Package ‘xhaz’

September 19, 2022

Title Excess Hazard Modelling Considering Inappropriate Mortality Rates

Version 2.0.1

Description Fits relative survival regression models with or without proportional excess hazards and with the additional possibility to correct for background mortality by one or more parameter(s). These models are relevant when the observed mortality in the studied group is not comparable to that of the general population or in population-based studies where the available life tables used for net survival estimation are insufficiently stratified. In the latter case, the proposed model by Touraine et al. (2020) <doi:10.1177/0962280218823234> can be used. The user can also fit a model that relax the proportional expected hazards assumption considered in the Touraine et al. excess hazard model. This extension was proposed by Mba et al. (2020) <doi:10.1186/s12874-020-01139-z> to allow non-proportional effects of the additional variable on the general population mortality. In non-population-based studies, researchers can identify non-comparability source of bias in terms of expected mortality of selected individuals. A proposed excess hazard model correcting this selection bias is presented in Goungounga et al. (2019) <doi:10.1186/s12874-019-0747-3>.

License AGPL (>= 3)
Depends R (>= 3.5.0), statmod, stats, survival
Imports gtools, numDeriv, optimParallel, splines, stringr, survexp.fr
Suggests knitr, rmarkdown, spelling, testthat (>= 2.1.0)
VignetteBuilder knitr
Encoding UTF-8
LazyData true
RoxygenNote 7.2.1
Language en-US

NeedsCompilation no

Author Juste Goungounga [aut, cre] (<https://orcid.org/0000-0002-9039-2639>),
Darlin Mba [aut] (<https://orcid.org/0000-0001-9768-0230>),
Nathalie Graffeo [aut] (<https://orcid.org/0000-0001-7227-7525>),
Roch Giorgi [aut] (<https://orcid.org/0000-0001-6135-3078>)

Maintainer Juste Goungounga <juste.goungounga@univ-amu.fr>
Repository CRAN
Date/Publication 2022-09-19 08:10:02 UTC
Description

Calculates the Akaike’s ‘An Information Criterion’ for fitted models from xhaz.

Usage

## S3 method for class 'bsplines'
AIC(object, ..., k = 2)

Arguments

object

a fitted model object obtained from xhaz function

...  
on Optionally more fitted model objects obtained from xhaz function

k

numeric, the penalty per parameter to be used; the default k = 2 is the classical AIC.
Value
the value corresponds to the AIC calculated from the total log-likelihood of the fitted model if just one object is provided. If multiple objects are provided, a data.frame with columns corresponding to the objects and rows representing the number of parameters in the model (df) and the AIC.

Examples

library("xhaz")

# Giorgi et al model: baseline excess hazard is a quadratic Bsplines
# function with two interior knots and allow here a linear and proportional effects for the covariates on baseline excess hazard.
levels(simuData$sex) <- c("male", "female")

fitphBS <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData,
ratetable = survexp.us,
interval = c(0, NA, NA, 6),
rmap = list(age = "age", sex = "sex", year = "date"),
baseline = "bsplines", pophaz = "classic")

fitphBS
AIC(fitphBS)

AIC.constant

Akaike's An Information Criterion for excess hazard model with baseline hazard following a piecewise constant function

Description
Calculates the Akaike's 'An Information Criterion' for fitted models from xhaz.

Usage

## S3 method for class 'constant'
AIC(object, ..., k = 2)

Arguments

object a fitted model object obtained from xhaz function
... optionally more fitted model objects obtained from xhaz function
k numeric, the penalty per parameter to be used; the default k = 2 is the classical AIC.
Value

the value corresponds to the AIC calculated from the total log-likelihood of the fitted model if just one object is provided. If multiple objects are provided, a data.frame with columns corresponding to the objects and rows representing the number of parameters in the model (df) and the AIC

Examples

library("xhaz")

# Esteve et al. model: baseline excess hazard is a piecewise function
# linear and proportional effects for the covariates on baseline excess hazard.
levels(simuData$sex) <- c("male", "female")

set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData2,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "classic")

fit.estv2
AIC(fit.estv2)

Description

This function compute an analysis of deviance table for two excess hazard models fitted using xhaz R package.

Usage

## S3 method for class 'bsplines'
anova(object, ..., test = "LRT")
Arguments

object an object of class bsplines
... an object of class bsplines
test a character string. The appropriate test is a likelihood-ratio test, all other choices result in Not yet implemented test.

Value

An object of class anova inheriting from class matrix. The different columns contain respectively the degrees of freedom and the log-likelihood values of the two nested models, the degree of freedom of the chi-square statistic, the chi-square statistic and the p-value of the likelihood ratio test.

Note

As expected, the comparison between two or more models by anova or more excess hazard models will only be valid if they are fitted to the same dataset, and if the compared models are nested. This may be a problem if there are missing values.

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Grafféo and Roch Giorgi

References


See Also

xhaz, summary.bsplines, print.constant
Examples

# load the data set in the package
library("survival")
library("numDeriv")
library("survexp.fr")
library("statmod")
data("dataCancer", package = "xhaz")  # load the data set in the package

fit.phBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
  data = dataCancer,
  ratetable = survexp.fr::survexp.fr,
  interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "bsplines", pophaz = "classic")

fit.nphBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + qbs(immuno_trt),
  data = dataCancer,
  ratetable = survexp.fr::survexp.fr,
  interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "bsplines", pophaz = "classic")

anova(fit.phBS, fit.nphBS)

anova.constant

anova.constant function used for likelihood-ratio Test of two models from xhaz function

Description

This function compute an analysis of deviance table for two excess hazard models fitted using xhaz R package.

