Package: lodr (via r-universe)

October 6, 2024

Type Package

Title Linear Model Fitting with LOD Covariates

Version 1.0

Date 2020-04-08

Author Kevin Donovan

Maintainer Kevin Donovan <kmdono02@ad.unc.edu>

Description Tools to fit linear regression model to data while taking into account covariates with lower limit of detection (LOD).

License MIT + file LICENSE

Imports Rcpp (>= 1.0.2), Rdpack

RdMacros Rdpack

LinkingTo Rcpp, RcppArmadillo

LazyData true

RoxygenNote 6.1.1

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-04-10 17:10:02 UTC

Contents

coef.lod_lm *Extract lod_lm Coefficients*

Description

Extracts estimates regression coefficients from object of class "lod_lm".

Usage

```
## S3 method for class 'lod_lm'
coef(object, ...)
```

Arguments

object	An object of class "lod_lm", usually, a result of a call to lod_lm
...	further arguments passed to or from other methods.

Value

Coefficients extracted from object as a named numeric vector.

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

fitted.lod_lm and residuals.lod_lm for related methods; lod_lm for model fitting.

The generic functions fitted and residuals.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

## nSamples set to 100 for computational speed/illustration purposes only.
## At least 250 is recommended.  Same for boots=0; results in NAs returned for standard errors
fit <- lod_lm(data=lod_data_ex, frmla=y~x1+x2+x3, lod=c(0,0),
                  var_LOD=c("x2", "x3"), nSamples=100, boots=0)
coef(fit)
```

`fitted.lod_lm` *Extract lod_lm residuals*

Description

Extracts fitted values from object of class "`lod_lm`".

Usage

```
## S3 method for class 'lod_lm'
fitted(object, ...)
```

Arguments

`object`	An object of class "`lod_lm`", usually, a result of a call to `lod_lm`
`...`	further arguments passed to or from other methods.

Details

For subjects with covariates outside of limits of detection, when computing fitted values the values for these covariates are set according to method specified by argument `fill_in_method` in call to `lod_lm`.

Value

Fitted values extracted from `object` as a named numeric vector.

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

`coef.lod_lm` and `residuals.lod_lm` for related methods; `lod_lm` for model fitting.

The generic functions `coef` and `residuals`.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

## nSamples set to 100 for computational speed/illustration purposes only.
## At least 250 is recommended. Same for boots=0; results in NAs returned for standard errors

fit <- lod_lm(data=lod_data_ex, frmla=y~x1+x2+x3, lod=c(0,0),
                  var_LOD=c("x2", "x3"), nSamples=100, boots=0)
 fitted(fit)
```

`LOD_bootstrap_fit` *Rcpp Code for Computing Standard Errors When Fitting Linear Models with Covariates Subject to a Limit of Detection (LOD)*

Description

`LOD_bootstrap_fit` calls Rcpp code to compute linear model regression parameter standard errors in C++, taking into account covariates with limits of detection per the method detailed in May et al. (2011).

Usage

```
LOD_bootstrap_fit(num_of_boots, y_data, x_data, no_of_samples, threshold,
max_iterations, LOD_u_l)
```

Arguments

`num_of_boots`	number denoting the number of bootstrap resamples to use to compute the regression parameter standard errors.
`y_data`	numeric vector consisting of data of the model's outcome variable.
`x_data`	column-named matrix consisting of data of the model's covariates with each column representing one covariate, with values outside of the limit(s) of detection marked as `NA`. A columns of ones must be included if the model has an intercept term. Note that for valid inference, order of the covariates/columns in the matrix must be as follows from left to right: those with no LOD followed by those with an LOD.
`no_of_samples`	an integer specifying the number of samples to generate for each subject with covariate values outside of their limits of detection. For more details, see May et al. (2011).
`threshold`	number denoting the minimum difference in the regression parameter estimates needed for convergence of the model fitting procedure.

`max_iterations` number denoting the maximum number of iterations allowed in the model fitting procedure.

`LOD_u_l` numeric matrix consisting of the lower and upper limits of detection for all covariates in the model as the columns, with each covariate containing its own row, in the same order as the covariates in `x_data`. If no limit of detection exists, the corresponding matrix entry is marked with an `NA`. An entry for the intercept (`NA` in each column) must be included if applicable.

Details

This function is used to complete the standard error computations done when fitting a linear model by calling lod_lm; the standard error computations are done in C++ to minimize computation time.

Value

`LOD_bootstrap_fit` returns a list which each component being a numeric vector consisting of the last iteration's regression parameter estimates when fitting the model on a bootstrap resample of the input data.

