Package ‘npROCRegression’

November 24, 2021

Type Package
Title Kernel-Based Nonparametric ROC Regression Modelling
Version 1.0-6
Date 2021-11-23
Imports lattice, graphics, stats
Suggests knitr, rmarkdown
Description Implements several nonparametric regression approaches for the inclusion of covariate information on the receiver operating characteristic (ROC) framework.
VignetteBuilder knitr
NeedsCompilation yes
License GPL
LazyLoad yes
Author Maria Xose Rodriguez-Alvarez [aut, cre],
Javier Roca-Pardinas [aut]
Maintainer Maria Xose Rodriguez-Alvarez <mxrodriguez@uvigo.es>
Repository CRAN
Date/Publication 2021-11-23 23:20:08 UTC

R topics documented:

npROCRegression-package ........................................... 2
controlDNPROCreg .................................................. 2
controlINPROCreg .................................................. 4
DNPROCreg .......................................................... 5
DNPROCregData ..................................................... 7
endosim ............................................................... 8
INPROCreg ............................................................ 9
plot.DNPROCreg ..................................................... 12
plot.INPROCreg ..................................................... 13
print.DNPROCreg ................................................... 14
print.INPROCreg ................................................... 15
summary.DNPROCreg ............................................... 16
summary.INPROCreg ............................................... 17
The npROCRegression package allows the user to apply in practice the nonparametric induced and direct ROC regression approaches presented in Rodriguez-Alvarez et al. (2011a) and Rodriguez-Alvarez et al. (2011b, 2016) respectively.

Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

Maintainer: Maria Xose Rodriguez - Alvarez <mxrodriguez@uvigo.es>

References


Function used to set several parameters controlling the ROC regression fitting process
Description

Function used to set several parameters controlling the ROC regression fitting process.

Usage

controlDNPROCreg(step.p = 0.02, card.P = 50, link = c("probit", "logit", "cloglog"), kbin = 30, p = 1, seed = NULL, nboot = 500, level = 0.95, resample.m = c("coutcome", "ncoutcome"))

Arguments

- **step.p**: a numeric value, defaulting to 0.02. ROC curves are calculated at a regular sequence of false positive fractions with step.p increment.
- **card.P**: an integer value specifying the cardinality of the set of false positive fractions used in the estimation process. By default 50.
- **link**: a character string specifying the link function ("probit", "logit" or "cloglog"). By default the link is the probit function.
- **kbin**: an integer value specifying the number of binning knots. By default 30.
- **p**: an integer value specifying the order of the local polynomial kernel estimator. By default 1.
- **seed**: an integer value specifying the seed for the bootstrap resamples. If NULL it is initialized randomly.
- **nboot**: an integer value specifying the number of bootstrap resamples for the construction of the confidence intervals. By default 500.
- **level**: a real value specifying the confidence level for the confidence intervals. By default 0.95.
- **resample.m**: a character string specifying if bootstrap resampling (for the confidence intervals) should be done with or without regard to the disease status ("coutcome" or "noutcome"). In both cases, a naive bootstrap is used. By default, the resampling is done conditionally on the disease status.

Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

See Also

See Also DNPROCreg

Examples

data(endosim)
# Fit a model including the interaction between age and gender.
m0 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)", formula.ROC = "~ gender + s(age) + s(age, by = gender)", group = "idf_status", tag.healthy = 0,
data = endosim,
control = controlINPROCreg(card.P=50, kbin=30, step.p=0.02))
summary(m0)
plot(m0)

controlINPROCreg

Function used to set several parameters controlling fitting process.

Description
Function used to set several parameters controlling fitting process.

Usage
controlINPROCreg(step.p = 0.02, kbin = 30, p = 1, h = c(-1, -1, -1, -1),
seed = NULL, nboot = 500, level = 0.95, resample.m = c("coutcome", "ncoutcome"))

Arguments
step.p  a numeric value, defaulting to 0.02. ROC curves are calculated at a regular sequence of false positive fractions with step.p increment.

kbin   an integer value specifying the number of binning knots. By default 30.

p    an integer value specifying the order of the local polynomial kernel estimator for the regression functions. By default 1.

h    a vector of length 4 specifying the bandwidths to be used for the estimation of the regression and variance functions in healthy population and the regression and variance functions in diseased populations (in this order). By default -1 (selected using cross-validation). A value of 0 would indicate a linear fit.

