Package ‘SignifReg’

June 26, 2019

Type Package

Title Consistent Significance Controlled Variable Selection in Linear Regression

Version 2.1

Date 2019-06-25

Author Jongwook Kim, Adriano Zanin Zambom

Maintainer Jongwook Kim <jongwook226@gmail.com>

Description Provide consistent significance controlled variable selection procedure with different directions (forward, backward, stepwise) based on diverse criteria (AIC, BIC, adjusted r-square, and p-value). The algorithm selects a final model with only significant variables based on a correction choice of False Discovery Rate, Bonferroni, or no correction.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2019-06-26 08:20:03 UTC

R topics documented:

SignifReg-package .......................................................... 2
add1SignifReg .............................................................. 2
add1summary ............................................................... 4
drop1SignifReg ............................................................ 5
drop1summary ............................................................. 6
SignifReg ................................................................. 7

Index 10
**Description**

Provide consistent significance controlled variable selection procedure with different directions (forward, backward, stepwise) based on diverse criteria (AIC, BIC, adjusted r-square, and p-value). The algorithm selects a final model with only significant variables based on a correction choice of False Discovery Rate, Bonferroni, or no correction.

**Details**

The DESCRIPTION file:

```
Package: SignifReg
Type: Package
Title: Consistent Significance Controlled Variable Selection in Linear Regression
Version: 2.1
Date: 2019-06-25
Author: Jongwook Kim, Adriano Zanin Zambom
Maintainer: Jongwook Kim <jongwook226@gmail.com>
Description: Provide consistent significance controlled variable selection procedure with different directions (forward, backward, stepwise) based on diverse criteria (AIC, BIC, adjusted r-square, and p-value). The algorithm selects a final model with only significant variables based on a correction choice of False Discovery Rate, Bonferroni, or no correction.
License: GPL (>=2)
```

**Author(s)**

Jongwook Kim, Adriano Zanin Zambom

Maintainer: Jongwook Kim <jongwook226@gmail.com>

**References**


**Description**

Add a single term to the model using the method Consistent Significance Controlled Variable Selection. The function also provides a summary table of the selection. `max_pvalue` indicates the maximum p-value from the multiple t-tests for each predictor. This value can be used as a criterion...
in the case criterion = "p-value". alpha_cut_off, Bonferroni, and FDR represent whether the entire model satisfies the significance correction. alpha_cut_off means no correction.

Usage

```r
add1SignifReg(fit, scope, alpha = 0.05, criterion = "p-value", correction = "FDR", override = FALSE)
```

Arguments

- **fit**: an lm object representing a model. It is an initial model for the variable selection.
- **scope**: The range of models examined in regression. It should be either a data.frame or formula containing predictors. When scope is data.frame, all variables except the response variable in the data.frame are considered for the variable selection.
- **alpha**: Significance level. Default value is 0.05.
- **criterion**: Criterion to select predictor variables. criterion = "AIC", criterion = "BIC", criterion = "r-adj" (adjusted r-square), and criterion = "p-value" are available. Default is p-value.
- **correction**: Correction criterion to reduce multiple testing error. correction = "FDR" (False Discovery Rate), correction = "Bonferroni", and correction = "None" (no correction) are available. Default is correction = "FDR". For Bonferroni correction, either correction = "Bonferroni" or correction = "Bonf" can be used.
- **override**: If override = TRUE, it returns a new lm object that adds a new variable according to criterion even if the new model does not pass the correction.

Author(s)

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

References


Examples

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

data(mtcars)

fit1 <- lm(mpg~1, mtcars)
add1SignifReg(fit1)

fit2 <- lm(mpg~disp+cyl+wt+qsec, mtcars)
add1SignifReg(fit2, criterion="AIC", override="TRUE")
```
**add1summary**

*Summaries of add1 or drop1*

**Description**

Offers summaries of information as every single 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. `alpha_cut_off`, Bonferroni, and FDR represent whether the model satisfies the significance by correction. `alpha_cut_off` means no correction.

**Usage**

```r
add1summary(fit, scope, alpha = 0.05)
```

**Arguments**

- **fit**: an `lm` object representing a model. It is an initial model for the variable selection.
- **scope**: The range of models examined in regression. It should be either a `data.frame` of formula containing predictors. When scope is `data.frame`, all variables except the response variable in the `data.frame` are considered for the variable selection. See the examples how they can be used.
- **alpha**: Significance level. Default value is 0.05.

**Author(s)**

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

**References**


**Examples**

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

data(mtcars)

fit1 <- lm(mpg~1, mtcars)
add1summary(fit1)
add1summary(fit1, scope=~.+disp+cyl+wt+qsec+cyl, alpha=0.025)

fit2 <- lm(mpg~disp+cyl+wt+qsec+cyl, mtcars)
add1summary(fit2, mtcars)
```
drop1SignifReg

Drop a possible single predictor to a regression model based on Consistent Significance Controlled Variable Selection

Description

Add or drop a single term to the model using the method Consistent Significance Controlled Variable Selection. The function also provides a summary table of the selection. max_pvalue indicates the maximum p-value from the multiple t-tests for each predictor. This value can be used as a criterion in the case criterion = "p-value". alpha_cut_off, Bonferroni, and FDR represent whether the entire model satisfies the significance correction. alpha_cut_off means no correction.

