Package 'logistic4p'

October 13, 2022

Type Package

Title Logistic Regression with Misclassification in Dependent Variables

Version 1.5

Date 2017-05-30

Depends R (>= 2.10), MASS

Author Haiyan Liu and Zhiyong Zhang

Maintainer Zhiyong Zhang <johnnyzhz@gmail.com>

Description Error in a binary dependent variable, also known as misclassification, has not drawn much attention in psychology. Ignoring misclassification in logistic regression can result in misleading parameter estimates and statistical inference. This package conducts logistic regression analysis with misspecification in outcome variables.

License GPL

LazyLoad yes

RoxygenNote 5.0.1

Repository CRAN

Repository/R-Forge/Project semdiag

Repository/R-Forge/Revision 120

Repository/R-Forge/DateTimeStamp 2016-05-16 20:22:19

Date/Publication 2017-05-31 06:20:10 UTC

NeedsCompilation no

R topics documented:

`logistic4p-package` *Logistic Regression with Misclassification in Dependent Variables*

Description

Error in a binary dependent variable, also known as misclassification, has not drawn much attention in psychology. Ignoring misclassification in logistic regression can result in misleading parameter estimates and statistical inference. This package conducts logistic regression analysis with misspecification in outcome variables.

Details

The DESCRIPTION file:

Package:	logistic4p
Type:	Package
Title:	Logistic Regression with Misclassification in Dependent Variables
Version:	1.5
Date:	2017-05-30
Depends:	R (>= 2.10), MASS
Author:	Haiyan Liu and Zhiyong Zhang
Maintainer:	Zhiyong Zhang <johnnyzhz@gmail.com>
Description:	Error in a binary dependent variable, also known as misclassification, has not drawn n
License:	GPL
LazyLoad:	yes
RoxygenNote:	5.0.1
Repository:	R-Forge
Repository/R-Forge/Project:	semdiag
Repository/R-Forge/Revision:	120
Repository/R-Forge/DateTimeStamp:	2016-05-16 20:22:19
Date/Publication:	2016-05-16 20:22:19
NeedsCompilation:	no
Packaged:	2016-05-16 20:25:20 UTC; rforge

Index of help topics:

```
logistic                Logistic Regression
logistic4p              Logistic Regressions with Misclassification
                        Correction
logistic4p-package      Logistic Regression with Misclassification in
                        Dependent Variables
```

logistic4p.e	Logistic regressions with constrained FP and FN misclassifications
logistic4p.fn	Logistic Regression Model with FN Misclassification Correction
logistic4p.fp	Logistic Regression with FP Misclassification Correction
logistic4p.fp.fn	Logistic Regression with both FP and FN Misclassification Correction
nlsy	An example data set
print.logistic4p	Printing Outputs of Logistic Regression with Misclassification Parameters

Author(s)

Haiyan Liu and Zhiyong Zhang

Maintainer: Zhiyong Zhang <johnnyzhz@gmail.com>

References

Liu, H. and Zhang, Z. (2016) Logistic Regression with Misclassification in Dependent Variables: Method and Software.(In preparation.)

Examples

```
## Not run:
data(nlsy)
x=nlsy[, -1]
y=nlsy[,1]
mod=logistic4p(x, y, model='fn')

## End(Not run)
```

---

logistic *Logistic Regression*

Description

Fit a logistic regression model.

Usage

```
logistic(x, y, initial, max.iter = 1000, epsilon = 1e-06, detail = FALSE)
```

Arguments

x, y	x is a data frame or data matrix containing the predictor variables and y is the vector of outcomes. The number of rows in x must be the same as the length of y.
initial	a vector of starting values for the parameters in the linear predictor; if not specified, the default initials are 0 for all parameters.
max.iter	a positive integer giving the maximal number of iterations; if it is reached, the algorithm will stop.
epsilon	a positive convergence tolerance epsilon; the iterations converge when max(|par-par_old|)<epsilon.
detail	logical indicating if output should be printed for each iteration.

Value

estimates	a named matrix of estimates including parameter estimates, standard errors, z-scores, and p-values.
n.iter	an integer giving the number of iteration used
d	the actual max absolute difference of the parameters of the last two iterations, d=max(|par.final-par_old|).
loglike	loglikelihood evaluated at the parameter estimates.
AIC	Akaike Information Criterion.
BIC	Bayesian Information Criterion.
converged	logical indicating whether the current procedure converged or not.