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

`lod_lm` is the recommended function for fitting a linear model with covariates subject to limits of detection, which uses `LOD_fit`. `LOD_fit` is used to compute the regression parameter estimates.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

# Replace values marked as under limit of detection using 0 with NA,
# add column of ones for intercept
lod_data_with_int <-
  as.matrix(cbind("Intercept"=rep(1, dim(lod_data_ex)[1]), lod_data_ex))

lod_data_ex_edit <-
  apply(lod_data_with_int, MARGIN = 2, FUN=function(x){ifelse(x==0, NA, x)})

# Fit model with bootstrap procedure, report regression parameter estimate standard errors
```

```
LOD_matrix <- cbind(c(NA, NA, -100, -100), c(NA, NA, 0, 0))

## no_of_samples set to 50 for computational speed/illustration purposes only.
## At least 250 is recommended.
## Same for num_of_boots=5; at least 25 is recommended

bootstrap_fit_object <-
  LOD_bootstrap_fit(num_of_boots=5, y_data=lod_data_ex_edit[,2],
                    x_data=lod_data_ex_edit[,-2],
                    no_of_samples=50,
                    threshold=0.001, max_iterations=100, LOD_u_l=LOD_matrix)

boot_SEs <- apply(do.call("rbind", bootstrap_fit_object), 2, sd)
names(boot_SEs) <- names(lod_data_with_int[,-2])

boot_SEs
```

`lod_data_ex` *Simulated data with covariates subject to limits of detection*

Description

A simulated dataset containing a generic outcome varible and three covariates, two of which are subject to a lower limit of detection of 0, with a sample size of 100. See Details for information on how these data were generated.

Usage

```
lod_data_ex
```

Format

A data frame with 100 rows and 4 variables:

y Outcome

x1 First covariate , no limits of detection

x2 Second covariate, lower limit of detection of 0

x3 Third covariate, lower limit of detection of 0

Details

Each of the covariates were generated independently from 100 independent draws from the standard normal distributon. The outcome variable was generated from a linear model with these three covariates, along with an intercept of 1, a residual variance of 1, and regression coefficients of 1 for each covariates. Then for two of the covariates, to reflect a lower limit of detection of 0, values below this limit were set to 0. This results in a 50 percent probability of being below the limit of detection for each of the two corresponding covariates.

LOD_fit *Rcpp Code for Fitting Linear Models with Covariates Subject to a Limit of Detection (LOD)*

Description

LOD_fit calls Rcpp code to compute linear model regression parameter estimates in C++, taking into account covariates with limits of detection per the method detailed in May et al. (2011).

Usage