Usage

drop1SignifReg(fit, scope, alpha = 0.05, criterion = "p-value", correction = "FDR", override = FALSE)

Arguments

fit an lm object representing a model. It is an initial model for the variable selection
scope The range of models examined in regression. It should be either a data.frame or formula containing predictors. When scope is data.frame, all variables except the response variable in the data.frame are considered for the variable selection.
alpha Significance level. Default value is 0.05.
criterion Criterion to select predictor variables. criterion = "AIC", criterion = "BIC", criterion = "r-adj" (adjusted r-square), and criterion = "p-value" are available. Default is p-value.
correction Correction criterion to reduce multiple testing error. correction = "FDR" (False Discovery Rate), correction = "Bonferroni", and correction = "None" (no correction) are available. Default is correction = "FDR". For Bonferroni correction, either correction = "Bonferroni" or correction = "Bonf" can be used.
override If override = TRUE, it returns a new lm object that excludes a new variable according to criterion even if the new model does not pass the correction.

Author(s)

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

References

Examples

```
# mtcars data is used as an example.
data(mtcars)
fit3 <- lm(mpg~., mtcars)
drop1SignifReg(fit3)
drop1SignifReg(fit3, scope=~.-disp-cyl-wt-hp-drat-qset-carb, correction="Bonf")
```

Description

Offers summaries of information as every single 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. `alpha_cut_off`, `Bonferroni`, and `FDR` represent whether the model satisfies the significance by correction. `alpha_cut_off` means no correction.

Usage

```
drop1summary(fit, scope, alpha = 0.05)
```

Arguments

- **fit**: an `lm` object representing a model. It is an initial model for the variable selection.
- **scope**: The range of models examined in regression. It should be either a `data.frame` of formula containing predictors. When `scope` is `data.frame`, all variables except the response variable in the `data.frame` are considered for the variable selection. See the examples how they can be used.
- **alpha**: Significance level. Default value is 0.05.

Author(s)

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

References

Examples

# mtcars data is used as an example.
data(mtcars)

fit3 <- lm(mpg ~ ., mtcars)
drop1(summary(fit3))
drop1(summary(fit3, scope = ~ .-disp-cyl-wt-qsec-cyl, alpha = 0.025))

Description

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

Usage

SignifReg(fit, scope, alpha = 0.05, direction = "forward", criterion = "p-value", correction = "FDR", trace = FALSE)

Arguments

- **fit**: an lm object representing a model. It is an initial model for the variable selection.
- **scope**: The range of models examined in regression. It should be either a data.frame or formula containing predictors. When scope is data.frame, all variables except the response variable in the data.frame are considered for the variable selection.
- **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), and criterion = "p-value" are available. Default is criterion = "p-value".
- **correction**: Correction for multiple testing accumulation of error. correction = "FDR" (False Discovery Rate), correction = "Bonferroni", and correction = "None" (no correction) are available. Default is correction = "FDR". For Bonferroni correction, either correction = "Bonferroni" or correction = "Bonf" can be used.
trace If true, information is printed for each step of variable selection. Default is FALSE. Offers summaries of information as every single 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. This value can be used as a criterion in the case criterion = "p-value". alpha_cut_off, Bonferroni, and FDR represent whether the model satisfies the significance by correction. alpha_cut_off means no correction and all p-values will be compared to it.

Details

SignifReg selects only significant predictors according to a designated criterion. Although a model has the best feature for a chosen criterion, for example, the smallest AIC, this model will be excluded from the selection if it includes insignificant predictors based on the chosen correction. When the criterion is "p-value", at each step, the variable to be added or substracted is the one that generates a model having the smallest maximum p-value of the t-tests. 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. For "r-adj(adjusted r-square), SignifReg selects, at each step, the model having the largest value of adjusted r-square among models having only significant predictors according to the chosen correction.

Value

SignifReg returns an object of the class lm for a regression model.

Author(s)

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

References


Examples

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

data(mtcars)

fit1 <- lm(mpg~1, mtcars)
summary(SignifReg(fit1))
summary(SignifReg(fit1, scope=mtcars, criterion="r-adj", direction="forward", correction="None", trace=TRUE))
summary(SignifReg(fit1, scope=~.+disp+cyl+wt+qsec, criterion="p-value", direction="both", correction="FDR", trace=TRUE))
```
fit2 <- lm(mpg~., mtcars)
summary(SignifReg(fit2, scope=~.disp-cyl-wt-hp-drat-qset-carb, direction="backward",
correction="Bonf", trace=TRUE))
summary(SignifReg(fit2, scope=mtcars,criterion="AIC", direction="both",
correction="None", trace=TRUE))
Index

add1SignifReg, 2
add1summary, 4

drop1SignifReg, 5
drop1summary, 6

SignifReg, 7
SignifReg-package, 2