Usage

## S3 method for class 'constant'
anova(object, ..., test = "LRT")
Arguments

object an object of class constant

... an object of class constant
test a character string. The appropriate test is a likelihood-ratio test, all other choices result in Not yet implemented test.

Value

An object of class anova inheriting from class matrix. The different columns contain respectively the degrees of freedom and the log-likelihood values of the two nested models, the degree of freedom of the chi-square statistic, the chi-square statistic and the p-value of the likelihood ratio test.

Note

As expected, the comparison between two or more models by anova or more excess hazard models will only be valid if they are fitted to the same dataset, and if the compared models are nested. This may be a problem if there are missing values.

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Grafféo and Roch Giorgi

References


See Also

xhaz, summary.bsplines, print.constant
Examples

# load the data set in the package
library("survival")
library("numDeriv")
library("survexp.fr")

data("dataCancer")  # load the data set in the package

fit.ph <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
  data = dataCancer,
  ratetable = survexp.fr::survexp.fr,
  interval = c(0, NA, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "constant", pophaz = "classic")

fit.ph2 <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre,
  data = dataCancer,
  ratetable = survexp.fr::survexp.fr,
  interval = c(0, NA, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "constant", pophaz = "classic")

anova(fit.ph2, fit.ph)

BIC.bsplines

Bayesian Information Criterion for excess hazard model with baseline
hazard following a B-splines functions

Description

Calculates the Bayesian Information Criterion’ for fitted models from xhaz.

Usage

## S3 method for class 'bsplines'
BIC(object, ...)

Arguments

object an a fitted model object obtained from xhaz function
... optionally more fitted model objects obtained from xhaz function
BIC.constant

Value

the value corresponds to the BIC calculated from the total log-likelihood of the fitted model if just one object is provided. If multiple objects are provided, a data.frame with columns corresponding to the objects and rows representing the number of parameters in the model (df) and the BIC.

Examples

library("xhaz")

# Giorgi et al model: baseline excess hazard is a quadratic Bsplines function with two interior knots and allow here a linear and proportional effects for the covariates on baseline excess hazard.

levels(simuData$sex) <- c("male", "female")

fitphBS <- xhaz(formula = Surv(time_year, status) ~ agec + race, 
                 data = simuData, 
                 ratetable = survexp.us, 
                 interval = c(0, NA, NA, 6), 
                 rmap = list(age = "Var", sex = "Var", year = "Var"),
                 baseline = "bsplines", pophaz = "classic")

fitphBS

BIC(fitphBS)

Description

Calculates the Bayesian Information Criterion’ for fitted models from xhaz.

Usage

## S3 method for class 'constant'
BIC(object, ...)

Arguments

object a fitted model object obtained from xhaz function

... optionally more fitted model objects obtained from xhaz function

Value

the value corresponds to the BIC calculated from the total log-likelihood of the fitted model if just one object is provided. If multiple objects are provided, a data.frame with columns corresponding to the objects and rows representing the number of parameters in the model (df) and the BIC.
Examples

```r
library("xhaz")

# Esteve et al. model: baseline excess hazard is a piecewise function
# linear and proportional effects for the covariates on
# baseline excess hazard.

levels(simuData$sex) <- c("male", "female")

set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]
fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
    data = simuData2,
    ratetable = survexp.us,
    interval = c(0, NA, NA, NA, NA, NA, 6),
    rmap = list(age = 'age', sex = 'sex', year = 'date'),
    baseline = "constant", pophaz = "classic")

fit.estv2
BIC(fit.estv2)
```

---

**dataCancer**

*Simulated data with cause death information with non comparability bias in term of individuals expected hazard*

**Description**

Simulated data

**Usage**

`data(dataCancer)`

**Format**

This dataset contains the following variables:

- **obs_time** Follow-up time (months)
- **obs_time_year** Follow-up time (years)
- **event** Vital status
- **age** Age at diagnosis
- **agegrp** "<30", "30_60" and ">=60" age groups
- **ageCentre** centered age at diagnosis
sexx  Sex(Female,Male).
immuno_trt  Treatment group
year_date  date of diagnosis.

References


Examples
data(dataCancer)
summary(dataCancer)

duplicate(status, event, data)

Description
Duplicate data for survival analysis in the context of competing risks, where an individual can experience only one of alternative events, using the Lunn & McNeil (Biometrics, 1995) approaches.

Duplication of data proceeds as follows: Suppose that we study \( J \) distinct types of events. Each observation concerning a given subject is duplicated \( J \) times, with one row for each type of event. In addition, \((J-1)\) dummy variables are created, each indicating the type of event in relation with that observation (\( \delta_{j}=1 \) if the event of type \( j \) is the observed one and \( 0 \) otherwise). Since, for a given subject, only the first occurring event is considered, the status indicator equals 1 for that event and 0 for all the others. In the case of a censored observation (dropout or administrative censoring), the same principle applies also: duplication of each subject’s data is made \( J \) times with \((J-1)\) dummy variables and a status indicator equal to 0 for all observations.

