Package 'logistiX'

November 30, 2012

Type Package

Title Exact logistic regression including Firth correction

Version 1.0

Date 2012-09-27

Author Georg Heinze and Tobias Ladner

Maintainer Georg Heinze <georg.heinze@meduniwien.ac.at>

Depends R (>= 2.15.1)

Description Exact logistic regression including Firth correction for binary covariates

License GPL

URL http://www.meduniwien.ac.at/cemsiis/kb/software/stat-software/logistiX

LazyLoad

Archs i386, x64

R topics documented:

`logistiX-package` *Exact logistic regression*

Description

Implements exact conditional logistic regression in R.

Details

Package:	logistiX
Type:	Package
Version:	1.0
Date:	2012-09-28
License:	GPL

`logistiX` implements exact conditional logistic regression in R, including maximum conditional likelihood, median unbiased and Firth-corrected estimation methods, and twice-smaller-tail, probability and scores methods for hypothesis tests, with optional P-mid adjustment. For confidence interval estimation, various methods are available, such as exact, P-mid and integrated randomized variants, as well as the profile penalized completely conditional likelihood method discussed in Heinze and Puhr, 2010. The construction of the exact conditional permutational distributions is based on the Multivariate Shift Algorithm by Hirji et al, 1989. In contrast to other, Monte-Carlo- or Markov-chain-Monte-Carlo approaches, this is an exact method, and does not depend on Monte Carlo sample size or on convergence of an MCMC chain.

Implemented Methods:

Algorithm to construct permutational distribution:	Multivariate Shift Algorithm
Estimation methods:	Maximum conditional likelihood estimation (MCLE)
	Median unbiased estimation (MUE)
	"LogXact" (LX, combination of MCLE and MUE)
	Firth-corrected estimation (FC)
Confidence interval methods:	Exact twice-smaller-tail (`TST`)
	Exact scores (inverting the scores test) (`scores`)
	Pmid adjustment (`pmid`)
	Integrated randomized confidence limits (`randomized`)
	Profile penalized completely conditional likelihood (`PCCL`)
Inference methods:	Exact twice-smaller-tail (`TST`)
	Exact scores (`scores`)
	Exact probability (`probability`)
	P-mid adjustment (`pmid`)

Author(s)

Georg Heinze and Tobias Ladner

Maintainer: Georg Heinze <georg.heinze@meduniwien.ac.at>

References

Hirji KF, Mehta CR and Patel NR (1987). Computing Distributions for Exact Logistic Regression. Journal of the American Statistical Association 82:1110-1117

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

See Also

Other packages for small-sample logistic regression are:

`elrm` (Approximate exact inference in logistic regression using MCMC)

`logistf` (Bias-reduced logistic regression using Firth's correction)

Examples

```
set.seed(123)
vars<-5
n<-50
beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
px<-0.2

x<-matrix(rbinom(vars*n,1,px),n,vars)

py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
y<-rbinom(n,1,py)

test1<-logistiX(x=x, y=y)
summary(test1)
coef(test1)
confint(test1)
```

`coef.logistiX` *Method to extract regression coefficients from a* `logistiX` *object*

Description

This method returns the estimated regression coefficients from a `logistiX` object. Available estimation methods are maximum conditional likelihood, median unbiased, their combination (median unbiased replaces infinite maximum likelihood estimates), and completely conditional Firth-type likelihood maximization (CCFL, Heinze and Puhr, 2010)

Usage

```
## S3 method for class 'logistiX'
coef(object, type = "LX", ...)
```

Arguments

`object`	a `logistiX` object
`type`	one of `c("LX", "CCFL", "MLE", "MUE")`. `"MLE"` returns the maximum likelihood estimates, `"MUE"` the median unbiased estimates, `"LX"` returns the MLE, and replaces infinite values by the MUE, and `"CCFL"` returns the Firth-corrected (CCFL) estimate.
`...`	additional argument(s) for methods.

Details

The parameter for the intercept term will not be computed and can not be extracted.

Value

A px1 vector of regression coefficients. Currently, the intercept estimate is not computed.

Note

Comparison to LogXact: basically, the same functionality is provided. In `logistiX`, the median unbiased estimate can also be displayed if the maximum conditional likelihood estimate is finite.

