, [6](#)

LogLikeFun, [6](#)

LogLikeFun_new, [7](#)

median_effect, [7](#)

ScoreFun, [8](#)

ScoreFun_new, [8](#)

svyglm_ms, [9](#)