Author(s)

Haiyan Liu and Zhiyong Zhang

Examples

```
## Not run:
data(nlsy)
y=nlsy[,1]
x=nlsy[, -1]

mod=logistic(x,y)

## End(Not run)
```

logistic4p *Logistic Regressions with Misclassification Correction*

Description

logistic4p is used to fit logistic regressions with correction of the misclassifications in the binary dependent variable. It is specified by

Usage

```
logistic4p(x, y, initial, model = c("lg", "fp.fn", "fp", "fn", "equal"),
max.iter = 1000, epsilon = 1e-06, detail = FALSE)
```

Arguments

x, y	x is a data frame or data matrix containing the predictor variables and y is the vector of outcomes. The number of rows in x must be the same as the length of y.
initial	starting values for the parameters in the model(FP,FN misclassification parameters and those in the linear predictor); if not specified, the default initials are 0 for the misclassification parameters and estimates obtained from the logistic regression for the parameters in the linear predictor.
model	a character string specifying the model to be used in the analysis. Currently available options are "lg" (logistic regression), "fp.fn" (logistic regression with both FP and FN parameters), "fp" (logistic regression with the FP parameter), "fn" (logistic regression with the FN parameter), "equal" (logistic regression with FN=FN). If it is not specified, the default one ('lg') will be used.
max.iter	a positive integer giving the maximal number of iterations; if it is reached, the algorithm will stop.
epsilon	a positive convergence tolerance epsilon; the iterations converge when max(|par-par_old|)<epsilon.
detail	logical indicating if the itermediate output should be printed after each iteration.

Details

This package implements the logistic regressions with misclassification corrections. There are five different models which can be specified by 'model'.

In the specification, x is a matrix of data frame of predictors fitted to the model; y is a numeric vector taking either 0 or 1.

The 'initial' is the vector of starting values for both misclassification and regression coefficients parameters in the model. It is suggested to provide 'initial', however if not, the default one will be used.

For the background to warning messages about 'fitted probabilities numerically 0 or 1 occurred', when the fitted probabilities of some individuals are either 0 or 1.

The package cannot handle missing data problems currently. If there are missing values in either x or y, there will be warning message.

Value

logistic4p returns a list of values inheriting from "logistic4p".

`estimates`	a named matrix of estimates including parameter estimates, standard errors, z-scores, and p-values.
`n.iter`	an integer giving the number of iteration used
`d`	the actual max absolute difference of the parameters of the last two iterations, d=max(|par.final-par_old|).
`loglike`	loglikelihood evaluated at the parameter estimates.
`AIC`	Akaike Information Criterion.
`BIC`	Bayesian Information Criterion.
`converged`	logical indicating whether the current procedure converged or not.

Author(s)

Haiyan Liu and Zhiyong Zhang

References

Liu, H. and Zhang, Z. (2016) Logistic Regression with Misclassification in Dependent Variables: Method and Software.(In preparation.)

Examples

```
## Not run:
data(nlsy)
y=nlsy[, 1]
x=nlsy[,-1]

mod1=logistic4p(x,y)
mod1
mod1$estimates

mod2=logistic4p(x,y, model='fp.fn')

mod3=logistic4p(x,y, model='fn')

## End(Not run)
```

`logistic4p.e` *Logistic regressions with constrained FP and FN misclassifications*

Description

Fit logistic regressions with misclassification correction. The FP and FN parameters are constrained to be equal.

Usage

```
logistic4p.e(x, y, initial, max.iter = 1000, epsilon = 1e-06, detail = FALSE)
```

Arguments

`x, y`	x is a data frame or data matrix containing the predictor variables and y is the vector of outcomes. The number of rows in x must be the same as the length of y.
`initial`	starting values for the parameters in the model(the misclassification parameter and those in the linear predictor); if not specified, the default initials are 0 for the misclassification parameters and estimates obtained from the logistic regression for the parameters in the linear predictor.
`max.iter`	a positive integer giving the maximal number of iterations; if it is reached, the algorithm will stop.
`epsilon`	a positive convergence tolerance epsilon; the iterations converge when max(|par-par_old|)<epsilon.
`detail`	logical indicating if the itermediate output should be printed after each iteration.

Value

`estimates`	a named matrix of estimates including parameter estimates, standard errors, z-scores, and p-values.
`n.iter`	an integer giving the number of iteration used
`d`	the actual max absolute difference of the parameters of the last two iterations, d=max(|par.final-par_old|).
`loglike`	loglikelihood evaluated at the parameter estimates.
`AIC`	Akaike Information Criterion.
`BIC`	Bayesian Information Criterion.
`converged`	logical indicating whether the current procedure converged or not.

Author(s)

Haiyan Liu and Zhiyong Zhang

Examples

```
## Not run:
data(nlsy)
y=nlsy[,1]
x=nlsy[, -1]

mod=logistic4p.e(x, y, max.iter = 1000, epsilon = 1e-06, detail = FALSE)

## End(Not run)
```

`logistic4p.fn` *Logistic Regression Model with FN Misclassification Correction*

Description

logistic4p.fn is used to fit logistic regressions with the false negative parameter in the model.

Usage

```
logistic4p.fn(x, y, initial, max.iter = 1000, epsilon = 1e-06, detail = FALSE)
```

Arguments

`x, y`	x is a data frame or data matrix containing the predictor variables and y is the vector of outcomes. The number of rows in x must be the same as the length of y.
`initial`	starting values for the parameters in the model(FP and those in the linear predictor); if not specified, the default initials are 0 for the misclassification parameter and estimates obtained from the logistic regression for the parameters in the linear predictor.
`max.iter`	a positive integer giving the maximal number of iterations; if it is reached, the algorithm will stop.
`epsilon`	a positive convergence tolerance epsilon; the iterations converge when max(|par-par_old|)<epsilon.
`detail`	logical indicating if output should be printed for each iteration.

