

Package ‘addhaz’

May 16, 2016

Type Package

Title Binomial and Multinomial Additive Hazards Models

Version 0.4

Date 2016-05-16

Author Renata T C Yokota, Caspar W N Looman, Wilma J Nusselder,
Herman Van Oyen, Geert Molenberghs

Maintainer Renata T C Yokota <Renata.Yokota@wiv-isp.be>

Description Functions to fit the binomial and multinomial additive hazards models and to calculate the contribution of diseases/conditions to the disability prevalence, as proposed by Nusselder and Looman (2004) <DOI:10.1353/dem.2004.0017>.

License GPL-3

LazyData TRUE

Depends R (>= 3.1.0), stats, boot (>= 1.3-17), Matrix (>= 1.2-3), MASS

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-05-16 19:38:50

R topics documented:

BinAddHaz	2
disabData	5
MultAddHaz	6

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

Description

This function fits binomial additive hazards models subject to linear inequality constraints using the function `constrOptim` in the `stats` package for binary outcomes. Additionally, it calculates the cause-specific contributions to the disability prevalence based on the attribution method, as proposed by Nusselder and Looman (2004).

Usage

```
BinAddHaz(formula, data, subset, weights, na.action, model = TRUE,
          contrasts = NULL, start = FALSE, start.val, attrib = FALSE,
          attrib.var, type.attrib = "abs", set.seed = FALSE, seed,
          bootstrap = FALSE, nbootstrap, parallel = FALSE,
          type.parallel = "snow", ncpus = 4,...)
```

Arguments

<code>formula</code>	A formula expression of the form <code>response ~ predictors</code> , similar to other regression models. In case of <code>attrib = TRUE</code> , the first predictor in the formula should be the <code>attrib.var</code> . See example.
<code>data</code>	an optional data frame or matrix containing the variables in the model. If not found in <code>data</code> , the variables are taken from <code>environment(formula)</code> , typically the environment from which <code>BinAddHaz</code> is called.
<code>subset</code>	an optional vector specifying a subset of observations to be used in the fitting process. All observations are included by default.
<code>weights</code>	an optional vector of survey weights to be used in the fitting process.
<code>na.action</code>	a function which indicates what should happen when the data contain NAs. The default is set by the <code>na.action</code> setting of options, and is <code>na.fail</code> if that is unset. The 'factory-fresh' default is <code>na.omit</code> .
<code>model</code>	logical. If <code>TRUE</code> , the model frame is included as a component of the returned object.
<code>contrasts</code>	an optional list to be used for some or all of the factors appearing as variables in the model formula.
<code>start</code>	logical. Will the user provide starting values? Default is <code>FALSE</code> , in which case the starting values will be the absolute values of the coefficients for a Poisson <code>glm</code> .
<code>start.val</code>	vector of the starting values. They should be provided if <code>start = TRUE</code> and lie inside the parameter space.
<code>attrib</code>	logical. Should the attribution of disability to chronic diseases/conditions be calculated? Default is <code>FALSE</code> .

<code>attrib.var</code>	character indicating the name of the attribution variable to be used if <code>attrib = TRUE</code> . If missing, the attribution variable is the intercept (no stratification of the attribution results). The attribution variable must be the first variable included in the linear predictor in formula. See example.
<code>type.attrib</code>	type of attribution to be calculated. The options are "abs" for absolute contribution, "rel" for relative contribution, or "both" for both absolute and relative contributions. Default is "abs".
<code>set.seed</code>	logical. Should a random seed be set? Default is FALSE.
<code>seed</code>	integer indicating the random seed. Only valid if <code>set.seed = TRUE</code> .
<code>bootstrap</code>	logical. Should bootstrap percentile confidence intervals be calculated for the parameters and attributions? Default is FALSE. See details.
<code>nbootstrap</code>	integer. Number of bootstrap replicates.
<code>parallel</code>	logical. Should parallel calculations be used to obtain the bootstrap percentile confidence intervals? Only valid if <code>bootstrap = TRUE</code> . Default is FALSE.
<code>type.parallel</code>	type of parallel operation to be used (if <code>parallel = TRUE</code>), with options: "multicore" and "snow". Default is "snow". See details.
<code>ncpus</code>	integer. Number of processes to be used in parallel operation: typically one would choose this to be the number of available CPUs. Default is 4.
<code>...</code>	other arguments passed to or from the other functions.

Details

The model is a generalized linear model with a non-canonical link function, which requires a restriction on the linear predictor ($\eta \geq 0$) to produce valid probabilities. This restriction is implemented in the optimization procedure, with an adaptive barrier algorithm, using the function `constrOptim` in the `stats` package.