Author(s)

Georg Heinze and Tobias Ladner

References

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

See Also

`confint.logistiX`

Examples

```
set.seed(123)
vars<-5
n<-50
beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
px<-0.2

x<-matrix(rbinom(vars*n,1,px),n,vars)

py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
y<-rbinom(n,1,py)

test1<-logistiX(x=x, y=y)

#LogXact estimate:
coef(test1, "LX")
#CCFL estimate:
coef(test1, "CCFL")
```

---

confint.logistiX *Method to extract confidence intervals from a* logistiX *object.*

Description

This method extracts confidence intervals from a `logistiX` object.

Usage

```
## S3 method for class 'logistiX'
confint(object, parm, level = 0.95,  type = "exact", method = "TST", ran.steps = seq(5/100, 0.95
```

Arguments

object	a logistiX object
parm	a specification of which parameters are to be given confidence intervals, a vector of numbers. If missing, all parameters are considered. vstart and vstop may be used alternatively.)
level	the confidence level (default=0.95)
type	one of c("exact", "pmid","randomized","PCCL") (see above)
method	one of c("TST", "scores") (see above)
ran.steps	the pmid-factor steps used for integration (randomized CL type). Default is 0.05 to 0.95 by 0.01 (ran.steps = seq(5/100, 0.95, 0.01)).
interval	the range of values for optimize to search the confidence limits in. If estimation fails (i.e., if the actual level is not equal the specified level), then the interval limits are halfened.
vstart	Optional: the first variable to compute confidence intervals (default=NULL). If NULL, the first variable contained in the obj object will be used.
vstop	Optional: the last variable to compute confidence intervals (default=NULL). If NULL, the last variable contained in the obj object will be used.
...	additional argument(s) for methods

Details

This confint method can be used to generate confidence intervals by various methods and 'types', that can be combined. Available methods are:

Method	type argument	method argument
Exact twice-smaller-tail	"exact"	"TST"
Pmid twice-smaller-tail	"pmid"	"TST"
Exact scores	"exact"	"scores"
Pmid scores	"pmid"	"scores"
Integrated randomized twice-smaller-tail (EXPERIMENTAL)	"randomized"	"TST"
Profile penalized completely conditional likelihood	"PCCL"	

The confidence level can be chosen, and for the "randomized" method, the number and density of integration steps can be specified. Furthermore, the set of parameters for which confidence limits are requested can be restricted.

PCCL intervals are as proposed by Heinze and Puhr (2010). Integrated randomized TST intervals are an experimental feature.

The confint method is also called by summary.

Value

A p x 2 matrix of confidence limits, where p = vstop-vstart+1.

Note

Comparison to LogXact: in LogXact, only exact TST and PCCL intervals are implemented.

Author(s)

Georg Heinze and Tobias Ladner

References

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

See Also

coef.logistiX

Examples

```
set.seed(123)
vars<-5
n<-50
beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
px<-0.2

x<-matrix(rbinom(vars*n,1,px),n,vars)

py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
y<-rbinom(n,1,py)

test1<-logistiX(x=x, y=y)
summary(test1)
confint(test1, type="exact")
confint(test1, type="pmid")
```

---

logistiX *Exact logistic regression*

Description

Main function for exact logistic regression in R.

Usage

```
logistiX(x = NULL, y = NULL, strat = NULL, tst = 1, sc = 1, pmid = 2, option = 2, noint = 0, vst
```

Arguments

x	A nxp matrix of covariates, these must be 0/1 coded. (Currently, it is not possible to use discrete/ordinal variables or continuous variables.)
y	A px1 vector of binary outcomes
strat	A px1 vector of stratification categories (e.g., matched set indicator in a matched case-control study). Not needed.
tst	Compute twice-smaller-tail method of inference (default=TRUE=yes)
sc	Compute scores method of inference (default=TRUE=yes)

pmid	Request pmid adjustment (default=2 meaning that pmid and exact results are computed. 1: only pmid, 0: only exact)
option	option=2: compute permutational distributions and estimate; option=1 only compute permutational distributions
noint	Omit intercept (NOT WORKING!) (default=0)
vstart	First variable to estimate, default=1
vstop	Last variable to estimate, default = p
alpha	1-confidence level, default=0.05
details	Request detailed printing of messages (default=0=no).

Details

logistiX computes the permutational null distribution of the sufficient statistics of either all or a subset of the variables of a model, using the multivariate shift algorithm (Hirji et al 1987). By default, logistiX also computes the parameter estimates by four methods: MLE (maximum conditional likelihood), MUE (median unbiased estimation), LX (the "LogXact" method, which uses MLE if defined, MUE otherwise), CCFL (Firth-corrected estimate, Heinze and Puhr, 2010). Confidence intervals are computed by the following methods: TST: twice-smaller-tail method, scores: scores method, pmid: adds pmid-adjustment (the observed sufficient statistic is downweighted by half (the "pmid-factor"), a correction to remove conservatism by exact inference; these confidence intervals are only approximately exact), randomized: computes integrated randomized confidence intervals, i.e., confidence limits are averaged over a series of values obtained by varying the pmid-factor from 0 to 1.

