Package 'SignifReg'

April 17, 2020

Type Package

Title Consistent Significance Controlled Variable Selection in Linear
    Regression

Version 3.0

Date 2020-04-16

Author Jongwook Kim, Adriano Zanin Zambom

Maintainer Jongwook Kim <jongwook226@gmail.com>

Description Provides significance controlled variable selection algorithms with different direc-
    tions (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 vari-
    ables based on a correction choice of False Discovery Rate, Bonferroni, or no correction.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2020-04-17 04:40:06 UTC

R topics documented:

SignifReg-package	*Consistent Significance Controlled Variable Selection in Linear Regression*

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 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:	3.0
Date:	2020-04-16
Author:	Jongwook Kim, Adriano Zanin Zambom
Maintainer:	Jongwook Kim <jongwook226@gmail.com>
Description:	Provides significance controlled variable selection algorithms with different directions (forward, backward, step)
License:	GPL (>=2)

Author(s)

Jongwook Kim, Adriano Zanin Zambom

Maintainer: Jongwook Kim <jongwook226@gmail.com>

References

Zambom A Z, Kim J. Consistent significance controlled variable selection in high-dimensional regression. Stat.2018;7:e210. https://doi.org/10.1002/sta4.210

---

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

Description

add1SignifReg includes in the model the predictor, out of the available predictors, which minimizes the criterion (AIC, BIC, r-ajd, 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). The function provides a summary table of the prospective models.

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 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",
 correction = "FDR", override = FALSE)
```

Arguments

fit	an lm object representing a linear regression model.
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.
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 multiple testing p-value correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>

Adriano Zanin Zambom <adriano.zambom@gmail.com>

References

Zambom A Z, Kim J. Consistent significance controlled variable selection in high-dimensional regression. Stat.2018;7:e210. https://doi.org/10.1002/sta4.210

See Also

SignifReg, add1summary, drop1summary, drop1SignifReg

Examples

```
##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 models when adding a predictor.*

Description

Offers summaries of prospective models as every available predictor in the scope is added to 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 correction. `alpha_cut_off` means no correction.

Usage

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

Arguments

fit	an lm object representing a model.
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

Zambom A Z, Kim J. Consistent significance controlled variable selection in high-dimensional regression. Stat.2018;7:e210. https://doi.org/10.1002/sta4.210

See Also

SignifReg, add1SignifReg, drop1summary, drop1SignifReg

Examples

```
##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 predictor to a 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-ajd, 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) 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 removed, and all
p-values of this prospective model. Then, the predictor to be removed 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",
 correction = "FDR", override = FALSE)
```

Arguments

fit	an lm object representing a model.
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 predictor according to criterion even if the prospective model passes the correction.

Author(s)

Jongwook Kim <jongwook226@gmail.com>

Adriano Zanin Zambom <adriano.zambom@gmail.com>

References

Zambom A Z, Kim J. Consistent significance controlled variable selection in high-dimensional regression. Stat.2018;7:e210. https://doi.org/10.1002/sta4.210

See Also

SignifReg, add1summary, add1SignifReg, drop1summary,

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")
```

---

drop1summary	*Summaries of models when removing a predictor.*

Description

Offers summaries of prospective models as every predictor in the model is 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 correction. `alpha_cut_off` means no correction.

Usage

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

Arguments

`fit`	an lm object representing a model.
`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

Zambom A Z, Kim J. Consistent significance controlled variable selection in high-dimensional regression. Stat.2018;7:e210. https://doi.org/10.1002/sta4.210

See Also

[SignifReg](#), [add1summary](#), [add1SignifReg](#), [drop1SignifReg](#),

Examples

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

data(mtcars)

fit3 <- lm(mpg~., mtcars)
drop1summary(fit3)
drop1summary(fit3, scope=~.-disp-cyl-wt-qsec-cyl, alpha=0.025)
```

SignifReg   *Significance Controlled Variable Selection in Linear Regression*

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, PRESS or 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 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 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. 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. 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", at each step, the variable 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

SifnifReg 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

Zambom A Z, Kim J. Consistent significance controlled variable selection in high-dimensional regression. Stat.2018;7:e210. https://doi.org/10.1002/sta4.210

See Also

add1SignifReg, drop1SignifReg, add1summary, drop1summary

Examples

```
##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