Package ‘SignifReg’

March 22, 2022

Type Package

Title Consistent Significance Controlled Variable Selection in Generalized Linear Regression

Version 4.3

Date 2022-03-21

Imports car

Author Jongwook Kim, Adriano Zanin Zambom

Maintainer Adriano Zanin Zambom <adriano.zambom@csun.edu>

Description Provides significance controlled variable selection algorithms with different directions (forward, backward, stepwise) based on diverse criteria (AIC, BIC, adjusted r-square, PRESS, or p-value). The algorithm selects a final model with only significant variables defined as those with significant p-values after multiple testing correction such as Bonferroni, False Discovery Rate, etc. See Zambom and Kim (2018) <doi:10.1002/sta4.210>.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2022-03-22 08:20:02 UTC

R topics documented:

SignifReg-package ................................................................. 2
add1SignifReg ................................................................. 2
add1summary ................................................................. 4
drop1SignifReg ................................................................. 5
drop1summary ................................................................. 7
SignifReg ................................................................. 8

Index 11
Description

Provides significance controlled variable selection algorithms with different directions (forward, backward, stepwise) based on diverse criteria (AIC, BIC, adjusted r-square, PRESS, or p-value). The algorithm selects a final model with only significant variables defined as those with significant p-values after multiple testing correction such as Bonferroni, False Discovery Rate, etc. See Zambom and Kim (2018) <doi:10.1002/sta4.210>.

Details

The DESCRIPTION file:

Package: SignifReg
Type: Package
Title: Consistent Significance Controlled Variable Selection in Generalized Linear Regression
Version: 4.3
Date: 2022-03-21
Imports: car
Author: Jongwook Kim, Adriano Zanin Zambom
Maintainer: Adriano Zanin Zambom <adriano.zambom@csun.edu>
Description: Provides significance controlled variable selection algorithms with different directions (forward, backward, stepwise) based on diverse criteria (AIC, BIC, adjusted r-square, PRESS, or p-value). The algorithm selects a final model with only significant variables defined as those with significant p-values after multiple testing correction such as Bonferroni, False Discovery Rate, etc. See Zambom and Kim (2018) <doi:10.1002/sta4.210>.
License: GPL (>=2)

Author(s)

Jongwook Kim, Adriano Zanin Zambom

Maintainer: Adriano Zanin Zambom <adriano.zambom@csun.edu>

References


add1SignifReg  Add a predictor to a (generalized) linear regression model using the forward step in the Significance Controlled Variable Selection method
add1SignifReg

Description

add1SignifReg adds to the model the predictor, out of the available predictors, which minimizes the criterion (AIC, BIC, r-adj, PRESS, max p-value) as long as all the p-values of the predictors in the prospective model (including the prospective predictor) are below the chosen correction method (Bonferroni, FDR, None, etc). The function returns the fitted model with the additional predictor if any. A summary table of the prospective models can be printed with print.step = TRUE.

max_pvalue indicates the maximum p-value from the multiple t-tests for each predictor. More specifically, the algorithm computes the prospective models with each predictor included, and all p-values of this prospective model. Then, the predictor selected to be added to the model is the one whose generating model has the smallest p-values, in fact, the minimum of the maximum p-values in each prospective model.

Usage

add1SignifReg(fit, scope, alpha = 0.05, criterion = "p-value", adjust.method = "fdr", override = FALSE, print.step = FALSE)

Arguments

fit an lm or glm object representing a linear regression model.
scope defines the range of models examined in the stepwise search. This should be either a single formula, or a list containing components upper and lower, both formulae. See the details for how to specify the formulae and how they are used.
alpha Significance level. Default value is 0.05.
criterion Criterion to select predictor variables. criterion = "AIC", criterion = "BIC", criterion = "r-adj" (adjusted r-square), criterion = "PRESS", and criterion = "p-value" are available. Default is p-value.
adjust.method Correction for multiple testing accumulation of error. See p.adjust.
override If override = TRUE, it returns a new lm or glm object that adds a new variable according to criterion even if the new model does not pass the multiple testing p-value correction.
print.step If true, information is printed for each step of variable selection. Default is FALSE.

