Package ‘StepReg’

December 17, 2022

Title  Stepwise Regression Analysis
Version  1.4.4
Date  2022-12-10
Author  Junhui Li,Xiaohuan Lu,Kun Cheng,Wenxin Liu
Maintainer  Junhui Li <junhuili@cau.edu.cn>
Description  Three most common types of stepwise regression including linear regression, logistic regression and cox proportional hazard regression can be performed to select best model with methods of forward selection, backward elimination, bidirectional selection and best subset selection. A widely used selection criteria are available for variable selection.
License  MIT + file LICENSE
VignetteBuilder  knitr
Suggests  knitr
Imports  dplyr, purrr, stringr, survival
Encoding  UTF-8
RoxygenNote  7.2.3
NeedsCompilation  no
Repository  CRAN
Date/Publication  2022-12-17 16:20:02 UTC

R topics documented:

modelFitStat .................................................. 2
print.StepReg ................................................. 3
stepwise ..................................................... 4
stepwiseCox ................................................... 6
stepwiseLogit ............................................... 9

Index  12
**modelFitStat**

**Fit Model Statistics**

**Description**

Fit Model Statistics with least square or likelihood method to return an information criteria value.

**Usage**

```r
modelFitStat(ic, fit, method = c("LeastSquare", "Likelihood"), cox = FALSE)
```

**Arguments**

- **ic**: Information criteria, including AIC, AICc, BIC, CP, HQ, HQc, Rsq, adjRsq and SBC.
- **fit**: Object of linear model or general linear model.
- **method**: Method to calculate information criteria value, including 'LeastSquare' and 'Likelihood'.
- **cox**: Compute model fit statistics for cox regression or not, where partial likelihood value will be used instead of the ordinary.

**Author(s)**

Junhui Li

**References**


### Examples

data(mtcars)
fit <- lm(mpg~wt+qsec+vs+am+gear+carb,data=mtcars)
modelFitStat("AIC",fit,"LeastSquare")

---

**print.StepReg**

*Prints from a StepReg object*

**Description**

print.StepReg prints to console the from an object of class StepReg

**Usage**

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

**Arguments**

- `x` each dataframe from outputlist
- `...` further parameters

**Value**

formatted dataframe
Stepwise Linear Model Regression

Description

Stepwise linear regression analysis selects models based on information criteria and F or approximate F test with 'forward', 'backward', 'bidirection' and 'score' model selection methods.

Usage

```r
stepwise(
  formula, 
  data, 
  include = NULL, 
  selection = c("forward", "backward", "bidirection", "score"), 
  select = c("AIC", "AICc", "BIC", "CP", "HQ", "HQc", "Rsq", "adjRsq", "SL", "SBC"), 
  sle = 0.15, 
  sls = 0.15, 
  multivarStat = c("Pillai", "Wilks", "Hotelling-Lawley", "Roy"), 
  weights = NULL, 
  best = NULL
)
```

Arguments

- `formula`: Model formulae. The models fitted by the `lm` functions are specified in a compact symbolic form. The basic structure of a formula is the tilde symbol (~) and at least one independent (righthand) variable. In most (but not all) situations, a single dependent (lefthand) variable is also needed. Thus we can construct a formula quite simple formula (y ~ x). Multiple independent variables by simply separating them with the plus (+) symbol (y ~ x1 + x2). Variables in the formula are removed with a minus(-) symbol (y ~ x1 - x2). One particularly useful feature is the . operator when modelling with lots of variables (y ~ .). The %in% operator indicates that the terms on its left are nested within those on the right. For example y ~ x1 + x2 %in% x1 expands to the formula y ~ x1 + x1:x2. A model with no intercept can be specified as y ~ x - 1 or y ~ x + 0 or y ~ 0 + x. Multivariate multiple regression can be specified as cbind(y1,y2) ~ x1 + x2.

- `data`: Data set including dependent and independent variables to be analyzed.

- `include`: Force vector of effects name to be included in all models.

- `selection`: Model selection method including 'forward', 'backward', 'bidirection' and 'score'. forward selection starts with no effects in the model and adds effects, backward selection starts with all effects in the model and removes effects, while bidirection regression is similar to the forward method except that effects already in the model do not necessarily stay there, and score method requests specifies the best-subset selection method, which uses the branch-and-bound technique to efficiently search for subsets of model effects that best predict the response variable.

