Package: addhazard (via r-universe)

August 22, 2024

Title Fit Additive Hazards Models for Survival Analysis

Version 1.2.0

Date 2020-07-15

Description Contains tools to fit the additive hazards model to a cohort study, a case-cohort study, or a general two-phase sampling study. It allows weight calibration to incorporate additional auxiliary information from the full cohort into the estimation procedure in a case-cohort study or from the phase I sample in a two-phase sampling study. It can be used for both Bernoulli and finite population sampling. This package provides regression parameter estimates and their model-based and robust standard errors. It also offers tools to make prediction of individual specific hazards at a specified time.

LazyData true

Depends R (>= 3.3.1)

Imports ahaz, survival, rootSolve

RoxygenNote 7.1.1

License GPL-2 | file LICENSE

BugReports `https://github.com/katehu/addhazard/issues`

Repository https://katehu.r-universe.dev

RemoteUrl https://github.com/katehu/addhazard

RemoteRef HEAD

RemoteSha e663211a74b36ff9a1cc57c0ea3103fb6a9bee11

Contents

---

ah *Fit Additive Hazards Regression Models*

Description

Fit a semiparametric additive hazard model '

$$\lambda(t|Z = z) = \lambda_0(t) + \beta' z.$$

The estimating procedures follow Lin & Ying (1994).

Usage

```
ah(formula, data, robust, weights, ties, seed = 20, ...)
```

Arguments

`formula`	a formula object for the regression model of the form response ~ predictors. The outcome is a survival object created by `Surv`.
`data`	a data frame. Input dataset.
`robust`	a logical variable. Robust standard errors are provided if robust == TRUE.
`weights`	a numeric vector. The weight of each observation.
`ties`	a string. If there are ties in the survival time, when ties = 'break' a small random number is added to the survival time to break the ties.
`seed`	an integer. Seed number used to generate random increment when breaking ties. The default number is 20.
`...`	additional arguments to be passed to the low level regression fitting functions.

Value

An object of class 'ah' representing the fit.

Note

The response variable is a survival object. The regression model can be univariate or multivariate. This function is built upon the function `ahaz` by Anders Gorst-Rasmussen.

References

Lin, D.Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. Biometrika; 81:61-71.

See Also

`predict.ah` for prediction based on fitted `ah` model, `nwtsco` for the description of nwtsco dataset

Examples

```
library(survival)
### using the first 100 rows in nwtsco to build an additive hazards model
nwts<- nwtsco[1:100,]

### fit the additive hazards model to the data
### the model-based standard errors are reported when setting robust = FALSE
fit1 <- ah(Surv(trel,relaps) ~ age + instit, data = nwts, robust = FALSE)
summary(fit1)

### fit the additive hazards model to the data with robust standard errors
fit2 <- ah(Surv(trel,relaps) ~ age + instit, data = nwts, robust = TRUE)
summary(fit2)

### when there are ties, break the ties by setting ties = 'break'
nwts_all <- nwtsco
fit3 <- ah(Surv(trel,relaps) ~ age + instit, ties = 'break', data = nwts_all, robust = TRUE)
summary(fit3)
### users could break the ties on their own by
nwts_all$trel <- nwtsco$trel + runif(dim(nwts_all)[1],0,1)*1e-10
fit3 <- ah(Surv(trel,relaps) ~ age + instit, data = nwts_all, robust = TRUE)
summary(fit3)
```

`ah.2ph` *Fit Additive Hazards Regression Models to Two-phase Sampling*

Description

The function fits a semiparametric additive hazards model

$$\lambda(t|Z=z) = \lambda_0(t) + \beta' z.$$

to two-phase sampling data. The estimating procedures follow Hu (2014).