```
LOD_fit(y_data, x_data, mean_x_preds, beta, sigma_2_y, sigma_x_preds, no_of_samples,
threshold, max_iterations, LOD_u_l)
```

Arguments

y_data	numeric vector consisting of data of the model's outcome variable
x_data	column-named matrix consisting of data of the model's covariates with each column representing one covariate, with values outside of the limit(s) of detection marked as NA. A columns of ones must be included if the model has an intercept term.
mean_x_preds	numeric vector consisting of initial estimates of the means for each covariate, in the same order as the covariates in x_data
beta	numeric vector consisting of initial estimates of the regression parameters for each covariate, in the same order as the covariates in x_data
sigma_2_y	an initial estimate of the variance of the outcome variable
sigma_x_preds	numeric matrix consisting of an initial estimate of the covariance matrix for the model's covariates, in the same order as the covariates in x_data
no_of_samples	an integer specifying the number of samples to generate for each subject with covariate values outside of their limits of detection. For more details, see May et al. (2011).
threshold	number denoting the minimum difference in the regression parameter estimates needed for convergence of the model fitting procedure.
max_iterations	number denoting the maximum number of iterations allowed in the model fitting procedure.
LOD_u_l	numeric matrix consisting of the lower and upper limits of detection for all covariates in the model as the columns, with each covariate containing its own row, in the same order as the covariates in x_data. If no limit of detection exists, the corresponding matrix entry is marked with an NA. An entry for the intercept (NA in each column) must be included if applicable.

Details

This function is used to complete the model fitting computations done when calling lod_lm; the fitting computations are done in C++ to minimize computation time.

Value

LOD_fit returns a list containing the following components:

y_expand_last_int
: a numeric vector consisting of the outcome data with duplicate entries for subjects with covariates outside of their limits of detection per the corresponding resampling procedure, from the last iteration of the model fitting procedure.

x_data_return_last_int
: a numeric matrix consisting of the covariate data with sampled values for covariates of subjects with covariates outside of their limits of detection, from the last iteration of the model fitting procedure.

beta_estimates
: a numeric matrix consisting of the regression parameter estimates from each iteration of the model fitting procedure.

beta_estimate_last_iteration
: a numeric vector consisting of the regression parameter estimates from the last iteration of the model fitting procedure.

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

lod_lm is the recommended function for fitting a linear model with covariates subject to limits of detection, which uses LOD_fit. LOD_bootstrap_fit is used to compute regression parameter estimate standard errors using bootstrap resampling.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

# Replace values marked as under limit of detection using 0 with NA,
# add column of ones for intercept
lod_data_with_int <-
  as.matrix(cbind("Intercept"=rep(1, dim(lod_data_ex)[1]), lod_data_ex))

lod_data_ex_edit <-
  data.frame(apply(lod_data_with_int, MARGIN = 2, FUN=function(x){ifelse(x==0, NA, x)}))

# Fit linear model to dataset with only subjects without covariates under
```

```
# limit of detection to get initial estimate for the regression parameters.
beta_inital_est <- coef(lm(y~x1+x2+x3, data=lod_data_ex_edit))

# Get initial estimates of mean vector and covariance matrix for covariates and variance of outcome,
# again using data from subjects without covariates under limit of detection

mean_x_inital <- colMeans(lod_data_ex_edit[,c(-1,-2)], na.rm = TRUE)
sigma_x_inital <- cov(lod_data_ex_edit[,c(-1,-2)], use="pairwise.complete.obs")
sigma_2_y_inital <- sigma(lm(y~x1+x2+x3, data=lod_data_ex_edit))^2

# Fit model, report regression parameter estimates from last iteration
LOD_matrix <- cbind(c(NA, NA, -100, -100), c(NA, NA, 0, 0))

## no_of_samples set to 100 for computational speed/illustration purposes only.
## At least 250 is recommended.

fit_object <-
LOD_fit(y_data=lod_data_ex_edit[,2],
        x_data=as.matrix(lod_data_ex_edit[,-2]),
        mean_x_preds=mean_x_inital, beta=beta_inital_est, sigma_2_y=sigma_2_y_inital,
        sigma_x_preds=sigma_x_inital, no_of_samples=100,
        threshold=0.001, max_iterations=100, LOD_u_l=LOD_matrix)

fit_object$beta_estimate_last_iteration
```

`lod_lm` *Fitting Linear Models with Covariates Subject to a Limit of Detection (LOD)*

Description

`lod_lm` is used to fit linear models while taking into account limits of detection for corresponding covariates. It carries out the method detailed in May et al. (2011) with regression coefficient standard errors calculated using bootstrap resampling.

Usage

```
lod_lm(data, frmla, lod=NULL, var_LOD=NULL, nSamples = 250,
fill_in_method="mean", convergenceCriterion = 0.001, boots = 25)

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

Arguments

`data`	a required data frame (or object coercible by as.data.frame to a data frame) containing the variables in the model. If not specified, a corresponding error is returned.
`x`	An object of class "`lod_lm`", usually, a result of a call to `lod_lm`

frmla	an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. The details of model specification are given under 'Details'.
lod	a numeric vector (or object coercible by as.numeric) specifying the limit of detection for each covariates specified in var_LOD (in the same order as the covariates in var_LOD). Default is NULL, representing no covariates having limits of detection, which calls lm.
var_LOD	a character vector specifying which covariates in the model (frmla) are subject to limits of detection. Default is NULL, representing no covariates having limits of detection, which calls lm.
nSamples	an integer specifying the number of samples to generate for each subject with covariate values outside of their limits of detection. For more details, see May et al. (2011). The default is 250.
fill_in_method	a string specifying how values outside of the limits of detection should be handled when calculating residuals and fitted values. Default is "mean", which uses the mean covariate value. Another choice is "LOD" which uses the lower limit of detection.
convergenceCriterion	a number specifying the smallest difference between iterations required for the regression coefficient estimation process to complete. The default is 0.001.
boots	a number specifying the number of bootstrap resamples used for the standard error estimation process for the regression coefficient estimates. The default is 25.
...	further arguments passed to or from other methods.

Details

Models for lod_lm are specified the same as models for lm. A typical model as the form response ~ terms where response is the (numeric) response vector and terms is a series of terms separated by + which specifies a linear predictor for response. A formula has an implied intercept term.

In the dataset used with lod_lm, values outside of the limits of detection need to be denoted by the value of the lower limit of detection. Observations with values marked as missing by NA are removed by the model fit procedure as done with lm.

Value

lod_lm returns an object of class) "lod_lm" if arguments lod and var_LOD are not NULL, otherwise it returns class) "lm". The function summary prints a summary of the results in the same format as with an object of class) "lm". The generic accessor functions coef, fitted and residuals extract various useful features of the value returned by lod_lm.

An object of class) "lod_lm" is a list containing the following components:

coefficients	a named vector of regression coefficient estimates.
boot_SE	a named vector of regression coefficient estimate bootstrap standard error estimates.

`fitted.values`	the fitted mean values for subjects with covariates within their limits of detection.
`rank`	the numeric rank of the fitted linear model
`residuals`	the residuals, that is response minus fitted values, for subjects with covariates within their limits of detection.
`df.residual`	the residual degrees of freedom.
`model`	the model frame used.
`call`	the matched call.
`terms`	the `terms` object used.

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

`summary.lod_lm` for summaries of the results from `lod_lm`

The generic functions `coef`, `fitted` and `residuals`.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

## nSamples set to 100 for computational speed/illustration purposes only.
## At least 250 is recommended.  Same for boots=0; results in NAs returned for standard errors

fit <- lod_lm(data=lod_data_ex, frmla=y~x1+x2+x3, lod=c(0,0),
                  var_LOD=c("x2", "x3"), nSamples=100, boots=0)
 summary(fit)
```