Usage
duplicate(status, event, data)
**Arguments**

status  
the censoring status indicator (numeric vector), 0=alive, 1=dead.

event  
the indicator of the event type (numeric vector). By default, the event==0 acts as the censoring indicator.

data  
a data frame containing the data to duplicate.

**Value**

A data.frame containing the duplicated data with the new dummy variables, named `delta.number_of_the_event`, indicating the type of event.

**Author(s)**

Roch Giorgi

**References**


**Examples**

```r
## Create the simplest test data set
data1 <- data.frame(futime = c(1, 2, 5, 2, 1, 7, 3, 4, 8, 8),
                   fustat = c(0, 1, 1, 0, 0, 1, 0, 1, 0, 1),
                   firstevent = c(0, 2, 1, 2, 0, 0, 1, 0, 2, 2),
                   x = c(1, 0, 0, 1, 0, 1, 1, 1, 0, 0))

## Duplicate data1 with firstevent == 0 as the censoring indicator.
dupli.data <- duplicate(status=fustat, event=firstevent, data=data1)

data2 <- data.frame(futime = c(10, 2, 7, 3, 4, 9, 13, 2, 5, 9),
                   fustat = c(0, 1, 1, 0, 0, 1, 0, 1, 0, 1),
                   firstevent = c(3, 2, 1, 2, 3, 3, 1, 3, 2, 2),
                   x = c(1, 0, 0, 1, 0, 1, 1, 1, 0, 0))

## Duplicate data1 with firstevent == 3 as the censoring indicator.
dupli.data <- duplicate(status = fustat, event = firstevent == 3, data = data2)

# Joint modeling
coxph(Surv(futime, fustat) ~ delta.2 + x + delta.2:(x), data = dupli.data)
coxph(Surv(futime, fustat) ~ delta.1 + x + delta.1:(x), data = dupli.data)
```
Description

Calculate the expected hazard and survival.

Usage

```r
exphaz(
  formula = formula(data),
  data = sys.parent(),
  ratetable,
  rmap = list(age = NULL, sex = NULL, year = NULL),
  ratedata = sys.parent(),
  only_ehazard = TRUE,
  subset,
  na.action,
  scale = 365.2425
)
```

Arguments

- **formula**: a formula object of the `Surv` function with the response on the left of a `~` operator and the terms on the right. The response must be a survival object as returned by the `Surv` function (time in first and status in second).
- **data**: a data frame in which to interpret the variables named in the formula.
- **ratetable**: a rate table stratified by age, sex, year (if missing, ratedata is used).
- **rmap**: a list that maps data set names to the ratetable names.
- **ratedata**: a data frame of the hazards mortality in general population.
- **only_ehazard**: a boolean argument (by default, `only_ehazard=TRUE`). If TRUE, the cumulative population hazard is not provided.
- **subset**: an expression indicating which subset of the rows in data should be used in the fit. All observations are included by default.
- **na.action**: a missing-data filter function. The default is `na.fail`, which returns an error if any missing values are found. An alternative is `na.exclude`, which deletes observations that contain one or more missing values.
- **scale**: a numeric argument specifying if the ratetable contains death rates per day (default `scale = 365.2425`) or death rates per year (`scale = 1`).
Value

An object of class list containing the following components:

- `ehazard`: expected hazard calculated from the matching ratetable.
- `ehazardInt`: cumulative expected hazard calculated from the matching ratetable. If `only_ehazard=TRUE`, this quantity is not provided.
- `dateDiag`: date of diagnosis

Note

Time is OBLIGATORY in YEARS.

References


Examples

```r
library(survexp.fr)
library(xhaz)
fit.haz <- exphaz(
  formula = Surv(obs_time_year, event) ~ 1,
  data = dataCancer,
  ratetable = survexp.fr, only_ehazard = TRUE,
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date')
)
```

Description

To plot the log hazard ratio functions for non-proportional hazards model
Usage

## S3 method for class 'bsplines'
plot(
x, 
cov, 
conf.int = TRUE, 
baseline = FALSE, 
xrange, 
yrange, 
xlegend, 
ylegend, 
glegend, 
xaxs = NULL, 
add = FALSE, 
col = 1, 
lty = 1, 
lwd = 1, 
...
)

Arguments

x An object of class xhaz

cov specify covariates for which a plot is required.

conf.int a vector of logical values indicating whether (if TRUE) confidence intervals will be plotted. The default is to do so if the plot concerns only one curve.

baseline a vector of logical values indicating whether (if baseline = TRUE) to plot the curve for the baseline group. Default is FALSE, except if cov is unspecified.

xrange vector indicating the minimum and the maximum values of the x axis. By default, these values are automatically calculated for the first plot (i.e before the use of add argument).

yrange vector indicating the minimum and the maximum values of the y axis. By default, these values are automatically calculated for the first plot (i.e before the use of add argument).

xlegend value indicating the location of the legend over x axis. By default, location at the left of the plot.

ylegend value indicating the location of the legend over y axis. By default, location at the top of the plot.

glegend vectors of names attributed to each lines of the excess hazard to be displayed in the plot. If (baseline = TRUE), glegend is "baseline".

xaxs the x axis style, as listed in 'par'. Survival curves are traditionally drawn with the curve touching the bounding box on the left edge, but not touching it on the right edge. This corresponds to neither of the two standard S axis styles of "e" (neither touches) or "i" (both touch). If xaxis is missing or NULL the internal axis style is used (xaxs= i) but only after the right endpoint has been extended.
add

A logical value indicating whether to add the survival curves to the current plot (if add = TRUE). Default is FALSE.

col

A vector of integers specifying colors for each curve. The default value is 1.

lty

A vector of integers specifying line types for each curve. The default value is fixed by the number of covariates (plus 1 if baseline = TRUE).

lwd

A vector of numeric values for line widths. The default value is 1.