Usage

```
ah.2ph(
  formula,
  data,
  R,
  Pi = NULL,
  weights = NULL,
  ties,
  robust = FALSE,
  calibration.variables = NULL,
  seed = 20,
  ...
)
```

Arguments

`formula`	a formula object for the regression model of the form response ~ predictors. The outcome is a survival object created by `Surv`.
`data`	a data frame. Input dataset.
`R`	a phase II membership indicator. A vector of values of 0 and 1. The subject is selected to phase II if R = 1.
`Pi`	the probability of a subject to be selected to the phase II subsample.
`weights`	weight assigned to each individual, inverse of the selection probability
`ties`	a string. If there are ties in the survival time, when ties = 'break' a small random number is added to the survival time to break the ties.
`robust`	a logical variable. Robust standard errors are provided if robust = TRUE.
`calibration.variables`	a vector of strings of some column names of the data. These are the variables available for every observation. They are used to calibrate the weight assigned to each subject
`seed`	an integer. Seed number used to generate random increment when breaking ties. The default number is 20.
`...`	additional arguments to be passed to the low level regression fitting functions.

Value

An object of class 'ah.2h' representing the fit.

Note

This function estimates both model-based and robust standard errors. It can be used to analyze case-cohort studies with subsampling among cases. It allows weight calibration with auxiliary information from the full cohort (phase I sample). By this means, more information is used and thus weight calibration potentially could further improve the precision of prediction or our estimation on the regression coefficients.

References

Jie Hu (2014) A Z-estimation System for Two-phase Sampling with Applications to Additive Hazards Models and Epidemiologic Studies. Dissertation, University of Washington.

See Also

`predict.ah.2ph` for prediction based on fitted additive hazards model with two-phase sampling and `nwtsco` for the description of nwtsco dataset.

Examples

```
library(survival)
### fit an additive hazards model to two-phase sampling data without calibration
fit1 <- ah.2ph(Surv(trel,relaps) ~ age + histol, data = nwts2ph, R = in.ph2, Pi = Pi,
 robust = FALSE,  ties = 'break')
```

```
summary(fit1)

### use weight instead of the selection probability Pi in the input
fit1 <- ah.2ph(Surv(trel,relaps) ~ age + histol, data = nwts2ph, R = in.ph2, weights = wts,
 robust = FALSE,  ties = 'break')
summary(fit1)

### fit an additve hazards model with calibration on age
fit2 <- ah.2ph(Surv(trel,relaps) ~ age + histol, data = nwts2ph, R = in.ph2,
            Pi = Pi, robust = FALSE, ties = 'break', calibration.variables = 'age')
summary(fit2)

### calibrate on age square
### note if users create a calibration variable, then
### the new variable should be added to the original data frame
nwts2ph$age2 <- nwts2ph$age^2
fit3 <- ah.2ph(Surv(trel,relaps) ~ age + histol,  data = nwts2ph,
 R = in.ph2, Pi = Pi, robust = FALSE, ties = 'break', calibratio.variables = 'age2')
summary(fit3)

#############################################################################
## When phase II samples are obtained by finite Sampling
#############################################################################

### calculating the sample size for each straum
### calculate the strata size
strt.size <- table(nwts2ph$strt)
ph2.strt.size <- table(subset(nwts2ph, in.ph2 == 1)$strt)
### fit an additve hazards model with finite stratified sampling
### calculate the sampling fractions
frac <- ph2.strt.size/strt.size
### treating the problem as bernoulli sampling coupled with calibration on strata sizes
### using frac as the sampling probilities
nwts2ph_by_FPSS <- nwts2ph
nwts2ph_by_FPSS$Pi <- NULL
for (i in 1:length(strt.size)){
  nwts2ph_by_FPSS$Pi[nwts2ph_by_FPSS$strt ==i] <- frac[i]
}

### create strt indicators, which become our calibration variables
for (i in 1:length(strt.size)){
   nwts2ph_by_FPSS$strt_ind <- as.numeric(nwts2ph_by_FPSS$strt ==i)
   names(nwts2ph_by_FPSS)[ncol(nwts2ph_by_FPSS)]= paste0('strt', i)
}
### fit an additve hazards model with finate sampling
fit4 <- ah.2ph(Surv(trel,relaps) ~ age + histol,
                                   data = nwts2ph_by_FPSS,
                                   R = in.ph2, Pi = Pi,
                                   robust = FALSE,
                                   ties = 'break',
                                   calibration.variables = c('strt1','strt2','strt3'))
summary(fit4)
```

