

Package ‘censCov’

April 26, 2017

Type Package

Title Linear Regression with a Randomly Censored Covariate

Version 1.0-0

Date 2017-04-25

Author Jing Qian, Sy Han (Steven) Chiou

Maintainer Sy Han (Steven) Chiou <schiou@hsph.harvard.edu>

Description Implementations of threshold regression approaches for linear regression models with a covariate subject to random censoring, including deletion threshold regression and completion threshold regression. Reverse survival regression, which flip the role of response variable and the covariate, is also considered.

License GPL (>= 3)

Imports survival

Encoding UTF-8

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2017-04-25 23:35:57 UTC

R topics documented:

thlm	2
Index	5

thlm

*Threshold regression with a censored covariate***Description**

This function fits a linear regression model when there is a censored covariate. The method involves thresholding the continuous covariate into a binary covariate. A collection of threshold regression methods are implemented to obtain the estimator of the regression coefficient as well as to test the significance of the effect of the censored covariate. When there is no censoring, the method reduces to the simple linear regression.

Usage

```
thlm(formula, data, cens = NULL, subset, method = "cc",
      B = 0, x.upplim = NULL, t0 = NULL, control = thlm.control())
```

Arguments

formula	a formula expression, of the form 'response ~ predictors'. The response variable is assumed to be fully observed. See the documentation of 'lm' or 'formula' for more details.
data	an optional data frame, list or environment contains variables in the 'formula', or in the 'subset' argument. If left unspecified, the variables are taken from 'environment(formula)', typically the environment from which 'thlm' is called.
cens	an optional vector of censoring indicator (0 = censoring, 1 = failure) for the censored covariate, which is assumed to be the first covariate specified in the 'formula'. When 'cens' is left unspecified or a vector of 1's, the model assumes all covariates are fully observed and the model reduced to simple linear regression, i.e. 'lm'.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
method	a character string specifying the threshold regression methods to be used. The following are permitted: "cc" for complete-cases regression, "rev" for reverse survival regression, "dt" for deletion threshold regression, "ct" for complete threshold regression, and "all" for all four approaches.
B	a numeric value specifies the bootstrap size for estimating the standard deviation of regression coefficient for the censored covariate when method = "dt" or "ct". When B = 0, only the beta estimate will be displayed.
x.upplim	an optional numeric value specifies the upper support of censored covariate. When left unspecified, the maximum of the censored covariate will be used.
t0	an optional numeric value specifies the threshold when method = "dt" or "ct". When left unspecified, an optimal threshold will be determined to optimize test power using the proposed procedure in Qian et al (2017).

control a list of parameters. 't0.interval' controls the end points of the interval to be searched for the optimal threshold when 't0' is left unspecified. 't0.plot' controls whether the objective function will be plotted. When 't0.plot' is true, both the raw values and the smoothed estimates (using local polynomial regression fitting) are plotted.

Details

The model assumes the linear regression model:

$$Y = a_0 + a_1X + a_2Z + e,$$

where X is the covariate of interest which is subject to right censoring, Z is a covariate matrix that are fully observed, Y is the response variable, and e is an independent random error term with mean 0 and finite variance.

The hypothesis test of association is based on the significance of the regression coefficient, a_1 . However, when deletion threshold regression or complete threshold regression is executed, an equivalent but easy-to-evaluate test is performed. Namely, given a threshold t^* , we define a derived binary covariate, X^* , such that $X^* = 1$ when $X > t^*$ and $X^* = 0$ when X is uncensored and $X < t^*$. The proposed linear regression can be expressed as

$$E(Y|X^*, Z) = b_0 + b_1X^* + b_2Z.$$

The proposed hypothesis test of association can be tested by the significance of b_1 . Under the assumption that X is independent of Z given X^* , b_2 is equivalent to a_2 .

Author(s)

Jing Qian, Sy Han Chiou

References

- Qian, J. et al (2017), Threshold regression with a censored covariate *Unpublished manuscript*.
- Atem, F., Qian, J., Maye J.E., Johnson, K.A. and Betensky, R.A. (2017) , Linear regression with a randomly censored covariate: Application to an Alzheimer's study. *Journal of the Royal Statistical Society: Series C*, 66(2):313-328.

Examples

```
simDat <- function(n) {
  X <- rexp(n, 3)
  Z <- runif(n, 1, 6)
  Y <- 0.5 + 0.5 * X - 0.5 * Z + rnorm(n, 0, .75)
  cstime <- rexp(n, .75)
  delta <- (X <= cstime) * 1
  X <- pmin(X, cstime)
  data.frame(Y = Y, X = X, Z = Z, delta = delta)
}
set.seed(0)
dat <- simDat(200)
```

```
## Falsely assumes all covariates are free of censoring
thlm(Y ~ X + Z, data = dat)

## Complete cases regression
thlm(Y ~ X + Z, cens = delta, data = dat)
thlm(Y ~ X + Z, data = dat, subset = delta == 1) ## same results

## reverse survival regression
thlm(Y ~ X + Z, cens = delta, data = dat, method = "reverse")

## threshold regression without bootstrap
thlm(Y ~ X + Z, cens = delta, data = dat, method = "dt")
thlm(Y ~ X + Z, cens = delta, data = dat, method = "ct", control =
list(t0.interval = c(0.2, 0.6), t0.plot = FALSE))

## threshold regression with bootstrap
thlm(Y ~ X + Z, cens = delta, data = dat, method = "dt", B = 100)
thlm(Y ~ X + Z, cens = delta, data = dat, method = "ct", B = 100)

## display all
thlm(Y ~ X + Z, cens = delta, data = dat, method = "all", B = 100)
```

Index

thlm, 2