- `select`: Selection criteria including 'AIC', 'AICc', 'BIC', 'CP', 'HQ', 'HQc', 'Rsq', 'adjRsq', 'SL', 'SBC'.

- `sle`: Selection level for exclusion.

- `sls`: Selection level for inclusion.

- `multivarStat`: Multivariate test statistics including 'Pillai', 'Wilks', 'Hotelling-Lawley', 'Roy'.

- `weights`: Weight vector.

- `best`: Force vector of effects name to be selected as the best model.
stepwise

**select** Specify the criterion that uses to determine the order in which effects enter and leave at each step of the specified selection method including "AIC", "AICc", "BIC", "CP", "HQ", "HQc", "R2", and "SL".

**sle** Specify the significance level for entry, default is 0.15

**sls** Specify the significance level for staying in the model, default is 0.15

**multivarStat** Statistic for multivariate regression analysis, including Wilks' lamda ("Wilks"), Pillai Trace ("Pillai"), Hotelling-Lawley’s Trace ("Hotelling"), Roy’s Largest Root ("Roy")

**weights** Numeric vector to provide a weight for each observation in the input data set. Note that weights should be ranged from 0 to 1, while negative numbers are forcibly converted to 0, and numbers greater than 1 are forcibly converted to 1. If you do not specify a weight vector, each observation has a default weight of 1.

**best** Control the number of models displayed in the output, default is NULL, which means all possible model will be displayed.

**Author(s)**

Junhui Li

**References**


Examples

data(mtcars)
mtcars$yes <- mtcars$wt
formula <- cbind(mpg,drat) ~ . + 0
stepwise(formula=formula,
data=mtcars,
   selection="bidirection",
   select="AIC")

---

stepwiseCox  

Stepwise Cox Proportional Hazards Regression

Description

Stepwise Cox regression analysis selects model based on information criteria and significant test with 'forward', 'backward', 'bidirection' and 'score' variable selection method.

Usage

stepwiseCox(
   formula,
data,
   include = NULL,
   selection = c("forward", "backward", "bidirection", "score"),
   select = c("SL", "AIC", "AICc", "SBC", "HQ", "HQc", "IC(3/2)", "IC(1)"),
   sle = 0.15,
   sls = 0.15,
   method = c("efron", "breslow", "exact"),
   weights = NULL,
   best = NULL
)
stepwiseCox

Arguments

**formula**
Model formulae. The models fitted by the coxph functions are specified in a compact symbolic form. The basic structure of a formula is the tilde symbol (~) and at least one independent (righthand) variable. In most (but not all) situations, a single dependent (lefthand) variable is also needed. Thus we can construct a formula quite simple formula (y ~ x). Multiple independent variables by simply separating them with the plus (+) symbol (y ~ x1 + x2). Variables in the formula are removed with a minus(-) symbol (y ~ x1 - x2). One particularly useful feature is the . operator when modelling with lots of variables (y ~ .). The %in% operator indicates that the terms on its left are nested within those on the right. For example y ~ x1 + x2 %in% x1 expands to the formula y ~ x1 + x1:x2.

**data**
Data set including dependent and independent variables to be analyzed

**include**
Force the effects vector listed in the data to be included in all models. The selection methods are performed on the other effects in the data set

**selection**
Model selection method including "forward", "backward", "bidirection" and 'score', forward selection starts with no effects in the model and adds effects, backward selection starts with all effects in the model and removes effects, while bidirection regression is similar to the forward method except that effects already in the model do not necessarily stay there, and score method requests best subset selection.

**select**
Specify the criterion that uses to determine the order in which effects enter and leave at each step of the specified selection method including AIC, AICc, SBC, IC(1), IC(3/2), HQ, HQc and Significant Levels(SL)

**sle**
Specify the significance level for entry, default is 0.15

**sls**
Specify the significance level for staying in the model, default is 0.15

**method**
Specify the method for tie handling. If there are no tied death times all the methods are equivalent. Nearly all Cox regression programs use the Breslow method by default, but not this one. The Efron approximation is used as the default here, it is more accurate when dealing with tied death times, and is as efficient computationally. The “exact partial likelihood is equivalent to a conditional logistic model, and is appropriate when the times are a small set of discrete values.

