

# Package ‘StepReg’

March 23, 2020

**Type** Package

**Title** Stepwise Regression Analysis

**Version** 1.4.1

**Date** 2020-03-23

**Author** Junhui Li,Xiaohuan Lu,Kun Cheng,Wenxin Liu

**Maintainer** Junhui Li <junhuili@cau.edu.cn>

**Description** Stepwise regression analysis for variable selection can be used to get the best candidate final regression model with the forward selection, backward elimination and bidirectional elimination approaches. Best subset selection fit a separate least squares regression for each possible combination of all predictors. Both the above two procedures in this package can use weighted data to get best regression model in univariate regression and multivariate regression analysis(Alsubaihi, A. A., (2002) <doi:10.18637/jss.v007.i12>). And continuous variables nested within class effect is also considered in both two procedures. Also stepwise logistic regression in this package can performed with binary dependent variable(Agresti, A. (1984) <doi:10.1002/9780470594001> and Agresti, A. (2014) <doi:10.1007/978-3-642-04898-2\_161>). A widely used selection criteria are available which includes Akaike information criterion(Darlington, R. B. (1968) <doi:10.1037/h0025471>, Judge, G. G. (1985) <doi:10.2307/1391738>), corrected Akaike information criterion(Hurvich, C. M., and Tsai, C. (1989) <doi:10.1093/biomet/76.2.297>), Bayesian information criterion(Sawa, T. (1978) <doi:10.2307/1913828>, Judge, G. G. (1985) <doi:10.2307/1391738>), Mallows Cp statistic(Mallows, C. L. (1973) <doi:10.1080/00401706.1995.10484370>, Hocking, R. R. (1976) <doi:10.2307/2529336>), Hannan and Quinn information criterion(Hannan, E. J. and Quinn, B. G. (1979) <doi:10.1111/j.2517-6161.1979.tb01072.x>, Mcquarrie, A. D. R. and Tsai, C. L. (1998) <doi:10.1142/3573>), corrected Hannan and Quinn information criterion(Mcquarrie, A. D. R. and Tsai, C. L. (1998) <doi:10.1142/3573>), Schwarz criterion(Schwarz, G. (1978) <doi:10.1214/aos/1176344136>, Judge, G. G. (1985) <doi:10.2307/1391738>), adjusted R-square statistic(Darlington, R. B. (1968) <doi:10.1037/h0025471>, Judge, G. G. (1985) <doi:10.2307/1391738>) and significance levels(Mckean, J. J. (1974) <doi:10.1093/biomet/61.2.381>, Harold Hotelling. (1992) <doi:10.1007/978-1-4612-0919-5\_4>, Pillai, K. C. S. (2006) <doi:10.1002/0471667196.ess1965.pub2>), where multicollinearity can be detected with checking tolerance value.

**License** GPL (>= 2)  
**Imports** Rcpp (>= 0.12.13)  
**LinkingTo** Rcpp,RcppEigen  
**Depends** R (>= 2.10)  
**NeedsCompilation** yes  
**Repository** CRAN  
**Date/Publication** 2020-03-23 15:30:05 UTC

## R topics documented:

StepReg-package . . . . .	2
bestsubset . . . . .	4
ModelFitStat . . . . .	6
scoretest . . . . .	7
stepOne . . . . .	8
stepwise . . . . .	10
stepwiselogit . . . . .	13
<b>Index</b>	<b>15</b>

---

StepReg-package	<i>Stepwise Regression Analysis</i>
-----------------	-------------------------------------

---

## Description

Stepwise regression analysis for variable selection can be used to get the best candidate final regression model with the forward selection, backward elimination and bidirectional elimination approaches. Best subset selection fit a separate least squares regression for each possible combination of all predictors. Both the above two procedures in this package can use weighted data to get best regression model in univariate regression and multivariate regression analysis(Alsubaihi, A. A., (2002) <doi:10.18637/jss.v007.i12>). And continuous variables nested within class effect is also considered in both two procedures. Also stepwise logistic regression in this package can performed with binary dependent variable(Agresti, A. (1984) <doi:10.1002/9780470594001> and Agresti, A. (2014) <doi:10.1007/978-3-642-04898-2\_161>). A widely used selection criteria are available which includes Akaike information criterion(Darlington, R. B. (1968) <doi:10.1037/h0025471>, Judge, G. G. (1985) <doi:10.2307/1391738>), corrected Akaike information criterion(Hurvich, C. M., and Tsai, C. (1989) <doi:10.1093/biomet/76.2.297>), Bayesian information criterion(Sawa, T. (1978) <doi:10.2307/1913828>, Judge, G. G. (1985) <doi:10.2307/1391738>), Mallows Cp statistic(Mallows, C. L. (1973) <doi:10.1080/00401706.1995.10484370>, Hocking, R. R. (1976) <doi:10.2307/2529336>), Hannan and Quinn information criterion(Hannan, E. J. and Quinn, B. G. (1979) <doi:10.1111/j.2517-6161.1979.tb01072.x>, Mcquarrie, A. D. R. and Tsai, C. L. (1998) <doi:10.1142/3573>), corrected Hannan and Quinn information criterion(Mcquarrie, A. D. R. and Tsai, C. L. (1998) <doi:10.1142/3573>), Schwarz criterion(Schwarz, G. (1978) <doi:10.1214/aos/1176344136>, Judge, G. G. (1985) <doi:10.2307/1391738>), adjusted R-square statistic(Darlington, R. B. (1968) <doi:10.1037/h0025471>, Judge, G. G. (1985) <doi:10.2307/1391738>) and significance levels(McKeon,