`nwts2ph` *An hypothetical two-phase sampling dataset based on nwtsco dataset from the National Wilms Tumor Study (NWTS)*

Description

An hypothetical two-phase sampling dataset based on nwtsco dataset from the National Wilms Tumor Study (NWTS)

Usage

```
nwts2ph
```

Format

A data frame with 3915 rows and 15 variables:

> We create a hypothetical two-phase sampling (stratified sampling) dataset. In this dataset, only the subjects who have relapses and some of the controls have their samples sent to the central laboratory for more accurate histology examination.
>
> Institutional histology is examined in the local hospital. It is usually available for all the samples. Central histology is more expensive to obtain since the samples have to be sent to the central laboratory and the work requires experienced lab scientists. It is reasonable to assume only some samples were tested for central histology. Following the two-phase sampling design, we selected subjects for central histology measurements based on their outcomes and institutional histology results.

trel Time to relapse or last date seen (yr), continuous

tsur Time to death or last date seen (yr), continuous

relaps Indicator of relapse, 0 = Alive no prior relapse when last seen, 1 = Relapsed after trel years

dead Indicator of death, 0 = Alive when last seen, 1 = Died after tsur years

study NWTS study, 3 = NWTS-3, 4 = NWTS-4

stage Stage of disease, 1=I, 2=II, 3=III, 4=IV

histol Central Path histology, 0 = Favorable (FH) and the subject is selected into the phase II subsample (in.ph2 = 1), 1 = Unfavorable (UH) and the subject is selected into the phase II subsample (in.ph2 = 1), NA = subject is NOT selected into the phase II subsample (in.ph2 = 1)

instit Institutional histology, 0 = Favorable (FH), 1 = Unfavorable (UH)

age Age at diagnosis (yr), continuous

yr Year of diagnosis, calendar year

specwgt Weight of tumor bearing specimen, in grams (continuous)

tumdiam Diameter of tumor, in centimeters (continuous)

strt Strata, 1 = Unfavorable Institutional histology and no relapse, 2 = favorable Institutional histology and no relapse, 3 = relapse

Pi Sampling probability, 0.5 for strata =1, 0.9 for strata = 2, 0.9 for strata = 3

wts weight assigned to each individual, inverse of the selection probability

in.ph2 Phase II membership, 1 = selected into the phase II subsample, 0 = not selected into the phase II subsample

Source

This dataset was created based on nwtsco dataset from the National Wilms Tumor Study (NWTS)

`nwts2ph.generate` *This file genreate the example dataset nwts2ph importFrom('stats', 'rbinom')*

Description

This file genreate the example dataset nwts2ph importFrom('stats', 'rbinom')

Usage

```
nwts2ph.generate(data, seed = 20)
```

Arguments

`data`	the full cohort data
`seed`	the random seed we use for generating this dataset

`nwtsco` *Dataset from the National Wilms Tumor Study (NWTS)*

Description

Dataset from the National Wilms Tumor Study (NWTS)

Usage

```
nwtsco
```

Format

A data frame with 3915 rows and 12 variables:

trel Time to relapse orlast date seen (yr), continuous

tsur Time to death or last date seen (yr), continuous

relaps Indicator of relapse, 0 = Alive no prior relapse when last seen, 1 = Relapsed after trel years

dead Indicator of death, 0 = Alive when last seen, 1 = Died after tsur years

study NWTS study, 3 = NWTS-3, 4 = NWTS-4

stage Stage of disease, 1=I, 2=II, 3=III, 4=IV

histol Central Path histology, 0 = Favorable (FH), 1 = Unfavorable (UH)

instit Institutional histology, 0 = Favorable (FH), 1 = Unfavorable (UH)

age Age at diagnosis (yr), continuous

yr Year of diagnosis, calendar year

specwgt Weight of tumor bearing specimen, in grams (continuous)

tumdiam Diameter of tumor, in centimeters (continuous)

Source

Originally used by M. Kulich and D.Y. Lin: Improving the efficiency of relative-risk estimation in case-cohort studies. J Amer Statis Assoc 99:832-844, 2004.