seed   an integer value specifying the seed for the bootstrap resamples. If NULL it is initialized randomly.

nboot   an integer value specifying the number of bootstrap resamples for the construction of the confidence intervals. By default 500.

level   a real value specifying the confidence level for the confidence intervals. By default 0.95.

resample.m   a character string specifying if bootstrap resampling (for the confidence intervals) should be done with or without regard to the disease status ("coutcome" or "ncoutcome"). When the resampling method is done conditionally on the disease status, the resampling is based on the residuals of the regression models in healthy and diseased populations. However, when the bootstrap resampling is done without regard to the disease status, a naive bootstrap is used. By default, the resampling is done conditionally on the disease status.

Author(s)
Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas
See Also

See Also INPROCreg

Examples

data(endosim)
# Evaluate the effect of age on the accuracy of the body mass index for males
m0.men <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endosim, gender == "Men"),
ci.fit = FALSE, test = FALSE,
accuracy = c("EQ","TH"),
accuracy.cal="AROC",
control=controlINPROCreg(p=1,kbin=30,step.p=0.01),
newdata = data.frame(age = seq(18,85,l=50)))

summary(m0.men)
plot(m0.men)

---

**DNPROCreg**

*Direct nonparametric ROC regression modelling*

**Description**

Estimates the covariate-specific ROC curve in the presence of multidimensional covariates by means of the ROC-GAM regression model presented in Rodriguez-Alvarez et al. (2011)

**Usage**

```
DNPROCreg(marker, formula.h = ~1, formula.ROC = ~1, group, tag.healthy, data,
           ci.fit = FALSE, test.partial = NULL, newdata = NULL,
           control = controlDNPROCreg(), weights = NULL)
```

**Arguments**

- **marker**: A character string with the name of the diagnostic test variable.
- **formula.h**: Right-hand formula(s) giving the mean and variance model(s) to be fitted in healthy population. Atomic values are also valid, being recycled.
- **formula.ROC**: Right-hand formula giving the ROC regression model to be fitted (ROC-GAM model).
- **group**: A character string with the name of the variable that distinguishes healthy from diseased individuals.
- **tag.healthy**: The value codifying the healthy individuals in the variable group.
- **data**: Data frame representing the data and containing all needed variables.
- **ci.fit**: A logical value. If TRUE, confidence intervals are computed.
test.partial  A numeric vector containing the position of the covariate components in the ROC-GAM formula to be tested for a possible effect. If NULL, no test is performed. If NULL, no test is performed.

newdata  A data frame containing the values of the covariate at which predictions are required.

control  Output of the controlDNPROCreg() function.

weights  An optional vector of ‘prior weights’ to be used in the fitting process.

Value
As a result, the function DNPROCreg() provides a list with the following components:

call  The matched call.

model  Data frame containing all variables and observations used in the fitting process.

fpf  Set of false positive fractions (FPF) at which the covariate-specific ROC curve has been estimated.

newdata  Data frame containing the values of the covariates at which estimates has been obtained.

pfunctions  Matrices containing the estimates of each component of the additive predictor of the ROC-GAM. One matrix contains the effects of the covariates, the other the effect of the FPF. Confidence intervals are returned if required.

coefficients  Vector of parametric coefficient of the fitted ROC-GAM.

ROC  Estimated covariate-specific ROC curve.

AUC  Estimated covariate-specific AUC, and corresponding confidence intervals if required.

pvalue  If required, p-values are obtained - with two different bootstrap-based tests - for each model component indicated in argument test.partial (T2: $L_2$-based test; and T1: $L_1$-based test). See Rodriguez-Alvarez et al. (2016).