J. J. (1974) <doi:10.1093/biomet/61.2.381>, Harold Hotelling. (1992) <doi:10.1007/978-1-4612-0919-5\_4>, Pillai, K. C. S. (2006) <doi:10.1002/0471667196.ess1965.pub2>), where multicollinearity can be detected with checking tolerance value.

## Details

Package: StepReg  
Type: Package  
Version: 1.4.1  
Date: 2020-3-23  
License: GPL (>= 2)

## Author(s)

Junhui Li, Xiaohuan Lu, Kun Cheng, Wenxin Liu

Maintainer: Junhui Li <junhuili@cau.edu.cn>

## References

- Agresti, A. (1984), Analysis of Ordinal Categorical Data, New York: John Wiley & Sons.
- Agresti, A. (2014), Categorical Data Analysis, New York: John Wiley & Sons.
- Alsubaihi, A. A., Leeuw, J. D., and Zeileis, A. (2002). Variable selection in multivariable regression using sas/iml. , 07(i12).
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69(3), 161.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society, 41(2), 190-195.
- Harold Hotelling. (1992). The Generalization of Student's Ratio. Breakthroughs in Statistics. Springer New York.
- Hocking, R. R. (1976). A biometrics invited paper. the analysis and selection of variables in linear regression. Biometrics, 32(1), 1-49.
- Hosmer, D. W., Jr. and Lemeshow, S. (2000), Applied Logistic Regression, 2nd Edition, New York: John Wiley & Sons.
- Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.
- Judge, & George G. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.
- Lovison, G. . (2005). On rao score and pearson x2 statistics in generalized linear models. Statistical Papers, 46(4), 555-574.
- Mallows, C. L. (1973). Some comments on cp. Technometrics, 15(4), 661-676.

- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. *Mathematical Gazette*, 37(1), 123-131.
- Mckeeon, J. J. (1974). F approximations to the distribution of hotelling's t20. *Biometrika*, 61(2), 381-383.
- Mcquarrie, A. D. R., & Tsai, C. L. (1998). *Regression and Time Series Model Selection. Regression and time series model selection /*. World Scientific.
- Pillai, K. C. S. (2006). Pillai's Trace. *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc.
- Robert Gilchrist. (1982). Glim 82: proceedings of the international conference on generalised linear models. *Lecture Notes in Statistics*, 14(9), 3008-11.
- R.S. Sparks, W. Zucchini, & D. Coutourides. (1985). On variable selection in multivariate regression. *Communication in Statistics- Theory and Methods*, 14(7), 1569-1587.
- Sawa, T. (1978). Information criteria for discriminating among alternative regression models. *Econometrica*, 46(6), 1273-1291.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), pages. 15-18.
- Smyth, G. K. . (2003). Pearson's goodness of fit statistic as a score test statistic. *Lecture Notes-Monograph Series*, 40, 115-126.

---

bestsubset

*Best subset selection*

---

## Description

This function uses information criteria to find a specified number of best models containing one, two, or three variables, and so on, up to the single model containing all of the explanatory variables.

## Usage

```
bestsubset(data, y, exclude = NULL, include = NULL, Class = NULL,
weights = c(rep(1, nrow(data))),select = "SBC", tolerance = 1e-07, best = 5)
```

## Arguments

data	Data set including dependent and independent variables to be analyzed
y	Numeric or character vector for dependent variables
exclude	Numeric or character vector for independent variables removed from stepwise regression analysis
include	Forces 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
Class	Class effect variable
weights	The weights names numeric vector to provide a weight for each observation in the input data set. And 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.

select	Specifies the criterion that uses to calculate all models including Akaike Information Criterion(AIC), the Corrected form of Akaike Information Criterion(AICc), Bayesian Information Criterion(BIC), Schwarz criterion(SBC), Hannan and Quinn Information Criterion(HQ), R-square statistic(Rsq), adjusted R-square statistic(adjRsq) and Mallows Cp statistic(CP)
tolerance	Tolerance value for multicollinearity, default is 1e-7
best	Controls the number of models displayed in the output, default is 5