The variance-covariance matrix is based on the observed information matrix.

This version of the the package only allows the calculation of non-parametric bootstrap percentile confidence intervals (CI), which are based on the 2.5 and 97.5 percentiles. Also, the function gives the user the option to do parallel calculation of the bootstrap CI. The `snow` parallel option is available for all operating systems (Windows, Linux, and Mac OS) while the `multicore` option is only available for Linux and Mac OS systems. These two calculations are done by calling the `boot` function in the `boot` package. For more details, see the documentation of the `boot` package.

More information about the binomial additive hazards model and the calculation of the contribution of chronic conditions to the disability prevalence can be found in the references.

Value

A list with arguments:

<code>coefficients</code>	numeric vector with the regression coefficients.
<code>ci</code>	confidence intervals calculated via bootstrapping (if <code>bootstrap = TRUE</code>) or as the inverse of the observed information matrix.
<code>resDeviance</code>	residual deviance.
<code>df</code>	degrees of freedom.

pvalue	numerical vector of p-values based on the Wald test. Only provided if bootstrap = FALSE.
stdError	Numerical vector with the standard errors for the parameter estimates based on the inverse of the observed information matrix. Only provided if bootstrap = FALSE.
vcov	variance-covariance (inverse of the observed information matrix). Only provided if bootstrap = FALSE.
contribution	for type.attrib = "abs" or "rel", a matrix is provided. For type.attrib = "both", a list with two matrices ("abs" and "rel") is provided. This represents the contribution of each predictor in the model (usually diseases or conditions) to the disability prevalence. Percentile bootstrap confidence intervals are provided if bootstrap = TRUE.
bootsRep	matrix with the bootstrap replicates.
bootstrap	logical. Was bootstrap CI requested?
fitted.values	the fitted mean values, obtained by transforming the linear predictor by the inverse of the link function.
residuals	difference between the response variable and the fitted values.
call	the matched call.

Author(s)

Renata T C Yokota. This function is based on the R code developed by Caspar W N Looman and Wilma J Nusselder for non R-users, with modifications. Original R code is available upon request to Wilma J Nusselder (w.nusselder@erasmusmc.nl).

References

Nusselder, W.J., Looman, C.W.N. (2004). Decomposition of differences in health expectancy by cause. *Demography*, 41(2), 315-334.

Nusselder, W.J., Looman, C.W.N. (2010). WP7: Decomposition tools: technical report on attribution tool. European Health Expectancy Monitoring Unit (EHEMU). Available at http://www.eurohex.eu/pdf/Reports_2010/2010TR7.2_TR%20on%20attribution%20tool.pdf.

See Also

[MultAddHaz](#)

Examples

```
data(disabData)

## Model without bootstrap percentile confidence intervals and no attribution

fit1 <- BinAddHaz(dis.bin ~ diab + arth + stro , data = disabData, weights = wgt)
summary(fit1)

## Model with bootstrap percentile confidence intervals and attribution without
## stratification, no parallel calculation
# Warning message due to the low number of bootstrap replicates
```

```
## Not run:
fit2 <- BinAddHaz(dis.bin ~ diab + arth + stro , data = disabData, weights = wgt,
                 attrib = TRUE, type.attrib = "both", set.seed = TRUE, seed = 111,
                 bootstrap = TRUE, nbootstrap = 5)
summary(fit2)

## Model with bootstrap percentile confidence intervals with attribution based on
## the age variable and parallel calculation of bootstrap CI
# Warning message due to the low number of bootstrap replicates

diseases <- as.matrix(disabData[,c("diab", "arth", "stro")])

fit3 <- BinAddHaz(dis.bin ~ factor(age) -1 + diseases:factor(age), data = disabData,
                 weights = wgt, attrib.var = age, attrib = TRUE, type.attrib = "both",
                 set.seed = TRUE, seed = 111, bootstrap = TRUE, nbootstrap = 10,
                 parallel = TRUE, type.parallel = "snow", ncpus = 4)
summary(fit3)
## End(Not run)
```

disabData

Example of disability data

Description

The `disabData` is a subset of the data from the 2013 National Health Survey in Brazil ("Pesquisa Nacional de Saude, 2013"). The data are restricted to women aged 60 years or older, resulting in 6294 individuals.