Value

add1SignifReg returns an object of the class lm or glm for a generalized regression model with the additional component steps.info, which shows the steps taken during the variable selection and model metrics: Deviance, Resid.Df, Resid.Dev, AIC, BIC, adj.rsq, PRESS, max_pvalue, max.VIF, and whether it passed the chosen p-value correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>
Adriano Zanin Zambom <adriano.zambom@gmail.com>
add1summary

Summaries of models when adding a predictor in (generalized) linear models

Description

Offers summaries of prospective models as every available predictor in the scope is added to the model.

Usage

add1summary(fit, scope, alpha = 0.05, adjust.method = "fdr", sort.by = "p-value")

Arguments

fit
an lm or glm object representing a model.

scope
defines the range of models examined in the stepwise search. This should be either a single formula, or a list containing components upper and lower, both formulae. See the details for how to specify the formulae and how they are used.

alpha
Significance level. Default value is 0.05.

adjust.method
Correction for multiple testing accumulation of error. See p.adjust.

sort.by
The criterion to use to sort the table of prospective models. Must be one of criterion = "AIC", criterion = "BIC", criterion = "r-adj" (adjusted r-square), criterion = "PRESS", and criterion = "p-value" are available. Default is p-value.

Examples

## mtcars data is used as an example.

data(mtcars)

nullmodel = lm(mpg~1, mtcars)
fullmodel = lm(mpg~, mtcars)
scope = list(lower=formula(nullmodel),upper=formula(fullmodel))

fit1 <- lm(mpg~1, data = mtcars)
add1SignifReg(fit1, scope = scope, print.step = TRUE)

fit2 <- lm(mpg~disp+cyl+wt+qsec, mtcars)
add1SignifReg(fit2, scope = scope, criterion="AIC", override="TRUE")

References


See Also

SignifReg, add1summary, drop1summary, drop1SignifReg

Examples

## mtcars data is used as an example.

data(mtcars)

nullmodel = lm(mpg~1, mtcars)
fullmodel = lm(mpg~, mtcars)
scope = list(lower=formula(nullmodel),upper=formula(fullmodel))

fit1 <- lm(mpg~1, data = mtcars)
add1SignifReg(fit1, scope = scope, print.step = TRUE)

fit2 <- lm(mpg~disp+cyl+wt+qsec, mtcars)
add1SignifReg(fit2, scope = scope, criterion="AIC", override="TRUE")

add1summary

Summaries of models when adding a predictor in (generalized) linear models
Details

max_pvalue indicates the maximum p-value from the multiple t-tests for each predictor.

Value

a table with the possible inclusions and the metrics of the prospective models: AIC, BIC, adj.rsq, PRESS, max_pvalue, max.VIF, and whether it passed the chosen p-value correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>
Adriano Zanin Zambom <adriano.zambom@gmail.com>

References


See Also

SignifReg, add1SignifReg, drop1summary, drop1SignifReg

Examples

```r
## mtcars data is used as an example.

data(mtcars)

nullmodel = lm(mpg~1, mtcars)
fullmodel = lm(mpg~., mtcars)
scope = list(lower=formula(nullmodel),upper=formula(fullmodel))
fit1 <- lm(mpg~1, mtcars)
add1summary(fit1, scope = scope)

fit2 <- lm(mpg~disp+cyl+wt+qsec+cyl, data = mtcars)
add1summary(fit2, scope = scope)
```

---

drop1SignifReg | Drop a predictor to a (generalized) linear regression model using the backward step in the Significance Controlled Variable Selection method
Description

drop1SignifReg removes from the model the predictor, out of the current predictors, which mini-
mizes the criterion (AIC, BIC, r-adj, PRESS, max p-value) when a) the p-values of the predictors 
in the current model do not pass the multiple testing correction (Bonferroni, FDR, None, etc) or b) 
when the p-values of both current and prospective models pass the correction but the criterion of 
the prospective model is smaller.

max_pvalue indicates the maximum p-value from the multiple t-tests for each predictor. More 
specifically, the algorithm computes the prospective models with each predictor included, and all 
p-values of this prospective model. Then, the predictor selected to be added to the model is the one 
whose generating model has the smallest p-values, in fact, the minimum of the maximum p-values 
in each prospective model.

