

# Package ‘My.stepwise’

January 20, 2025

**Type** Package

**Title** Stepwise Variable Selection Procedures for Regression Analysis

**Version** 0.1.0

**Author** International-Harvard Statistical Consulting Company <nhsc2010@hotmail.com>

**Maintainer** Fu-Chang Hu <fuchang.hu@gmail.com>

**Description** The stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) can be used to obtain the best candidate final regression model in regression analysis. All the relevant covariates are put on the 'variable list' to be selected. The significance levels for entry (SLE) and for stay (SLS) are usually set to 0.15 (or larger) for being conservative. Then, with the aid of substantive knowledge, the best candidate final regression model is identified manually by dropping the covariates with p value > 0.05 one at a time until all regression coefficients are significantly different from 0 at the chosen alpha level of 0.05.

**License** GPL (>= 3)

**Encoding** UTF-8

**LazyData** True

**Depends** R (>= 3.3.3)

**Imports** car, lmtest, survival, stats

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2017-06-29 09:13:47 UTC

## Contents

|                             |   |
|-----------------------------|---|
| My.stepwise.coxph . . . . . | 2 |
| My.stepwise.glm . . . . .   | 4 |
| My.stepwise.lm . . . . .    | 5 |

|              |          |
|--------------|----------|
| <b>Index</b> | <b>8</b> |
|--------------|----------|

---

My.stepwise.coxph      *Stepwise Variable Selection Procedure for Cox's Proportional Hazards Model and Cox's Model*

---

### Description

This stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) can be applied to obtain the best candidate final Cox's proportional hazards model or Cox's proportional hazards model with time-dependent covariates (called the Cox's model).

### Usage

```
My.stepwise.coxph(Time = NULL, T1 = NULL, T2 = NULL, Status = NULL,
  variable.list, in.variable = "NULL", data, sle = 0.15, sls = 0.15,
  vif.threshold = 999)
```

### Arguments

|               |  |
|---------------|--|
| Time          | The 'Time' (time to an event) for the specified Cox's proportional hazards model as in <code>coxph()</code> .  |
| T1            | The 'T1' (Start) of the long-form data for the specified Cox's model as in <code>coxph()</code> .              |
| T2            | The 'T2' (Stop) of the long-form data for the specified Cox's model as in <code>coxph()</code> .               |
| Status        | The 'Status' (event indicator) for the specified Cox's proportional hazards model as in <code>coxph()</code> . |
| variable.list | A list of covariates to be selected.   |
| in.variable   | A list of covariate(s) to be always included in the regression model.  |
| data          | The data to be analyzed.   |
| sle           | The chosen significance level for entry (SLE).   |
| sls           | The chosen significance level for stay (SLS).  |
| vif.threshold | The chosen threshold value of variance inflating factor (VIF).   |

### Details

The goal of regression analysis is to find one or a few parsimonious regression models that fit the observed data well for effect estimation and/or outcome prediction. To ensure a good quality of analysis, the model-fitting techniques for (1) variable selection, (2) goodness-of-fit assessment, and (3) regression diagnostics and remedies should be used in regression analysis. The stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) is one of the best ways to obtaining the best candidate final regression model. All the bivariate significant and non-significant relevant covariates and some of their interaction terms (or moderators) are put on the 'variable list' to be selected. The significance levels for entry (SLE) and for stay (SLS) are suggested to be set at 0.15 or larger for being conservative. Then, with the aid of substantive knowledge, the best candidate final regression model is identified manually by dropping the covariates with  $p$  value  $> 0.05$  one at a time until all regression coefficients are significantly different from 0 at

the chosen alpha level of 0.05. Since the statistical testing at each step of the stepwise variable selection procedure is conditioning on the other covariates in the regression model, the multiple testing problem is not of concern. Any discrepancy between the results of bivariate analysis and regression analysis is likely due to the confounding effects of uncontrolled covariates in bivariate analysis or the masking effects of intermediate variables (or mediators) in regression analysis.