Usage

```
data(disabData)
```

Format

This data set has information about disability and chronic conditions. The disability outcomes were defined based on limitations on instrumental activities of daily living (IADL). Individuals with missing data were excluded. The data frame contains 7 variables:

- **dis.bin**: disability as a binary variable, with 2 categories: 0 (no disability), 1 (disability).
- **dis.mult**: disability as a multinomial variable, with 3 categories: 0 (no disability), 1 (mild disability), and 2 (severe disability).
- **wgt**: survey weights.
- **age**: binary variable for age: 0 (60-79 years) or 1 (80+ years).
- **diab**: binary variable for diabetes: 0 (no) or 1 (yes).
- **arth**: binary variable for arthritis: 0 (no) or 1 (yes).
- **stro**: binary variable for stroke: 0 (no) or 1 (yes).

Source

The data were obtained from the National Health Survey 2013, Brazil. For more information about the survey, see references.

References

Szwarcwald, C.L., Malta, D.C., Pereira, C.A., Vieira, M.L., Conde, W.L., Souza-Junior, P.R., et al. (2013). National Health Survey in Brazil: design and methodology of application. *Cien Saude Colet.*, 19(2): 333:42 [Article in Portuguese].

Instituto Brasileiro de Geografia e Estatística (IBGE). Pesquisa Nacional de Saude 2013. Available at <<http://www.ibge.gov.br/home/estatistica/populacao/pns/2013/>>.

Examples

```
data(disabData)
str(disabData)
```

MultAddHaz

Fit Multinomial Additive Hazards Models

Description

This function fits multinomial additive hazards models subject to linear inequality constraints using the function `constrOptim` in the `stats` package for multinomial (multi-category) outcomes. It also calculates the cause-specific contributions to the disability prevalence for each category of the response variable based on the extension of the attribution method, as proposed by Yokota, Van Oyen, and Molenberghs (2015).

Usage

```
MultAddHaz(formula, data, subset, weights, na.action, model = TRUE,
            contrasts = NULL, start.val, attrib = FALSE, attrib.var,
            type.attrib = "abs", set.seed = FALSE, seed, bootstrap = FALSE,
            nbootstrap, parallel = FALSE, type.parallel = "snow", ncpus = 4,...)
```

Arguments

<code>formula</code>	A formula expression of the form <code>response ~ predictors</code> , similar to other regression models. In case of <code>attrib = TRUE</code> , the first predictor in the formula should be the <code>attrib.var</code> . See example.
<code>data</code>	an optional data frame or matrix containing the variables in the model. If not found in <code>data</code> , the variables are taken from <code>environment(formula)</code> , typically the environment from which <code>MultAddHaz</code> is called.
<code>subset</code>	an optional vector specifying a subset of observations to be used in the fitting process. All observations are included by default.
<code>weights</code>	an optional vector of survey weights to be used in the fitting process.

<code>na.action</code>	a function which indicates what should happen when the data contain NAs. The default is set by the <code>na.action</code> setting of options, and is <code>na.fail</code> if that is unset. The 'factory-fresh' default is <code>na.omit</code> .
<code>model</code>	logical. If TRUE, the model frame is included as a component of the returned object.
<code>contrasts</code>	an optional list to be used for some or all of the factors appearing as variables in the model formula.
<code>start.val</code>	vector of the starting values. Starting values are required and they should lie inside the parameter space.
<code>attrib</code>	logical. Should the attribution of chronic diseases/conditions for each disability level be calculated? Default is FALSE.
<code>attrib.var</code>	character indicating the name of the attribution variable to be used if <code>attrib = TRUE</code> . If missing, the attribution variable is the intercept (no stratification of the attribution results). The attribution variable must be the first variable included in the linear predictor in formula. See example.
<code>type.attrib</code>	type of attribution to be calculated. The options are "abs" for absolute contribution, "rel" for relative contribution, or "both" for both absolute and relative contributions. Default is "abs".
<code>set.seed</code>	logical. Should a random seed be set? Default is FALSE.
<code>seed</code>	integer indicating the random seed. Only valid if <code>set.seed = TRUE</code> .
<code>bootstrap</code>	logical. Should bootstrap percentile confidence intervals be calculated for the parameters and attributions? Default is FALSE. See details.
<code>nbootstrap</code>	integer. Number of bootstrap replicates.
<code>parallel</code>	logical. Should parallel calculations be used to obtain the bootstrap percentile confidence intervals? Default is FALSE.
<code>type.parallel</code>	type of parallel operation to be used (if <code>parallel = TRUE</code>), with options: "multicore" and "snow". Default is "snow". See details.
<code>ncpus</code>	integer. Number of processes to be used in parallel operation: typically one would choose this to the number of available CPUs. Default is 4.
<code>...</code>	other arguments passed to or from the other functions.