Usage

drop1SignifReg(fit, scope, alpha = 0.05, criterion = "p-value", 
                 adjust.method = "fdr", override = FALSE, print.step = FALSE)

Arguments

fit       an lm or glm object representing a model.
scope     defines the range of models examined in the stepwise search. This should be 
either a single formula, or a list containing components upper and lower, both 
formulae. See the details for how to specify the formulae and how they are used.
alpha     Significance level. Default value is 0.05.
criterion Criterion to select predictor variables. criterion = "AIC", criterion = "BIC", 
criterion = "r-adj" (adjusted r-square), criterion = "PRESS", and criterion 
= "p-value" are available. Default is p-value.
adjust.method Correction for multiple testing accumulation of error. See p.adjust.
override  If override = TRUE, it returns a new lm or glm object that adds a new variable 
according to criterion even if the new model does not pass the multiple testing 
p-value correction.
print.step If true, information is printed for each step of variable selection. Default is 
FALSE.

Value

drop1SignifReg returns an object of the class lm or glm for a generalized regression model with the 
additional component steps.info, which shows the steps taken during the variable selection and 
model metrics: Deviance, Resid.Df, Resid.Dev, AIC, BIC, adj.rsq, PRESS, max_pvalue, max.VIF, 
and whether it passed the chosen p-value correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>
Adriano Zanin Zambom <adriano.zambom@gmail.com>
References


See Also

SignifReg, add1summary, add1SignifReg, drop1summary.

Examples

```r
## mtcars data is used as an example.

data(mtcars)
fit <- lm(mpg~., mtcars)
drop1SignifReg(fit, print.step = TRUE)
```

---

**drop1summary**

*Summaries of models when removing a predictor in a (generalized) linear model*

**Description**

Offers summaries of prospective models as every predictor in the model is removed from the model.

**Usage**

```r
drop1summary(fit, scope, alpha = 0.05, adjust.method = "fdr", sort.by = "p-value")
```

**Arguments**

- `fit`: an lm or glm object representing a model.
- `scope`: defines the range of models examined in the stepwise search. This should be either a single formula, or a list containing components upper and lower, both formulae. See the details for how to specify the formulae and how they are used.
- `alpha`: Significance level. Default value is 0.05.
- `adjust.method`: Correction for multiple testing accumulation of error. See `p.adjust`.
- `sort.by`: The criterion to use to sort the table of prospective models. Must be one of `criterion = "AIC", criterion = "BIC", criterion = "r-adj"` (adjusted r-square), `criterion = "PRESS", and criterion = "p-value"` are available. Default is p-value.

**Details**

`max_pvalue` indicates the maximum p-value from the multiple t-tests for each predictor.
Value

A table with the possible exclusions and the metrics of the prospective models: AIC, BIC, adj.rsq, PRESS, max_pvalue, max.VIF, and whether it passed the chosen p-value correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>
Adriano Zanin Zambom <adriano.zambom@gmail.com>

References


See Also

SignifReg, add1summary, add1SignifReg, drop1SignifReg.

Examples

```r
## mtcars data is used as an example.

data(mtcars)

fit <- lm(mpg~., mtcars)
drop1summary(fit)
```

---

**SignifReg**

Significance Controlled Variable Selection in (Generalized) Linear Regression

Description

Significance controlled variable selection selects variables in a generalized linear regression model with different directions of the algorithm (forward, backward, stepwise) based on a chosen criterion (AIC, BIC, adjusted r-square, PRESS or p-value). The algorithm selects a final model with only significant variables based on a correction choice of False Discovery Rate, Bonferroni, etc from the p.adjust().