Author(s)
Maria Xose Rodriguez-Alvarez and Javier Roca-Pardinas

References


See Also
See Also as INPROCreg, summary.DNPROCreg, plot.DNPROCreg, controlDNPROCreg, DNPROCregData.
Examples

data(endosim)
# Fit a model including the interaction between age and gender.
m0 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
            formula.ROC = "~ gender + s(age) + s(age, by = gender)",
            group = "idf_status",
            tag.healthy = 0,
            data = endosim,
            control = list(card.P=50, kbin=30, step.p=0.02))
summary(m0)
plot(m0)

## Not run:
# For confidence intervals
set.seed(123)
m1 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
            formula.ROC = "~ gender + s(age) + s(age, by = gender)",
            group = "idf_status",
            tag.healthy = 0,
            data = endosim,
            control = list(card.P=50, kbin=30, step.p=0.02),
            ci.fit = TRUE)
summary(m1)
plot(m1)

# For testing the presence of interaction between age and gender
set.seed(123)
m2 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
            formula.ROC = "~ gender + s(age) + s(age, by = gender)",
            group = "idf_status",
            tag.healthy = 0,
            data = endosim,
            control = list(card.P=50, kbin=30, step.p=0.02),
            test.partial = 3)
summary(m2)
plot(m2)

## End(Not run)

---

DNPROCregData

Selects an adequate set of points from a data set for obtaining predictions or plots.

Description

Selects an adequate set of points from a data set to be used as a default dataset for obtaining predictions or plots.

Usage

DNPROCregData(data, names.cov, group)
Arguments

data            Data set from which the new set of covariate values is obtained.
names.cov       Character vector with the names of the covariates to be included in the new data set.
group           A character string with the name of the variable in the original data set that distinguishes healthy from diseased individuals.

Value

a data frame containing selected values of all needed covariates. For those that are continuous, 30 different values are selected.

Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

See Also

See Also DNPROCreg.

Examples

data(endosim)
  # Fit a model including the interaction between age and gender.
  m0 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
                  formula.ROC = "~ gender + s(age) + s(age, by = gender)",
                  group = "idf_status",
                  tag.healthy = 0,
                  data = endosim,
                  control = list(card.P=50, kbin=30, step.p=0.02),
                  ci.fit = FALSE,
                  test.partial = NULL,
                  newdata = NULL)
  summary(m0)
  plot(m0)

Description

The endosim data set was simulated based on the data analyzed in Rodriguez-Alvarez et al. (2011a,b) and presented in Botana et al. (2007) and Tome et al. (2008). The aim of these studies was to use the Body Mass Index (BMI) to detect patients having a higher risk of cardiovascular problems, ascertaining the possible effect of age and gender on the accuracy of this measure.

Usage

data(endosim)
INPROCreg

Format

A data frame with 2840 observations on the following 4 variables.

gender  patient’s gender. Factor with Male and Female levels.
age  patient’s age.
idf_status true disease status (presence/absence of two of more cardiovascular risk factors according to the International Diabetes Federation). Numerical vector (0=absence, 1=presence).
bmi  patient’s body mass index.

Source


References


Examples

data(endosim)
summary(endosim)

INPROCreg

Induced nonparametric ROC regression modelling

Description

Estimates the covariate-specific ROC curve (and related measures) in the presence of a one-dimensional continuous covariate based on the induced nonparametric ROC regression approach as presented in Rodriguez-Alvarez et al. (2011).

Usage

INPROCreg(marker, covariate, group, tag.healthy, data, ci.fit = FALSE, test = FALSE, accuracy = NULL, accuracy.cal = c(“ROC”, “AROC”), newdata = NULL, control = controlINPROCreg(), weights = NULL)
Arguments

- **marker**: A character string with the name of the diagnostic test variable.
- **covariate**: A character string with the name of the continuous covariate.
- **group**: A character string with the name of the variable that distinguishes healthy from diseased individuals.
- **tag.healthy**: The value codifying the healthy individuals in the variable group.
- **data**: Data frame representing the data and containing all needed variables.
- **ci.fit**: A logical value. If TRUE, confidence intervals are computed.
- **test**: A logical value. If TRUE, the bootstrap-based test for detecting covariate effect is performed.
- **accuracy**: A character vector indicating if the Youden index (“YI”), the value for which the TPF and the TNF coincides (“EQ”), and/or optimal threshold (“TH”) based on these two criteria should be computed.
- **accuracy.cal**: A character string indicating if the accuracy measures should be calculated based on the covariate-specific ROC curve or on the covariate-adjusted ROC curve (AROC).
- **newdata**: A data frame containing the values of the covariate at which predictions are required.
- **control**: Output of the controlINROCreg() function.
- **weights**: An optional vector of 'prior weights' to be used in the fitting process.

