Package 'FMCCSD'

April 14, 2020

Type Package

Title Efficient Estimation of Clustered Current Status Data

Version 1.0

Date 2020-04-05

Author Tong Wang [aut, cre],
Kejun He [aut],
Wei Ma [aut],
Dipankar Bandyopadhyay [aut],
Samiran Sinha [aut]

Maintainer Tong Wang <tong@stat.tamu.edu>

Description Current status data abounds
in the field of epidemiology and public health, where the only observable data for a subject is
the random inspection time and the event status at inspection. Motivated by such a current status
data from a periodontal study where data are inherently clustered,
we propose a unified methodology to analyze such complex data.

License GPL-2

Encoding UTF-8

Imports Rcpp (>= 0.12.18), numDeriv, splines2, orthopolynom

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 6.1.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-04-14 16:30:05 UTC

R topics documented:

FMCCSD-package   *Efficient Estimation of Clustered Current Status Data*

Description

Current status data abounds in the field of epidemiology and public health, where the only observable data for a subject is the random inspection time and the event status at inspection. Motivated by such a current status data from a periodontal study where data are inherently clustered, we propose a unified methodology to analyze such complex data.

Author(s)

Tong Wang [aut, cre], Kejun He [aut], Wei Ma [aut], Dipankar Bandyopadhyay [aut], Samiran Sinha [aut]

Maintainer:Tong Wang <tong@stat.tamu.edu>

CSDfit   *Analysing clustered current status data*

Description

CSDfit is used to analyze clustered current status data. The function provides parameter estimates, the maximum log likelihood value and the corresponding AIC value.

Usage

```
CSDfit(Rawdata, n_subject.raw, n_within.raw, r, n_quad = 30,
  lambda = 0, tolerance = 0.5, knots.num = 2,
  degree = 2, scale.numr = TRUE, clustering=TRUE)
```

Arguments

Rawdata   This is a dataframe of the current status data. The first column should be the index of the subject (cluster). The second column is the inspection time. The next n_subjec.raw columns are the subject (cluster-specifie) level covariates. Then the next n_within.raw columns are the within subject covariates. The last column is the indicator of the event where 1 or 0 indicate if the event has or has not happened by the inspection time, respectively. All the covariates are assumed to be either numerical or binary, and our program automatically detects if a covariate is a binary or numerical variable.

n_subject.raw   The number of subject (cluster-specifie) level covariates.

n_within.raw   The number of within cluster covariates.

r	The index of the Generalized odds ratio (GOR) model. This index is a non-negative number and it must be specified by the user. Here r=0 and 1 imply the proportional hazard and the proportional odds model, respectively.
n_quad	The number of Gauss-Hermite quadrature nodes used in numerical integration. The default value is 30.
lambda	The tuning parameter of the roughness penalty used for estimating the non-parametric component of the GOR model. The default value is 0. One must use the roughness penalty when the number of basis functions in the non-parametric component of the GOR model is large.
tolerance	This denotes the summation of the absolute values of the relative tolerance of all parameters in the model. It is used to define the convergence of the parameter estimates. The default value is 0.5.
knots.num	The number of equidistant interior knots for the integrated B-spline approximation of the nonparametric component of the GOR model. The default value is 2.
degree	The degree of integrated B-splines. The default value is 2.
scale.numr	logical. If TRUE, then all numeric covariates (cluster specifie and within cluster) are scaled with mean zero and standard deviation one. The default value is TRUE.
clustering	logical. TRUE and FALSE indicate assume there is clustering effect or not, respectively. The default value is TRUE.

Value

Function CSDfit returns a list containing the following components:

parameter.est	It is a matrix. The column "par.est" contains the estimate of the regression parameters. The column "SE" contains the standard error of these estimators. The columns "Z" and "p-value" are the corresponding Z-statistic and p-value. The last two columns are the lower and upper bound of the 95% Wald's confidence interval for the parameters.
log_likelihood	The maximum log likelihood value. This includes the logarithm of the roughness penalty and the Cauchy penalty if Cauchy.pen=TRUE.
AICvalue	The AIC value.
coefs	The estimates of the coefficients of the integrated B-spline basis functions.

Examples

```
data(PD)
PDfit=CSDfit(PD,3,1,0,n_quad=5)
```

PD   *A clustered current status dataset.*

Description

A clustered current status dataset arises from a periodontal disease (PD) study where tooth level data are clusterd within subjects. The first and second columns are the index for patients and the inspection time for each tooth, respectively. The next 3 column are the three subject level covariates (gender, smoking and Hba1c). After that, it is the tooth level covariate (jaw). The last column is the indicator for the event.

Usage

```
data(PD)
```

See Also

CSDfit

Index