### Value

A model object representing the identified "Stepwise Final Model" with the values of variance inflating factor (VIF) for all included covariates is displayed.

### Warning

The value of variance inflating factor (VIF) is bigger than 10 in continuous covariates or VIF is bigger than 2.5 in categorical covariates indicate the occurrence of multicollinearity problem among some of the covariates in the fitted regression model.

### See Also

[My.stepwise.lm](#)

[My.stepwise.glm](#)

### Examples

```
## Not run:
The data 'lung' is available in the 'survival' package.
## End(Not run)

if (requireNamespace("survival", quietly = TRUE)) {
  lung <- survival::lung
}

names(lung)
dim(lung)
my.data <- na.omit(lung)
dim(my.data)
head(my.data)
my.data$status1 <- ifelse(my.data$status==2,1,0)
my.variable.list <- c("inst", "age", "sex", "ph.ecog", "ph.karno", "pat.karno")
My.stepwise.coxph(Time = "time", Status = "status1", variable.list = my.variable.list,
  in.variable = c("meal.cal", "wt.loss"), data = my.data)

my.variable.list <- c("inst", "age", "sex", "ph.ecog", "ph.karno", "pat.karno", "meal.cal",
  "wt.loss")
My.stepwise.coxph(Time = "time", Status = "status1", variable.list = my.variable.list,
  data = my.data, sle = 0.25, sls = 0.25)
```

**Description**

This stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) can be applied to obtain the best candidate final generalized linear model.

**Usage**

```
My.stepwise.glm(Y, variable.list, in.variable = "NULL", data, sle = 0.15,
  sls = 0.15, myfamily, myoffset = "NULL")
```

**Arguments**

|               |   |
|---------------|---|
| Y             | The response variable.  |
| variable.list | A list of covariates to be selected.                                  |
| in.variable   | A list of covariate(s) to be always included in the regression model. |
| data          | The data to be analyzed.  |
| sle           | The chosen significance level for entry (SLE).                        |
| sls           | The chosen significance level for stay (SLS).                         |
| myfamily      | The 'family' for the specified generalized linear model as in glm().  |
| myoffset      | The 'offset' for the specified generalized linear model as in glm().  |

**Details**

The goal of regression analysis is to find one or a few parsimonious regression models that fit the observed data well for effect estimation and/or outcome prediction. To ensure a good quality of analysis, the model-fitting techniques for (1) variable selection, (2) goodness-of-fit assessment, and (3) regression diagnostics and remedies should be used in regression analysis. The stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) is one of the best ways to obtaining the best candidate final regression model. All the bivariate significant and non-significant relevant covariates and some of their interaction terms (or moderators) are put on the 'variable list' to be selected. The significance levels for entry (SLE) and for stay (SLS) are suggested to be set at 0.15 or larger for being conservative. Then, with the aid of substantive knowledge, the best candidate final regression model is identified manually by dropping the covariates with p value > 0.05 one at a time until all regression coefficients are significantly different from 0 at the chosen alpha level of 0.05. Since the statistical testing at each step of the stepwise variable selection procedure is conditioning on the other covariates in the regression model, the multiple testing problem is not of concern. Any discrepancy between the results of bivariate analysis and regression analysis is likely due to the confounding effects of uncontrolled covariates in bivariate analysis or the masking effects of intermediate variables (or mediators) in regression analysis.

**Value**

A model object representing the identified "Stepwise Final Model" with the values of variance inflating factor (VIF) for all included covarites is displayed.

**Warning**

The value of variance inflating factor (VIF) is bigger than 10 in continuous covariates or VIF is bigger than 2.5 in categorical covariates indicate the occurrence of multicollinearity problem among some of the covariates in the fitted regression model.