`residuals.lod_lm` *Extract lod_lm residuals*

Description

Extracts residuals from object of class "`lod_lm`".

Usage

```
## S3 method for class 'lod_lm'
residuals(object, ...)
```

Arguments

`object`	An object of class "`lod_lm`", usually, a result of a call to `lod_lm`
`...`	further arguments passed to or from other methods.

Details

For subjects with covariates outside of limits of detection, when computing residuals the values for these covariates are set according to method specified by argument `fill_in_method` in call to `lod_lm`.

Value

Residuals extracted from `object` as a named numeric vector.

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

`fitted.lod_lm` and `coef.lod_lm` for related methods; `lod_lm` for model fitting.

The generic functions `coef` and `fitted`.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

## nSamples set to 100 for computational speed/illustration purposes only.
## At least 250 is recommended.  Same for boots=0; results in NAs returned for standard errors

fit <- lod_lm(data=lod_data_ex, frmla=y~x1+x2+x3, lod=c(0,0),
                  var_LOD=c("x2", "x3"), nSamples=100, boots=0)
 residuals(fit)
```

`summary.lod_lm` *Summarizing Linear Model Fits with Covariates Subject to a Limit of Detection*

Description

summary method for class "lod_lm"

Usage

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

## S3 method for class 'summary.lod_lm'
print(x, ...)
```

Arguments

object	An object of class "lod_lm", usually, a result of a call to lod_lm
x	An object of class "summary.lod_lm", usually, a result of a call to summary.lod_lm
...	further arguments passed to or from other methods.

Details

print.summary.lod_lm prints a table containing the coefficient estimates, standard errors, etc. from the lod_lm fit.

Value

The function summary.lod_lm returns a list of summary statistics of the fitted linear model given in object, using the components (list elements) "call" and "terms" from its argument, plus

residuals	residuals computed by lod_lm
coefficients	a p x 4 matrix for the estimated coefficient, its standard error, t-statistic and corresponding (two-sided) p-value.

sigma	the square root of the estimated variance of the random error.
df	degrees of freedom, a vector (p, n-p), where p is the number of regression coefficients and n is the sample size of the data used in the model fitting

Author(s)

Kevin Donovan, kmdono02@ad.unc.edu.

Maintainer: Kevin Donovan <kmdono02@ad.unc.edu>

References

May RC, Ibrahim JG, Chu H (2011). "Maximum likelihood estimation in generalized linear models with multiple covariates subject to detection limits." *Statistics in medicine*, **30**(20), 2551–2561.

See Also

The model fitting function lod_lm, summary.

Examples

```
library(lodr)
## Using example dataset provided in lodr package: lod_data_ex
## 3 covariates: x1, x2, x3 with x2 and x3 subject to a lower limit of
## detection of 0

## nSamples set to 100 for computational speed/illustration purposes only.
## At least 250 is recommended.  Same for boots=0; results in NAs returned for standard errors

fit <- lod_lm(data=lod_data_ex, frmla=y~x1+x2+x3, lod=c(0,0),
                  var_LOD=c("x2", "x3"), nSamples=100, boots=0)
 summary(fit)
```

Index