Value

`estimates`	a named matrix of estimates including parameter estimates, standard errors, z-scores, and p-values.
`n.iter`	an integer giving the number of iteration used
`d`	the actual max absolute difference of the parameters of the last two iterations, d=max(|par.final-par_old|).
`loglike`	loglikelihood evaluated at the parameter estimates.
`AIC`	Akaike Information Criterion.
`BIC`	Bayesian Information Criterion.
`converged`	logical indicating whether the current procedure converged or not.

Author(s)

Haiyan Liu and Zhiyong Zhang

Examples

```
## Not run:
data(nlsy)
y=nlsy[,1]
x=nlsy[,-1]

mod=logistic4p.fn(x, y, max.iter = 1000, epsilon = 1e-06, detail = FALSE)

## End(Not run)
```

---

logistic4p.fp *Logistic Regression with FP Misclassification Correction*

Description

logistic4p.fp is used to fit logistic regression models with correction of the false positive misclassification in the binary dependent variable.

Usage

```
logistic4p.fp(x, y, initial, max.iter = 1000, epsilon = 1e-06, detail = FALSE)
```

Arguments

x, y	x is a data frame or data matrix containing the predictor variables and y is the vector of outcomes. The number of rows in x must be the same as the length of y.
initial	starting values for the parameters in the model(FP,FN misclassification parameters and those in the linear predictor); if not specified, the default initials are 0 for the misclassification parameters and estimates obtained from the logistic regression for the parameters in the linear predictor.
max.iter	a positive integer giving the maximal number of iterations; if it is reached, the algorithm will stop.
epsilon	a positive convergence tolerance epsilon; the iterations converge when max(|par-par_old|)<epsilon.
detail	logical indicating if output should be printed for each iteration.

Value

estimates	a named matrix of estimates including parameter estimates, standard errors, z-scores, and p-values.
n.iter	an integer giving the number of iteration used
d	the actual max absolute difference of the parameters of the last two iterations, d=max(|par.final-par_old|).
loglike	loglikelihood evaluated at the parameter estimates.
AIC	Akaike Information Criterion.
BIC	Bayesian Information Criterion.
converged	logical indicating whether the current procedure converged or not.

Author(s)

Haiyan Liu and Zhiyong Zhang

Examples

```
## Not run:
data(nlsy)
y=nlsy[,1]
x=nlsy[, -1]

mod.fp=logistic4p.fp(x, y, max.iter = 1000, epsilon = 1e-06, detail = FALSE)

## End(Not run)
```

---

logistic4p.fp.fn *Logistic Regression with both FP and FN Misclassification Correction*

Description

logistic4p.fp.fn is used to fit a logistic regression model with both FP and FN misclassification parameters to a binary dependent variable.

Usage

```
logistic4p.fp.fn(x, y, initial, max.iter = 1000, epsilon = 1e-06, detail = FALSE)
```

Arguments

x, y	x is a data frame or data matrix containing the predictor variables and y is the vector of outcomes. The number of rows in x must be the same as the length of y.

initial	starting values for the parameters in the model(FP,FN misclassification parameters and those in the linear predictor); if not specified, the default initials are 0 for the misclassification parameters and estimates obtained from the logistic regression for the parameters in the linear predictor.
max.iter	a positive integer giving the maximal number of iterations; if it is reached, the algorithm will stop.
epsilon	a positive convergence tolerance epsilon; the iterations converge when max(|par-par_old|)<epsilon.
detail	logical indicating if the output should be printed for each iteration.

Value

estimates	a named matrix of estimates including parameter estimates, standard errors, z-scores, and p-values.
n.iter	an integer giving the number of iteration used
d	the actual max absolute difference of the parameters of the last two iterations, d=max(|par.final-par_old|).
loglike	loglikelihood evaluated at the parameter estimates.
AIC	Akaike Information Criterion.
BIC	Bayesian Information Criterion.
converged	logical indicating whether the current procedure converged or not.

Author(s)

Haiyan Liu and Zhiyong Zhang

Examples

```
## Not run:
data(nlsy)
y=nlsy[,1]
x=nlsy[, -1]

mod=logistic4p.fp.fn(x,y)

## End(Not run)
```

---

nlsy *An example data set*

Description

Data set used in Liu & Zhang (2016).

marijuana: binary; 1=used, 0=not used

gender: binary; 1=female, 0=male

smoke: binary; 1=smoke, 0=not smoke

residence: binary; 1=urban areas, 0=rural areas

peer: comprised score on peers life style; the higher score, the healthier the peers live.

Usage

```
data(nlsy)
```

---

print.logistic4p *Printing Outputs of Logistic Regression with Misclassification Parameters*

Description

This is an function to print the inherit outputs of. logistic4p

Usage

```
## S3 method for class 'logistic4p'
print(x, ...)
```

Arguments

x	An object of class 'logistic4p'.
...	further arguments passed to or from other methods.

Author(s)

Haiyan Liu and Zhiyong Zhang

Examples

```
## Not run:
data(nlsy)
y=nlsy[,1]
x=nlsy[,-1]

mod=logistic4p(x,y)
print(mod)

## End(Not run)
```

Index