**See Also**

[My.stepwise.lm](#)

[My.stepwise.coxph](#)

**Examples**

```
data("iris")
names(iris)
my.data <- iris[51:150, ]
my.data$Width <- (my.data$Sepal.Width + my.data$Petal.Width)/2
names(my.data)
dim(my.data)
my.data$Species1 <- ifelse(my.data$Species == "virginica", 1, 0)
my.variable.list <- c("Sepal.Length", "Petal.Length")
My.stepwise.glm(Y = "Species1", variable.list = my.variable.list,
  in.variable = c("Width"), data = my.data, myfamily = "binomial")

my.variable.list <- c("Sepal.Length", "Sepal.Width", "Width")
My.stepwise.glm(Y = "Species1", variable.list = my.variable.list,
  data = my.data, sle = 0.25, sls = 0.25, myfamily = "binomial")
```

---

My.stepwise.lm

*Stepwise Variable Selection Procedure for Linear Regression Model*

---

**Description**

This stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) can be applied to obtain the best candidate final linear regression model.

**Usage**

```
My.stepwise.lm(Y, variable.list, in.variable = "NULL", data, sle = 0.15,
  sls = 0.15)
```

**Arguments**

|                            |   |
|----------------------------|---|
| <code>Y</code>             | The response variable.  |
| <code>variable.list</code> | A list of covariates to be selected.                                  |
| <code>in.variable</code>   | A list of covariate(s) to be always included in the regression model. |
| <code>data</code>          | The data to be analyzed.  |
| <code>sle</code>           | The chosen significance level for entry (SLE).                        |
| <code>sls</code>           | The chosen significance level for stay (SLS).                         |

**Details**

The goal of regression analysis is to find one or a few parsimonious regression models that fit the observed data well for effect estimation and/or outcome prediction. To ensure a good quality of analysis, the model-fitting techniques for (1) variable selection, (2) goodness-of-fit assessment, and (3) regression diagnostics and remedies should be used in regression analysis. The stepwise variable selection procedure (with iterations between the 'forward' and 'backward' steps) is one of the best ways to obtaining the best candidate final regression model. All the bivariate significant and non-significant relevant covariates and some of their interaction terms (or moderators) are put on the 'variable list' to be selected. The significance levels for entry (SLE) and for stay (SLS) are suggested to be set at 0.15 or larger for being conservative. Then, with the aid of substantive knowledge, the best candidate final regression model is identified manually by dropping the covariates with  $p$  value  $> 0.05$  one at a time until all regression coefficients are significantly different from 0 at the chosen alpha level of 0.05. Since the statistical testing at each step of the stepwise variable selection procedure is conditioning on the other covariates in the regression model, the multiple testing problem is not of concern. Any discrepancy between the results of bivariate analysis and regression analysis is likely due to the confounding effects of uncontrolled covariates in bivariate analysis or the masking effects of intermediate variables (or mediators) in regression analysis.

**Value**

A model object representing the identified "Stepwise Final Model" with the values of variance inflating factor (VIF) for all included covarites is displayed.

**Warning**

The value of variance inflating factor (VIF) is bigger than 10 in continuous covariates or VIF is bigger than 2.5 in categorical covariates indicate the occurrence of multicollinearity problem among some of the covariates in the fitted regression model.

**See Also**

[My.stepwise.glm](#)

[My.stepwise.coxph](#)

**Examples**

```
data("LifeCycleSavings")
names(LifeCycleSavings)
dim(LifeCycleSavings)
my.variable.list <- c("pop15", "pop75", "dpi")
My.stepwise.lm(Y = "sr", variable.list = my.variable.list, in.variable = c("ddpi"),
  data = LifeCycleSavings)

my.variable.list <- c("pop15", "pop75", "dpi", "ddpi")
My.stepwise.lm(Y = "sr", variable.list = my.variable.list,
  data = LifeCycleSavings, sle = 0.25, sls = 0.25)
```

# Index

My.stepwise.coxph, 2, 5, 6

My.stepwise.glm, 3, 4, 6

My.stepwise.lm, 3, 5, 5