Value

As a result, the function INPROCreg() provides a list with the following components:

- **call**: The matched call.
- **X**: The data frame used in the predictions.
- **fpf**: Set of false positive fractions at which the covariate-specific ROC curve has been estimated.
- **h**: Estimated regression and variance functions in healthy population.
- **d**: Estimated regression and variance functions in diseased population.
- **ROC**: Estimated covariate-specific ROC curve.
- **AUC**: Estimated covariate-specific AUC, and corresponding confidence intervals if required.
- **AROC**: Estimated covariate-adjusted ROC curve.
- **YI/EQ**: If required, estimated covariate-specific YI (or values at which the TPF and the TNF coincide), and corresponding bootstrap confidence intervals.
- **TH**: If required, estimated optimal threshold values based on either the YI or the criterion of equality of TPF and TNF, and corresponding bootstrap confidence intervals.
- **pvalue**: If required, p-value obtained with the test for checking the effect of the continuous covariate on the ROC curve.
Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

References


See Also

See Also as DNPROCreg, summary.INPROCreg, plot.INPROCreg, controlINPROCreg.

Examples

data(endosim)
# Evaluate the effect of age on the accuracy of the body mass index for males
m0.men <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endosim, gender == "Men"),
ci.fit = FALSE, test = FALSE,
accuracy = c("EQ","TH"),
accuracy.cal="AROC",
control=controlINPROCreg(p=1,kbin=30,step.p=0.01),
newdata = data.frame(age = seq(18,85,l=50)))

summary(m0.men)
plot(m0.men)
# Evaluate the effect of age on the accuracy of the body mass index for females
m0.women <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endosim, gender == "Women"),
ci.fit = FALSE, test = FALSE,
accuracy = c("EQ","TH"),
accuracy.cal="ROC",
control=controlINPROCreg(p=1,kbin=30,step.p=0.01),
newdata = data.frame(age = seq(18,85,l=50)))

summary(m0.women)
plot(m0.women)
## Not run:
# For computing confidence intervals and testing covariate effect
set.seed(123)
m1.men <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endosim, gender == "Men"),
ci.fit = TRUE, test = TRUE,
plot.DNPROCreg

Default DNPROCreg plotting

Description
Takes a fitted DNPROCreg object produced by DNPROCreg() and plots the covariate-specific ROC curve and associated AUC.

Usage
## S3 method for class 'DNPROCreg'
plot(x, ask = TRUE, ...)

Arguments

x
an object of class DNPROCreg as produced by DNPROCreg()

ask
a logical value. If TRUE, the default, the user is asked for confirmation, before a new figure is drawn

... further arguments passed to or from other methods.

Author(s)
Maria Xose Rodriguez-Alvarez and Javier Roca-Pardinas

See Also
See Also DNPROCreg, summary.DNPROCreg.

Examples

data(endosim)
# Fit a model including the interaction between age and gender.
m0 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
formula.ROC = "~ gender + s(age) + s(age, by = gender)",
group = "idf_status",
tag.healthy = 0,
data = endosim,
control = list(card.P=50, kbin=30, step.p=0.02))
summary(m0)
plot(m0)
plot.INPROCreg

Description

Default INPROCreg plotting (see details)

Usage

```r
## S3 method for class 'INPROCreg'
plot(x, ask = TRUE, ...)
```

Arguments

- `x`: an object of class INPROCreg as produced by INPROCreg().
- `ask`: a logical value. If TRUE, the default, the user is asked for confirmation, before a new figure is drawn.
- `...`: further arguments passed to or from other methods.

Details

The function produces the following plots:

- **(a)** the estimated regression and variance functions in both the healthy and diseased populations.
- **(b)** the covariate-specific ROC curve and AUC.
- **(c)** the covariate-adjusted ROC curve (AROC).
- **(d)** (optionally) the Youden Index (YI) or the value for which the TPF and the TNF coincides (EQ).
- **(e)** (optionally) the optimal thresholds based on these criteria (TH).

Author(s)

Maria Xose Rodriguez-Alvarez and Javier Roca-Pardinas

See Also

See Also INPROCreg.