... additional arguments affecting the plot function

Value

The return of this function produce graphics of log hazard ratio functions for non-proportional hazards model

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Grafféo and Roch Giorgi

References


Examples

```r
# load the data set in the package
library("xhaz")
library("survexp.fr")

# load the data set in the package
data("dataCancer", package = "xhaz")

fit.nphBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + qbs(immuno_trt),
  data = dataCancer,
  ratetable = survexp.fr,
  # load the data set in the package
```

plot.predxhaz

interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
rmap = list(age = 'age', sex = 'sex', year = 'year_date'),
baseline = "bsplines", pophaz = "classic")

plot(fit.nphBS, cov = "immuno_trt", col = "blue", baseline = FALSE)

---

**Description**

Function to plot excess hazard or net survival

**Usage**

```r
## S3 method for class 'predxhaz'
plot(x, what = "survival", ...)
```

**Arguments**

- `x` An object of class predxhaz
- `what` allow to choose between excess hazard (what="hazard") or net survival (what="survival").
- `...` additional arguments affecting the plot function

**Value**

The return of this function produce graphics of excess hazard or net survival, or time-dependent effects, when times.pts argument is provided in prediction call.

**Author(s)**

Juste Goungounga, Robert Darlin Mba, Nathalie Grafféo and Roch Giorgi

**References**


Examples

```r
data("dataCancer")  # load the data set in the package
library("survival")
library("numDeriv")
library("survexp.fr")
data("simuData", package = "xhaz")  # load the data sets 'simuData'
levels(simuData$sex) <- c("male", "female")  # define the levels of variable sex

fit.estv1 <- xhaz(formula = Surv(time_year, status) ~ agec + race,  
data = simuData, ratetable = survexp.us,  
interval = c(0, NA, NA, NA, NA, max(simuData$time_year)),  
rmap = list(age = 'age', sex = 'sex', year = 'date'),  
baseline = "constant", pophaz = "classic")

predict_est <- predict(object = fit.estv1,  
new.data = simuData,  
times.pts = c(seq(0, 4, 0.1)),  
baseline = TRUE)

plot(predict_est, what = "survival",  
xlab = "time since diagnosis (year)",  
ylab = "net survival", ylim = c(0, 1))
data("dataCancer", package = "xhaz")  # load the data set in the package

fit.phBS <- xhaz(  
formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,  
data = dataCancer, ratetable = survexp.fr::survexp.fr,  
interval = c(0, NA, NA, max(dataCancer$obs_time_year)),  
rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),  
baseline = "bsplines", pophaz = "classic")

predict_mod1 <- predict(object = fit.phBS, new.data = dataCancer,  
times.pts = c(seq(0, 10, 0.1)), baseline = FALSE)

old.par <- par(no.readonly = TRUE)
par(mfrow = c(2, 1))

plot(predict_mod1, what = "survival",  
xlab = "time since diagnosis (year)",  
ylab = "net survival", ylim = c(0, 1))

plot(predict_mod1, what = "hazard",  
xlab = "time since diagnosis (year)",  
ylab = "hazard", ylim = c(0, 1))
```
Function to predict excess hazard and net survival based on an object of class `bsplines`. The function allows the predictions at several time points but not exceeding the maximum time of follow-up from the baseline model.

### Usage

```r
## S3 method for class 'bsplines'
predict(object, new.data = NULL, times.pts = NULL, baseline = TRUE, ...)
```

### Arguments

- **object**: an object of class `bsplines`
- **new.data**: new.data where is covariates
- **times.pts**: time in year scale to calculate the excess hazard. The default value is `NULL`. In this case, time variable must be provided in the `new.data`
- **baseline**: default is survival baseline; put `baseline = FALSE` to estimate the net survival with covariates
- **...**: additional arguments affecting the predictions of excess hazard and net survival

### Value

An object of class `predxhaz`, which is a list of data.frame. Each element of the list contains the estimates of hazard and survival at a fixed time point. The return of this function can be used to produce graphics of excess hazard or net survival, when `times.pts` argument is provided. This object contains:

- **times.pts**: the times value in year at which the excess hazard and or the net survival have been estimated
- **hazard**: the excess hazard values based on the model of interest
- **survival**: the net survival values based on the model of interest

### Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Grafféo and Roch Giorgi
**References**


**See Also**

`xhaz, print.bsplines, print.constant`

**Examples**

```r
library("survival")
library("numDeriv")
library("survexp.fr")
library("splines")
data("dataCancer", package = "xhaz")  # load the data set in the package

fit.phBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
  data = dataCancer, ratetable = survexp.fr,
  interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "bsplines", pophaz = "classic")

print(fit.phBS)

predicted <- predict(object = fit.phBS,
  new.data = dataCancer[1:10,,
  times.pts = c(seq(0,10,1)),
  baseline = TRUE)

# a list of predicted hazard and survival at different time points
print(predicted)

# predicted hazard and survival at time points 10 years
```
predict.constant

    print(predicted[[10]])

predict.constant  Predictions of excess hazard and net Survival from an constant object

Description

Function to predict excess hazard and net survival based on an object of class constant. The function allows the predictions at several time points but not exceeding the maximum time of follow-up from the baseline model.