Usage

```r
SignifReg(fit, scope, alpha = 0.05, direction = "forward", criterion = "p-value", adjust.method = "fdr", trace=FALSE)
```
Arguments

fit  
an lm or glm object representing a model. It is an initial model for the variable selection.

scope  
defines the range of models examined in the stepwise search. This should be either a single formula, or a list containing components upper and lower, both formulae. See the details for how to specify the formulae and how they are used.

alpha  
Significance level. Default value is 0.05.

direction  
Direction in variable selection: direction = "both", direction = "forward", and direction = "backward" are available. direction = "both" is a stepwise selection. Default is direction = "forward".

criterion  
Criterion to select predictor variables. criterion = "AIC", criterion = "BIC", criterion = "r-adj" (adjusted r-square), criterion = "PRESS", and criterion = "p-value" are available. Default is p-value.

adjust.method  
Correction for multiple testing accumulation of error. See p.adjust.

trace  
If true, information is printed for each step of variable selection. Default is FALSE. Offers summaries of prospective models as each predictor in the scope is added to or removed from the model. max_pvalue indicates the maximum p-value from the multiple t-tests for each predictor in the model.

Details

SignifReg selects only significant predictors according to a designated criterion. A model with the best criterion, for example, the smallest AIC, will not be considered if it includes insignificant predictors based on the chosen correction. When the criterion is "p-value", a predictor can be dropped only if the current model has an insignificant predictor, and a predictor can be added as long as the prospective model has all predictors significant (including the one to be added). The predictor to be added or removed is the one that generates a model having the smallest maximum p-value of the t-tests in the prospective models. This step is repeated as long as every predictor is significant according to the correction criterion. In the case that the criterion is "AIC", and "BIC", SignifReg selects, at each step, the model having the smallest value of the criterion among models having only significant predictors according to the chosen correction.

Value

SignifReg returns an object of the class lm or glm for a generalized regression model with the additional component steps.info, which shows the steps taken during the variable selection and model metrics: Deviance, Resid.Df, Resid.Dev, AIC, BIC, adj.rsq, PRESS, max_pvalue, max.VIF, and whether it passed the chosen p-value correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>
Adriano Zanin Zambom <adriano.zambom@gmail.com>
References


See Also

add1SignifReg, drop1SignifReg, add1summary, drop1summary

Examples

## mtcars data is used as an example.

data(mtcars)
nullmodel = lm(mpg~1, mtcars)
fullmodel = lm(mpg~., mtcars)
scope = list(lower=formula(nullmodel), upper=formula(fullmodel))

fit1 <- lm(mpg~1, mtcars)
select.fit = SignifReg(fit1, scope = scope, direction = "forward", trace = TRUE)
select.fit$steps.info

fit = lm(mpg ~cyl + hp + am + gear, data = mtcars)
select.fit = SignifReg(fit, scope=scope, alpha = 0.05,direction = "backward",
 criterion = "p-value",adjust.method = "fdr",trace=TRUE)
select.fit$steps.info

fit = lm(mpg ~ cyl + hp + am + gear + disp, data = mtcars)
select.fit = SignifReg(fit,scope=scope, alpha = 0.5,direction = "both",
 criterion = "AIC",adjust.method = "fdr",trace=TRUE)
select.fit$steps.info
Index

add1SignifReg, 2, 5, 7, 8, 10
add1summary, 4, 4, 7, 8, 10

drop1SignifReg, 4, 5, 5, 8, 10
drop1summary, 4, 5, 7, 7, 10

p.adjust, 3, 4, 6, 7, 9

SignifReg, 4, 5, 7, 8, 8
SignifReg-package, 2