Examples

```r
data(endsim)
data(endsim)
# Evaluate the effect of age on the accuracy of the body mass index for males
m0.men <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endsim, gender == "Men"),
clip.fit = FALSE, test = FALSE,
accuracy = c("EQ","TH"),
accuracy.cal="AROC",
```
control=controlDNPROCreg(p=1,kbin=30,step.p=0.01),
newdata = data.frame(age = seq(18,85,l=50))
summary(m0.men)
plot(m0.men)

print.DNPROCreg

Print method for DNPROCreg objects

Description
Print method for DNPROCreg objects

Usage
## S3 method for class 'DNPROCreg'
print(x, ...)

Arguments
x an object of class DNPROCreg as produced by DNPROCreg()
... further arguments passed to or from other methods. Not yet implemented

Author(s)
Maria Xose Rodriguez-Alvarez and Javier Roca-Pardinas

See Also
See Also as DNPROCreg.

Examples
data(endosim)
# Fit a model including the interaction between age and gender.
m0 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
formula.ROC = "~ gender + s(age) + s(age, by = gender)",
group = "idf_status",
tag.healthy = 0,
data = endosim,
control = list(card.P=50, kbin=30, step.p=0.02),
ci.fit = FALSE,
test.partial = NULL,
newdata = NULL)
m0
summary(m0)
plot(m0)
Description

Print method for INPROCreg objects

Usage

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

Arguments

x an object of class INPROCreg as produced by INPROCreg()

... further arguments passed to or from other methods. Not yet implemented

Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

See Also

See Also as INPROCreg.

Examples

data(endosim)
# Evaluate the effect of age on the accuracy of the body mass index for males
m0.men <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endosim, gender == "Men"),
ci.fit = FALSE, test = FALSE,
accuracy = c("EQ","TH"),
accuracy.cal="AROC",
control=controlINPROCreg(p=1,kbin=30,step.p=0.01),
newdata = data.frame(age = seq(18,85,l=50)))
m0.men
summary(m0.men)
plot(m0.men)
Summary method for DNPROCreg objects.

### Usage

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

### Arguments

- `object`: an object of class `DNPROCreg` as produced by `DNPROCreg()`.
- `...`: further arguments passed to or from other methods. Not yet implemented.

### Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

### See Also

See Also `DNPROCreg`, `plot.DNPROCreg`.

### Examples

```r
data(endosim)
# Fit a model including the interaction between age and gender.
m0 <- DNPROCreg(marker = "bmi", formula.h = "~ gender + s(age) + s(age, by = gender)",
formula.ROC = "~ gender + s(age) + s(age, by = gender)",
group = "idf_status",
tag.healthy = 0,
data = endosim,
control = list(card.P=50, kbin=30, step.p=0.02),
ci.fit = FALSE,
test.partial = NULL,
newdata = NULL)
summary(m0)
plot(m0)
```
Summary method for INPROCreg objects.

Usage

## S3 method for class 'INPROCreg'
summary(object, ...)

Arguments

object an object of class INPROCreg as produced by INPROCreg().
...
    further arguments passed to or from other methods. Not yet implemented.

Author(s)

Maria Xose Rodriguez - Alvarez and Javier Roca-Pardinas

See Also

See Also INPROCreg.

Examples

data(endosim)
# Evaluate the effect of age on the accuracy of the body mass index for males
m0.men <- INPROCreg(marker = "bmi", covariate = "age", group = "idf_status",
tag.healthy = 0,
data = subset(endosim, gender == "Men"),
ci.fit = FALSE, test = FALSE,
accuracy = c("EQ","TH"),
accuracy.cal="AROC",
control=controlINPROCreg(p=1,kbin=30,step.p=0.01),
newdata = data.frame(age = seq(18,85,l=50)))
summary(m0.men)
plot(m0.men)
Index

* datasets
  endosim, 8
* package
  npROCRegression-package, 2

  controlDNPROCreg, 2, 6
  controlINPROCreg, 4, 11

  DNPROCreg, 3, 5, 8, 11, 12, 14, 16
  DNPROCregData, 6, 7

  endosim, 8

  INPROCreg, 5, 6, 9, 13, 15, 17

  npROCRegression
    (npROCRegression-package), 2
    npROCRegression-package, 2

  plot.DNPROCreg, 6, 12, 16
  plot.INPROCreg, 11, 13
  print.DNPROCreg, 14
  print.INPROCreg, 15

  summary.DNPROCreg, 6, 12, 16
  summary.INPROCreg, 11, 17