Usage

## S3 method for class 'constant'

predict(object, new.data = NULL, times.pts = NULL, baseline = TRUE, ...)

Arguments

object          An object of class constant
new.data       new.data where is covariates
times.pts      time in year scale to calculate the excess hazard. The default value is NULL. In this case, time variable must be provided in the new.data
baseline       default is survival baseline; put baseline = FALSE to estimate the net survival with covariates
...            additional arguments affecting the predictions of excess hazard and net survival

Value

An object of class predxhaz. The return of this fonction can be used to produce graphics of excess hazard or net survival, when times.pts argument is provided. This object contains:

times.pts      the times value in year at which the excess hazard and or the net survival have been estimated
hazard         the excess hazard values based on the model of interest
survival       the net survival values based on the model of interest

Author(s)

Juste Goungounga, Robert Darlin Mba, Nathalie Grafféo and Roch Giorgi
References


See Also

xhaz, print.bsplines, print.constant

Examples

# load the data set in the package
library("xhaz")
library("numDeriv")

# load the data sets 'simuData'
data("simuData", package = "xhaz")

#define the levels of variable sex
levels(simuData$sex) <- c("male", "female")

# Esteve et al. model
set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData2,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "classic")

predict_est <- predict(object = fit.estv2,
new.data = simuData2,
times.pts = c(seq(0, 4, 1)),
baseline = TRUE)

predict_est
print.bsplines  A print.bsplines Function used to print a object of class bsplines

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for
the covariate effects

Usage

## S3 method for class 'bsplines'
print(x, digits = max(options()$digits - 4, 3), ...)

Arguments

x  an object of class bsplines

digits  minimal number of significant digits.

...  addionnal parameters which can be used in the print function

Value

Estimated parameters of the model in different scales for interpretation purposes.

References

Goungounga JA, Touraine C, Grafféo N, Giorgi R; CENSUR working survival group. Correcting
for misclassification and selection effects in estimating net survival in clinical trials. BMC Med Res
PMC6524224. (PubMed)

Touraine C, Grafféo N, Giorgi R; CENSUR working survival group. More accurate cancer-related
excess mortality through correcting background mortality for extra variables. Stat Methods Med
30674229. (PubMed)

Mba RD, Goungounga JA, Grafféo N, Giorgi R; CENSUR working survival group. Correcting
inaccurate background mortality in excess hazard models through breakpoints. BMC Med Res
PMC7596976. (PubMed)

See Also

xhaz, plot.predxhaz, print.constant
Examples

```r
library("xhaz")
library("survival")
library("numDeriv")
library("survexp.fr")
library("splines")
data("dataCancer", package = "xhaz") # load the data set in the package

fit.phBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
  data = dataCancer, ratetable = survexp.fr,
  interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "bsplines", pophaz = "classic")

print(fit.phBS)
```

---

**print.constant**  
*A print.constant Function used to print a object of class constant*

**Description**

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

**Usage**

```r
## S3 method for class 'constant'
print(x, ci_type = "lognormal", digits = max(options()$digits - 4, 3), ...)
```

**Arguments**

- `x` an object of class xhaz.constant  
- `ci_type` method for confidence intervals calculation  
- `digits` minimal number of significant digits.  
- `...` additionnal parameters which can be used in the print function

**Value**

Estimated parameters of the model in different scales for interpretation purposes.

**See Also**

`xhaz`, `summary.constant`, `print.bsplines`
Examples

```r
library("numDeriv")
library("survexp.fr")

data("simuData", "rescaledData", "dataCancer")
# load the data sets 'simuData', 'rescaledData' and 'dataCancer'.

# Esteve et al. model: baseline excess hazard is a piecewise function
# linear and proportional effects for the covariates on
# baseline excess hazard.

levels(simuData$sex) <- c("male", "female")
set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData2,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "classic")

print(fit.estv2)
```

---

**print.predxhaz**

*A print.predxhaz Function used to print a object of class predxhaz*

---

**Description**

This function present the print of the predict function

**Usage**

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

**Arguments**

- `x` an object of class predxhaz
- `...` other parameters used for print function
Value

an object of class data.frame containing the following components:

- times.pts The time at which the estimations of excess hazard and net survival are predicted
- hazard the predicted excess hazard at the fixed times
- survival the predicted net survival at the fixed times

Examples

```r
library("xhaz")
library("survexp.fr")
library("splines")

data("dataCancer", package = "xhaz") # load the data set in the package

fit.phBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
  data = dataCancer, ratetable = survexp.fr,
  interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
  baseline = "bsplines", pophaz = "classic")

fit.phBS

predicted <- predict(object = fit.phBS,
  new.data = dataCancer[1:10,],
  times.pts = c(seq(0,10,1)),
  baseline = TRUE)

# a list of predicted hazard and survival at different time points
print(predicted)

# predicted hazard and survival at time points 10 years
print(predicted[[10]])
```

qbs function
Description

A function indicating which covariates have a time-dependent effect in the formula.