**weights**
Numeric vector to provide a weight for each observation in the input data set. Note that weights should be ranged from 0 to 1, while negative numbers are forcibly converted to 0, and numbers greater than 1 are forcibly converted to 1. If you do not specify a weight vector, each observation has a default weight of 1.

**best**
Control the number of models displayed in the output, default is NULL which means all possible model will be displayed

Author(s)

Junhui Li

References


Examples

```r
lung <- survival::lung
my.data <- na.omit(lung)
my.data$status1 <- ifelse(my.data$status==2,1,0)
data <- my.data
formula = Surv(time, status1) ~ . ~ status

stepwiseCox(formula=formula, data=my.data, selection="bidirection", select="HQ", method="efron")
```
Stepwise logistic regression analysis selects model based on information criteria and Wald or Score test with 'forward', 'backward', 'bidirection' and 'score' model selection method.

### Usage

```r
codes
stepwiseLogit(
  formula,
  data,
  include = NULL,
  selection = c("forward", "backward", "bidirection", "score"),
  select = c("SL", "AIC", "AICc", "SBC", "HQ", "HQc", "IC(3/2)", "IC(1)"),
  sle = 0.15,
  sls = 0.15,
  sigMethod = c("Rao", "LRT"),
  weights = NULL,
  best = NULL
)
```

### Arguments

- **formula**: Model formulae. The models fitted by the glm functions are specified in a compact symbolic form. The basic structure of a formula is the tilde symbol (~) and at least one independent (righthand) variable. In most (but not all) situations, a single dependent (lefthand) variable is also needed. Thus we can construct a formula quite simple formula (y ~ x). Multiple independent variables by simply separating them with the plus (+) symbol (y ~ x1 + x2). Variables in the formula are removed with a minus(-) symbol (y ~ x1 - x2). One particularly useful feature is the . operator when modelling with lots of variables (y ~ .). The %in% operator indicates that the terms on its left are nested within those on the right. For example y ~ x1 + x2 %in% x1 expands to the formula y ~ x1 + x1:x2. A model with no intercept can be specified as y ~ x - 1 or y ~ x + 0 or y ~ 0 + x.

- **data**: Data set including dependent and independent variables to be analyzed

- **include**: Force the effects vector listed in the data to be included in all models. The selection methods are performed on the other effects in the data set

- **selection**: Model selection method including "forward", "backward", "bidirection" and 'score'.forward selection starts with no effects in the model and adds effects, backward selection starts with all effects in the model and removes effects, while bidirection regression is similar to the forward method except that effects already in the model do not necessarily stay there, and score method requests best subset selection.
select  Specify the criterion that uses to determine the order in which effects enter and leave at each step of the specified selection method including AIC, AICc, SBC, IC(1), IC(3/2), HQ, HQc and Significant Levels(SL)

sle    Specify the significance level for entry, default is 0.15

sls    Specify the significance level for staying in the model, default is 0.15

sigMethod  Specify the method of significant test for variable to be entered in the model. "Rao" and "LRT" cab be chosen for Rao's efficient score test and likelihood ratio test.

weights Numeric vector to provide a weights for each observation in the input data set. Note that weights should be ranged from 0 to 1, while negative numbers are forcibly converted to 0, and numbers greater than 1 are forcibly converted to 1. If you do not specify a weights vector, each observation has a default weights of 1.

best   Control the number of models displayed in the output, default is NULL which means all possible model will be displayed

Author(s)

Junhui Li

References


Examples

data(mtcars)
formula <- vs ~ .
stepwiseLogit(formula,
data=mtcars,
  selection="bidirection",
  select="SL",
  sle=0.15,
  sls=0.15,
  sigMethod="Rao")
Index

* logistic
  stepwiseLogit, 9
* regression
  stepwise, 4
  stepwiseLogit, 9
* stepwise
  stepwise, 4
  stepwiseLogit, 9

modelFitStat, 2
print.StepReg, 3

stepwise, 4
stepwiseCox, 6
stepwiseLogit, 9