By calling the summary, coef or confint methods, one can obtain results by the specified methods.

Value

distout	Conditional permutational distribution for variables vstart:vstop)
estout	Parameter estimates by the methods MLE, MUE, LX and CCFL
ciout	Confidence intervals by the methods TST, TST-pmid, SC and SC-pmid
tobs	Observed values of the sufficient statistics for all variables (including intercept)
call	The call to logistiX

Warning

Memory requirements can be excessive already with a moderately-sized data set (e.g., with N=200 and 5 variables). R's 64-bit version allows for more allocatable memory, but can be slow if virtual memory is on a conventional harddisk drive. In such cases, the elrm package may be an alternative.

Note

Currently, the implementation has two major limitations: (1) only binary covariates can be processed, and (2) it is not possible to compute the joint distribution of several parameters and consequently no simultaneous testing is possible

Author(s)

Georg Heinze and Tobias Ladner (georg.heinze@meduniwien.ac.at)

References

Hirji KF, Mehta CR and Patel NR (1987). Computing Distributions for Exact Logistic Regression. Journal of the American Statistical Association 82:1110-1117

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

See Also

print.logistiX, summary.logistiX, plot.logistiX, profile.logistiX, coef.logistiX, confint.logistiX

Examples

```
set.seed(123)
 vars<-5
 n<-50
 beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
 px<-0.2

 x<-matrix(rbinom(vars*n,1,px),n,vars)

 py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
 y<-rbinom(n,1,py)

 test1<-logistiX(x=x, y=y)
 summary(test1)

 #LogXact estimate:
 coef(test1, "LX")
 confint(test1, type="exact")


 # CCFL estimate:
 coef(test1, "CCFL")
 # PCCL confidence limits:
 confint(test1, type="PCCL")
```

---

plot.logistiX *Plot method for logistiX objects*

Description

Plots the exact null or alternative distribution of the sufficient statistic for a variable of interest. Computes the likelihood or penalized likelihood for the observed value of that variable.

Usage

```
## S3 method for class 'logistiX'
plot(x, var, beta = 0, plot = TRUE, firth = FALSE, ...)
```

Arguments

x	a logistiX object
var	The variable number, for which the distribution of sufficient statistic should be plotted.
beta	The value of the regression coefficient for which the exact distribution should be computed. (default=0, will plot the null distribution)
plot	Set to TRUE for plotting (default=TRUE). Set to FALSE to suppress plotting.
firth	Set to TRUE for computing the penalized instead of the ordinary likelihood.
...	further arguments

Details

This method can be used to visualize the exact distribution of the sufficient statistic of a variable of interest, and/or it can be used to compute the likleihood or penalized likelihood for the sufficient statistic of that variable.

The method can also display the distribution under a non-0 beta coefficient.

Value

A list with entries:

$dist	a data frame with columns t.stat: the value of the sufficient statistic score: the probability for the sufficient statistic) penalized.score: the penalized likelihood of each value of the sufficient statistic
$likelihood	the value of the (penalized) likelihood at the observed value of the sufficient statistic

Author(s)

Georg Heinze and Tobias Ladner

References

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

Examples

```
set.seed(123)
vars<-5
n<-50
beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
px<-0.2

x<-matrix(rbinom(vars*n,1,px),n,vars)

py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
y<-rbinom(n,1,py)

test1<-logistiX(x=x, y=y)
```

```
plot(test1,1) # plots the distribution for the first variable
plot(test1,1, plot=FALSE, firth=TRUE)$likelihood # returns the penalized null likelihood for the first va
```

`print.logistiX` *Print method for* `logistiX`

Description

This method prints the main results from a `logistiX` analysis.

Usage

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

Arguments

`x`	a `logistiX` object.
`...`	further arguments

Value

returns a 4p x 5 matrix of estimates and confidence limits. For each variable, four methods are applied:

for estimation: MLE, MUE, LX, and CCFL; for confidence limits: TST, TST-pmid, scores, scores-pmid

Note

`summary.logistiX` provides a nicer summary with user-specifiable contents.