Usage

`qbs(x)`

Arguments

`x`  
a covariate to be considered in the `xhaz` formula with a time-dependent effect. Quadratic B-splines with two interior knots are used.

Value

No return value, called for side effects.

Examples

```r
library("xhaz")
library("numDeriv")
library("survexp.fr")
library("splines")

fit.tdphBS <- xhaz(
  formula = Surv(obs_time_year, event) ~ ageCentre + qbs(immuno_trt),
  data = dataCancer, ratetable = survexp.fr,
  interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
  rmap = list(age = 'age', sex = 'sex', year = 'year_date'),
  baseline = "bsplines", pophaz = "classic")

print(fit.tdphBS)
```

Simulation of individual expected hazards

**Description**

Simulated data with cause death information with non-comparability bias in terms of individuals expected hazard.

**Usage**

`data(rescaledData)`
**Format**

This dataset contains the following variables:

- **time**: Follow-up time (months)
- **status**: Vital status
- **age**: Age at diagnosis
- **age.c**: Centred age
- **sex**: Sex (Female, Male)
- **hormTh**: Treatment group variable
- **date**: Date of diagnosis

**References**


**Examples**

```r
data(rescaledData)
summary(rescaledData)
```

---

**simuData**

*Simulated data with cause death information in long term follow-up setting without non comparability bias in term of individuals expected hazard*

---

**Description**

Simulated data

**Usage**

```r
data(simuData)
```

**Format**

This dataset contains the following variables:

- **age**: Age at diagnosis
- **agec**: Centered age
- **sex**: Sex (Female, Male)
- **race**: Race
- **date**: Date of diagnosis.
- **time**: Follow-up time (months)
- **time_year**: Follow-up time (years)
- **status**: Vital status
References


Examples

```r
data(simuData)
summary(simuData)
```

summary.bsplines  
A `summary.bsplines` Function used to print a object of class `bsplines`

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

Usage

```r
## S3 method for class 'bsplines'
summary(object, ...) 
```

Arguments

- `object` an object of class `bsplines`
- `...` additionnal parameters which can be used in the `summary` function

Value

Estimated parameters of the model in different scales for interpretation purposes.

References


See Also

xhaz, summary.bsplines, plot.bsplines

Examples

library("xhaz")
library("survival")
library("numDeriv")
library("survexp.fr")
library("splines")

data("dataCancer", package = "xhaz") # load the data set in the package

fit.phBS <- xhaz(
    formula = Surv(obs_time_year, event) ~ ageCentre + immuno_trt,
    data = dataCancer, ratetable = survexp.fr,
    interval = c(0, NA, NA, max(dataCancer$obs_time_year)),
    rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
    baseline = "bsplines", pophaz = "classic")

summary(fit.phBS)

summary.constant A summary.constant Function used to print a object of class xhaz.constant

Description

This function present the estimated coefficients for the excess hazard baseline coefficient and for the covariate effects

Usage

## S3 method for class 'constant'
summary(object, ci_type = "lognormal", ...)

Arguments

object an object of class xhaz.constant
ci_type method for confidence intervals calculation
... additionnal parameters which can be used in the print function
Value

Estimated parameters of the model in different scales for interpretation purposes.

See Also

xhaz, summary.constant, print.bsplines

Examples

library("xhaz")
library("numDeriv")
data("simuData", package = "xhaz")  # load the data sets 'simuData'

# Esteve et al. model: baseline excess hazard is a piecewise function
# linear and proportional effects for the covariates on
# baseline excess hazard.

levels(simuData$sex) <- c("male", "female")
set.seed(1980)
simuData2 <- simuData[sample(nrow(simuData), size = 500), ]

fit.estv2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData2,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "classic")

summary(fit.estv2)

xhaz

xhaz function

Description

Usage

xhaz(
    formula = formula(data),
    data = sys.parent(),
    ratetable,
    rmap = list(age = NULL, sex = NULL, year = NULL),
    baseline = c("constant", "bsplines"),
    pophaz = c("classic", "rescaled", "corrected"),
    only_ehazard = FALSE,
    add.rmap = NULL,
    add.rmap.cut = list(breakpoint = FALSE, cut = NA, probs = NULL, criterion = "BIC",
        print_stepwise = FALSE),
    interval,
    ratedata = sys.parent(),
    subset,
    na.action,
    init,
    control = list(eps = 1e-04, iter.max = 800, level = 0.95),
    optim = TRUE,
    scale = 365.2425,
    trace = 0,
    speedy = FALSE,
    nghq = 12,
    ...
)