`predict.ah` *Prediction Based on the Fitted Additive Hazards Model*

Description

This function predicts a subject's overall hazard rates at given time points based on this subject's covariate values. The prediction function is an additive hazards model fitted using `ah`.

Usage

```
## S3 method for class 'ah'
predict(object, newdata, newtime, ...)
```

Arguments

`object`	an object of class inhering from `ah`.
`newdata`	a dataframe of an individual's predictors.
`newtime`	a given sequence of time points at which the prediction is performed. The time should be on the same scale as the survival time in `Surv`.
`...`	further arguments passed to or from other methods.

Value

A dataframe including the time points for prediction, predicted values and their standard errors.

See Also

`ah` for fitting the additive hazards model, `nwtsco` for the description of nwtsco dataset

Examples

```
library(survival)
###  fit the additive hazards model to the data
nwts<- nwtsco[1:100,]
fit <- ah(Surv(trel,relaps) ~ age + instit, data = nwts,  ties = 'break', robust = FALSE)

###  see the covariate names in the prediction function
fit$call
###  the newdata should be a dataframe
###  the variable names are the same as the covariate names in
###  the prediction function
newdata <- data.frame(age=60, instit =1)

###  an alternative way to give the newdata
newdata <- nwtsco[101,]

###  based on this subject's covariate values, the function predicts  individual specific
###  hazard rates at time points 3 and 5
predict(fit, newdata, newtime = c(3,5))
```

`predict.ah.2ph` *Prediction Based on the Additive Hazards Model Fitted from Two-phase Sampling*

Description

This function predicts a subject's overall hazard rates at given time points based on this subject's covariate values. The prediction function is an object from `ah.2ph`. The estimating procedures follow Hu (2014).

Usage

```
## S3 method for class 'ah.2ph'
predict(object, newdata, newtime, ...)
```

Arguments

`object`	an object of class inhering from 'ah.2ph'.
`newdata`	a dataframe of an individual's predictors.
`newtime`	a given sequence of time points at which the prediction is performed.
`...`	further arguments passed to or from other methods.

Value

A dataframe including the given time points, predicted hazards, their standard errors, their variances, the phase I component of the variance for predicted hazards and the phase II component of the variance.

References

Jie Hu (2014) A Z-estimation System for Two-phase Sampling with Applications to Additive Hazards Models and Epidemiologic Studies. Dissertation, University of Washington.

See Also

`ah.2ph` for fitting the additive hazards model with two-phase

Examples

```
library(survival)
### fit an additive hazards model to  two-phase sampling data without calibration
fit1 <- ah.2ph(Surv(trel,relaps) ~ age + histol, data = nwts2ph,  R = in.ph2,
               Pi = Pi, robust = FALSE, ties = 'break')

###  input the new data for prediction
newdata <- nwtsco[101,]
###  based on the fitted model fit1, perform prediction at time points t =3 and t= 5
predict(fit1, newdata, newtime = c(3,5))

### fit an additve hazards model to  two-phase sampling data with calibration
### The calibration variable is instit
fit2 <- ah.2ph(Surv(trel,relaps) ~ age + histol, data = nwts2ph, R = in.ph2, Pi = Pi,
                         ties = 'break', robust = FALSE, calibration.variables = "instit")
### based on the fitted model fit2, perform prediction at time points t =3 and t= 5
predict(fit2, newdata, newtime = c(3,5))
```

Index