Details

The model is a generalized linear model with a non-canonical link function, which requires a restriction on the linear predictor ($\eta \geq 0$) to produce valid probabilities. This restriction is implemented in the optimization procedure, with an adaptive barrier algorithm, using the function `constrOptim` in the `stats` package.

The variance-covariance matrix is based on the observed information matrix.

This version of the the package only allows the calculation of non-parametric bootstrap percentile confidence intervals (CI), which are based on the 2.5 and 97.5 percentiles. Stratified bootstrap is applied to each category of the outcome. Also, the function gives the user the option to do parallel calculation of the bootstrap CI. The `snow` parallel option is available for all operating systems (Windows, Linux, and Mac OS) while the `multicore` option is only available for Linux and Mac

OS systems. These two calculations are done by calling the boot function in the boot package. For more details see the documentation of the boot package.

More information about the multinomial additive hazards model and the calculation of the contribution of chronic conditions to the disability prevalence can be found in the references.

Value

A list with arguments:

coefficients	column matrix with the regression coefficients.
ci	matrix with confidence intervals calculated via bootstrapping (if bootstrap = TRUE) or as the inverse of the observed information matrix.
resDeviance	residual deviance.
df	degrees of freedom.
pvalue	column matrix of p-values based on the Wald test. Only provided if bootstrap = FALSE.
stdError	column matrix with the standard errors for the parameter estimates based on the inverse of the observed information matrix. Only provided if bootstrap = FALSE.
vcov	variance-covariance matrix (inverse of the observed information matrix). Only provided if bootstrap = FALSE.
contribution	for type.attrib = "abs" or "rel", a matrix is provided. For type.attrib = "both", a list with two matrices ("abs" and "rel") is provided. This represents the contribution of each predictor in the model (usually diseases or conditions) to the disability prevalence. Percentile bootstrap confidence intervals are provided if bootstrap = TRUE.
bootsRep	matrix with the bootstrap replicates.
bootstrap	logical. Was bootstrap CI requested?
call	the matched call.

Author(s)

Renata T C Yokota. This function is based on the R code developed by Caspar W N Looman and Wilma J Nusselder for the binomial additive hazards model with modifications and adaptations for the multinomial case.

References

Yokota, R.T.C., Van Oyen, H., Molenberghs, G. (2015). Use of a multinomial additive hazards model to assess the disability burden using cross-sectional data. Available at <<https://uhdspace.uhasselt.be/dspace/bitstream/1942/19407/1/11321252014093.pdf>>.

See Also

[BinAddHaz](#)

Examples

```
data(disabData)

## Model without bootstrap percentile confidence intervals and no attribution
## Not run:
fit1 <- MultAddHaz(dis.mult ~ diab + arth + stro , data = disabData, weights = wgt,
                  start = TRUE, start.val = c(rep(0.5, 4), rep(0.3, 4)))
summary(fit1)

## Model with bootstrap percentile confidence intervals and attribution without
## stratification, no parallel calculation
# Warning message due to the low number of bootstrap replicates

fit2 <- MultAddHaz(dis.mult ~ diab + arth + stro , data = disabData, weights = wgt,
                  start = TRUE, start.val = c(rep(0.5, 4), rep(0.3, 4)), attrib = TRUE,
                  type.attrib = "both", set.seed = TRUE, seed = 111, bootstrap = TRUE,
                  nbootstrap = 5)
summary(fit2)

## Model with bootstrap percentile confidence intervals with attribution based
## on the age variable and parallel calculation of bootstrap CI
# Warning message due to the low number of bootstrap replicates

diseases <- as.matrix(disabData[,c("diab", "arth", "stro")])

fit3 <- MultAddHaz(dis.mult ~ factor(age) -1 + diseases: factor(age),
                  data = disabData, weights = wgt, start = TRUE,
                  start.val = c(rep(0.5, 8), rep(0.3, 8)), attrib.var = age,
                  attrib = TRUE, type.attrib = "both", set.seed = TRUE,
                  seed = 111, bootstrap = TRUE, nbootstrap = 5, parallel = TRUE,
                  type.parallel = "snow", ncpus = 4)

summary(fit3)
## End(Not run)
```

Index

- *Topic **datasets**
 - disabData, [5](#)
 - *Topic **models**
 - BinAddHaz, [2](#)
 - MultAddHaz, [6](#)
 - *Topic **regression**
 - BinAddHaz, [2](#)
 - MultAddHaz, [6](#)
 - *Topic **survey**
 - BinAddHaz, [2](#)
 - disabData, [5](#)
 - MultAddHaz, [6](#)
- BinAddHaz, [2](#), [8](#)
- disabData, [5](#)
- MultAddHaz, [4](#), [6](#)