Arguments

formula a formula object of the function with the response on the left of a ~ operator and
the terms on the right. The response must be a survival object as returned by the
Surv function (time in first and status in second).
data a data frame in which to interpret the variables named in the formula
ratetable a rate table stratified by age, sex, year (if missing, ratedata is used)
rmap a list that maps data set names to the ratetable names.
baseline an argument to specify the baseline hazard: if it follows a piecewise constant,
baseline = "constant" is used and corresponds to the baseline in Esteve et al.
model; if the baseline follows a quadratic b-splines, baseline = "bsplines" is
used, corresponding to the baseline excess hazard in Giorgi et al model.
pophaz indicates three possibles arguments in character: classic or rescaled and cor-
rected. If pophaz = "classic" chosen, fits the model that do not require to
rescale or to correct the background mortality (i.e. the Esteve et al. model or
Giorgi et al. model); if pophaz = "rescaled" or pophaz = "corrected" chosen,
fits the models that require to rescale or to correct the background mortality.
only_ehazard a boolean argument (by default, only_ehazard=FALSE). If only_ehazard =
TRUE, pophaz = "classic" must be provided and the total value of the log-
likelihood will not account for the cumulative population hazard.
add.rmap character that indicates the name in character of the additional demographic variable from data to be used for correction of the life table, in particular when one is in the presence of an insufficiently stratified life table (see Touraine et al. model). This argument is not used if pophaz = "classic" or pophaz = "rescaled".

add.rmap.cut a list containing arguments to specify the modeling strategy for breakpoint positions, which allows a non-proportional effect of the correction term acting on the background mortality. By default list(breakpoint = FALSE), i.e. a proportional effect of the correction term acting on the background mortality is needed; in this case, all the other argument of the list are not working for the model specification;

if list(breakpoint = TRUE, cut = c(70)), the chosen cut-point(s) is (are) the numeric value(s) proposed. If list(breakpoint = TRUE, cut = NA), there is the same number of breakpoints as the number of NA, with their possible positions specified as here by probs, i.e. list(breakpoint = TRUE, cut = NA, probs = seq(0, 1, 0.25)). That corresponds to a numeric vector of probabilities with values between 0 and 1 as in quantile function. criterion is used to choose the best model, using the AIC or the BIC (the default criterion). If needed, all the fitted models are printed by the user by adding in the list print_stepwise = FALSE.

interval a vector indicating either the location of the year-scale time intervals for models with piecewise constant function, or the location of the knots for models with B-splines functions for their baseline hazard (see the appropriate specification in baseline argument). The first component of the vector is 0, and the last one corresponds to the maximum time fellow-up of the study.

ratedata a data frame of the hazards mortality in general population.

subset an expression indicating which subset of the data should be used in the modeling. All observations are included by default

na.action as in the coxph function, a missing-data filter function.

init a list of initial values for the parameters to estimate. For each elements of the list, give the name of the covariate followed by the vector of the fixed initials values

control a list of control values used to control the optimization process. In this list, eps, is a convergence criteria (by default, eps=10^-4), iter.max is the maximum number of iteration (by default, iter.max=15), and level, is the level used for the confidence intervals (by default, level=0.95).

optim a Boolean argument (by default, optim = FALSE). If optim = TRUE), the maximization algorithm uses the optim function

scale a numeric argument to specify whether the life table contains death rates per day (default scale = 365.2425) or death rates per year (scale = 1).

trace a Boolean argument, if trace = TRUE), tracing information on the progress of the optimization is produced

speedy a Boolean argument, if speedy = TRUE, optimization is done in a parallel mode

nghq number of nodes and weights for Gaussian quadrature

... other parameters used with the xhaz function
Details

Use the \texttt{Surv(time\_start, time\_stop, status)} notation for time dependent covariate with the appropriate organization of the data set (see the help page of the \texttt{Surv} function).

Only two interior knots are possible for the model with B-splines functions to fit the baseline (excess) hazard. Determination of the intervals might be user’s defined or automatically computed according to the quantile of the distribution of deaths. Use \texttt{NA} for an automatic determination (for example, \texttt{interval = c(0, NA, NA, 5)}).

Value

An object of class \texttt{xhaz.constant} or \texttt{xhaz.bsplines}, according to the type of functions chosen to fit the baseline hazard of model (see details for argument \texttt{baseline}). This object is a list containing the following components:

\begin{itemize}
  \item \texttt{coefficients} estimates found for the model
  \item \texttt{varcov} the variance-covariance matrix
  \item \texttt{loglik} for the Esteve et al. model: the log-likelihood of the null model, i.e without covariate, and the log-likelihood of the full model, i.e with all the covariates declared in the formula; for the Giorgi et al. model: the log-likelihood of the full model
  \item \texttt{cov.test} for the Esteve et al. model: the log-likelihood of the null model, i.e without covariate, and the log-likelihood of the full model, i.e with all the covariates declared in the formula; for the Giorgi et al. model: the log-likelihood of the full model
  \item \texttt{message} a character string returned by the optimizer see details in \texttt{optim} help page
  \item \texttt{convergence} an integer code as in \texttt{optim} when "L-BFGS-B" method is used.
  \item \texttt{n} the number of individuals in the dataset
  \item \texttt{n.events} the number of events in the dataset. Event are considered as death whatever the cause
  \item \texttt{level} the confidence level used
  \item \texttt{interval} the intervals used to split time for piecewise baseline excess hazard, or knots positions for Bsplines baseline
  \item \texttt{terms} the representation of the terms in the model
  \item \texttt{call} the function \texttt{call} based on model
  \item \texttt{pophaz} the assumption considered for the life table used in the excess hazard model
  \item \texttt{add.rmap} the additional variable for which the life table is not stratified
  \item \texttt{ehazardInt} the cumulative hazard of each individuals calculated from the ratetable used in the model
  \item \texttt{ehazard} the individual expected hazard values from the ratetable used to fit the model
  \item \texttt{data} the dataset used to run the model
  \item \texttt{time\_elapsed} the time to run the model
\end{itemize}
Note

time is OBLIGATORY in YEARS.