Author(s)

Georg Heinze and Tobias Ladner

References

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

See Also

`summary.logistiX`

Examples

```
set.seed(123)
 vars<-5
 n<-50
 beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
 px<-0.2

 x<-matrix(rbinom(vars*n,1,px),n,vars)

 py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
 y<-rbinom(n,1,py)

 test1<-logistiX(x=x, y=y)

 print(test1)
```

`profile.logistiX` *Profile likelihood analysis for* `logistiX`

Description

Computes the profile (penalized) likelihood at a sequence of beta values

Usage

```
## S3 method for class 'logistiX'
profile(fitted, parm = 1, firth = FALSE, type = "loglike", steps = 101, betaseq = NULL, normaliz
```

Arguments

`fitted`	A fitted `logistiX` object
`parm`	Variable number for which the profile likelihood should be computed
`firth`	if `TRUE`, requests penalized profile likelihood analysis
`type`	one of `c("loglike","like"`. By default, returns profile log likelihood; otherwise, profile likelihood
`steps`	number of steps for profiling (default=101).
`betaseq`	Optional: specifiy the steps at which the profile likelihood should be computed. If left `NULL`, the program will define a range on its own slightly extending the estimated confidence limits.
`normalize`	If `TRUE`, will normalize the profile to be 0 (for log likelihood) or 1 (for likelihood) at the maximum
`...`	additional argment(s) for methods.

Details

The profile likelihod can be checked for symmetry to see if normal approximation would be appropriate, or to evaluate whether the likelihood is flat around the estimated value. The beta values used for computation of the profile likelihood are generated automatically, but could also be user-supplied.

Value

A steps x 2 matrix of beta and profile likelihood values

Note

Not implemented in LogXact.

Author(s)

Georg Heinze and Tobias Ladner

References

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

See Also

confint.logistiX

Examples

```
set.seed(123)
vars<-5
n<-50
beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
px<-0.2

x<-matrix(rbinom(vars*n,1,px),n,vars)

py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
y<-rbinom(n,1,py)

test1<-logistiX(x=x, y=y)

# overlay unpenalized and penalized log likelihood of variable 1:
plot(profile(test1, 1, normalize=TRUE), type="l")
lines(profile(test1, 1, firth=TRUE, normalize=TRUE), lty=2)
legend("topright",lty=c(1,2),legend=c("Unpenalized", "Penalized"))
```

---

summary.logistiX *Summary method for* logistiX

Description

This methods provides a table of parameter estimates, confidence intervals, and p-values from a logistiX analysis.

Usage

```
## S3 method for class 'logistiX'
summary(object, esttype = "LX", citype = "exact", cimethod = NULL, pmid = FALSE, testtype = "TST
```

Arguments

object	A logistiX object.
esttype	Type of estimate: see documentation of coef.logistiX
citype	Type of confidence interval: see documentation of confint.logistiX
cimethod	Method for confidence interval estimation: see documentation of confint.logistiX
pmid	pmid adjustment for testing (if TRUE)
testtype	Type of test (one of "TST", "scores", "probability", "PCLR"; default="TST"): "TST" requests twice-smaller-tail test "scores" requests scores test "probability" requests sum-of-smaller-probabilities test "PCLR" requests penalized completely conditional likelihood ratio test (not exact)
cilevel	the confidence level (default=0.95)
...	further arguments

Details

The summary method allows to provide a comprehensive table of estimates, confidence intervals and p-values, where the methods used for these statistics can be chosen by the user.

Value

Returns some summary of the function call, the requested methods and a p x 6 matrix with:

column 1	estimates (depending on esttype)
columns 2-3	confidence limits (depending on citype and cimethod)
column 4	test statistics (depending on testtype)
column 5	p-values (depending on testtype)
column 6	cardinality of the exact distribution of sufficient statistic for the respective parameter

Author(s)

Georg Heinze and Tobias Ladner

References

Heinze G and Puhr R (2010). Bias-reduced and separation-proof conditional logistic regression with small or sparse data sets. Statistics in Medicine 29:770-777

Examples

```
set.seed(123)
  vars<-5
  n<-50
  beta<-c(-2,3,2,-2,1,2,2,1,0,-1,2)
  px<-0.2

  x<-matrix(rbinom(vars*n,1,px),n,vars)
```

```
py<-1/(1+exp(-cbind(rep(1,n),x) %*% beta[1:(vars+1)]))
y<-rbinom(n,1,py)

test1<-logistiX(x=x, y=y)

summary(test1)

summary(test1, esttype="CCFL", citype="PCCL")
```

Index