Author(s)

Juste Goungounga, Darlin Robert Mba, Nathalie Graffeo, Roch Giorgi

References


Examples

```r
library("numDeriv")
library("survexp.fr")
library("splines")
library("statmod")
data("simuData","rescaledData", "dataCancer")
# load the data sets 'simuData', 'rescaledData' and 'dataCancer'.

# Esteve et al. model: baseline excess hazard is a piecewise function
# linear and proportional effects for the covariates on
# baseline excess hazard.
levels(simuData$sex) <- c("male", "female")

fit.estv1 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "classic")

fit.estv1
```
# Touraine et al. model: baseline excess hazard is a piecewise function
# with a linear and proportional effects for the
# covariates on the baseline excess hazard.
# An additional covariate (here race) missing in the life table is
# considered by the model.

fit.corrected1 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "corrected",
add.rmap = "race")

fit.corrected1

# extension of Touraine et al model: baseline excess hazard is a piecewise
# constant function with a linear and proportional effects for the covariates
# on the baseline excess hazard.
# An additional covariate (here race) missing in the life table is
# considered by the model with a breakpoint at 75 years

fit.corrected2 <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData,
ratetable = survexp.us,
interval = c(0, NA, NA, NA, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "constant", pophaz = "corrected",
add.rmap = "race",
add.rmap.cut = list(breakpoint = TRUE, cut = 75))

fit.corrected2

# Giorgi et al model: baseline excess hazard is a quadratic Bsplines
# function with two interior knots and allow here a
# linear and proportional effects for the covariates on
# baseline excess hazard.

fitphBS <- xhaz(formula = Surv(time_year, status) ~ agec + race,
data = simuData,
ratetable = survexp.us,
interval = c(0, NA, NA, 6),
rmap = list(age = 'age', sex = 'sex', year = 'date'),
baseline = "phBS", pophaz = "corrected")
baseline = "bsplines", pophaz = "classic")
fitphBS

# Application on `dataCancer`
# Giorgi et al model: baseline excess hazard is a quadratic Bspline
# function with two interior knots and allow here a
# linear and proportional effect for the variable
# "immuno_trt" plus a non-proportional effect
# for the variable "ageCentre" on baseline excess hazard.
fittdphBS <- xhaz(formula = Surv(obs_time_year, event) ~ qbs(ageCentre) + immuno_trt,
data = dataCancer,
ratetable = survexp.fr,
interval = c(0, 0.5, 12, 15),
rmap = list(age = 'age', sex = 'sexx', year = 'year_date'),
baseline = "bsplines", pophaz = "classic")
fittdphBS

# Application on `rescaledData`
# rescaled model: baseline excess hazard is a piecewise function with a
# linear and proportional effects for the covariates on baseline excess hazard.
# A scale parameter on the expected mortality of general population is
# considered to account for the non-comparability source of bias.
rescaledData$timeyear <- rescaledData$time/12
rescaledData$agecr <- scale(rescaledData$age, TRUE, TRUE)
fit.res <- xhaz(formula = Surv(timeyear, status) ~ agecr + hormTh,
data = rescaledData,
ratetable = survexp.fr,
interval = c(0, NA, NA, NA, NA, NA, max(rescaledData$timeyear)),
rmap = list(age = 'age', sex = 'sexx', year = 'date'),
baseline = "constant", pophaz = "rescaled")
fit.res
Index

* **anova.bsplines**
  anova.bsplines, 4
* **anova.constant**
  anova.constant, 6
* **datasets**
  dataCancer, 10
  rescaledData, 27
  simuData, 28
* **duplicate**
  duplicate, 11
* **plot.bsplines**
  plot.bsplines, 14
* **plot.predxhaz**
  plot.predxhaz, 17
* **predict.bsplines**
  predict.bsplines, 19
* **predict.constant**
  predict.constant, 21
* **print.bsplines**
  print.bsplines, 23
* **print.constant**
  print.constant, 24
* **print.predxhaz**
  print.predxhaz, 25
* **qbs**
  qbs, 26
* **summary.bsplines**
  summary.bsplines, 29
* **summary.constant**
  summary.constant, 30
* **xhaz**
  xhaz, 31

AIC.bsplines, 2
AIC.constant, 3
anova.bsplines, 4
anova.constant, 6

BIC.bsplines, 8
BIC.constant, 9

dataCancer, 10
duplicate, 11
exphaz, 13
plot.bsplines, 14, 30
plot.predxhaz, 17, 23
predict.bsplines, 19
predict.constant, 21
print.bsplines, 20, 22, 23, 24, 31
print.constant, 5, 7, 20, 22, 23, 24
print.predxhaz, 25
qbs, 26
rescaledData, 27
simuData, 28
summary.bsplines, 5, 7, 29, 30
summary.constant, 24, 30, 31
Surv, 13
xhaz, 5, 7, 20, 22–24, 30, 31, 31