### Author(s)

Junhui Li

### References

- Alsubaihi, A. A., Leeuw, J. D., and Zeileis, A. (2002). Variable selection in multivariable regression using sas/iml. , 07(i12).
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69(3), 161.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, 41(2), 190-195.
- Harold Hotelling. (1992). *The Generalization of Student's Ratio. Breakthroughs in Statistics.* Springer New York.
- Hocking, R. R. (1976). A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1-49.
- Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297-307.
- Judge, & George G. (1985). *The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /.* Wiley.
- Mallows, C. L. (1973). Some comments on cp. *Technometrics*, 15(4), 661-676.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. *Mathematical Gazette*, 37(1), 123-131.
- Mckean, J. J. (1974). F approximations to the distribution of hotelling's t20. *Biometrika*, 61(2), 381-383.
- Mcquarrie, A. D. R., & Tsai, C. L. (1998). *Regression and Time Series Model Selection. Regression and time series model selection /.* World Scientific.
- Pillai, K. C. S. (2006). Pillai's Trace. *Encyclopedia of Statistical Sciences.* John Wiley & Sons, Inc.
- R.S. Sparks, W. Zucchini, & D. Coutourides. (1985). On variable selection in multivariate regression. *Communication in Statistics- Theory and Methods*, 14(7), 1569-1587.
- Sawa, T. (1978). Information criteria for discriminating among alternative regression models. *Econometrica*, 46(6), 1273-1291.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 15-18.

**Examples**

```

set.seed(4)
dfY <- data.frame(matrix(c(rnorm(20,0,2),c(rep(1,10),rep(2,10))),rnorm(20,2,3)),20,3))
colnames(dfY) <- paste("Y",1:3,sep="")
dfX <- data.frame(matrix(c(rnorm(100,0,2),rnorm(100,2,1)),20,10))
colnames(dfX) <- paste("X",1:10,sep="")
yx <- cbind(dfY,dfX)
bestsubset(yx,y="Y1",exclude="Y3",include="Y2",Class="Y2",
weights=c(rep(0.5,2),rep(1,18)),select="SBC",tolerance=1e-7,best=5)

```

---

ModelFitStat

*Compute model fit statistics*


---

**Description**

Compute model fit statistics adjRsq, AIC, AICc, BIC, CP, HQ, HQc, Rsq and SBC

**Usage**

```
ModelFitStat(Stattype, SSE, SST, n, nY, p, sigmaVal)
```

**Arguments**

Stattype	Model fit statistics adjRsq, AIC, AICc, BIC, CP, HQ, HQc, Rsq and SBC
SSE	Sum of squares of error
SST	Total sum of squares corrected for the mean for the dependent variable
n	Number of observation
nY	Number of dependent variable
p	Number of independent variable in the model including the intercept
sigmaVal	Estimate of pure error variance from fitting the full model

**Author(s)**

Junhui Li Xiaohuan Lu

**References**

Alsubaihi, A. A., Leeuw, J. D., and Zeileis, A. (2002). Variable selection in multivariable regression using sas/iml. , 07(i12).

Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69(3), 161.

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society, 41(2), 190-195.

Harold Hotelling. (1992). The Generalization of Student's Ratio. Breakthroughs in Statistics. Springer New York.

- Hocking, R. R. (1976). A biometrics invited paper. the analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1-49.
- Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297-307.
- Judge, & GeorgeG. (1985). *The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /*. Wiley.
- Mallows, C. L. (1973). Some comments on cp. *Technometrics*, 15(4), 661-676.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. *Mathematical Gazette*, 37(1), 123-131.
- Mckeon, J. J. (1974). F approximations to the distribution of hotelling's t20. *Biometrika*, 61(2), 381-383.
- Mcquarrie, A. D. R., & Tsai, C. L. (1998). *Regression and Time Series Model Selection. Regression and time series model selection /*. World Scientific.
- Pillai, K. C. S. (2006). Pillai's Trace. *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc.
- R.S. Sparks, W. Zucchini, & D. Coutsourides. (1985). On variable selection in multivariate regression. *Communication in Statistics- Theory and Methods*, 14(7), 1569-1587.
- Sawa, T. (1978). Information criteria for discriminating among alternative regression models. *Econometrica*, 46(6), 1273-1291.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), pags. 15-18.

---

scoretest

*Compute score test statistics and probability value for generalized linear model*

---

### Description

This function can compute score test statistics and probability value for linear model by adding an independent variable.

### Usage

```
scoretest(model, x)
```

### Arguments

model	Generalized linear model object
x	Vector or matrix of independent variable

### Author(s)

Junhui Li

## References

- Lovison, G. . (2005). On rao score and pearson x2 statistics in generalized linear models. *Statistical Papers*, 46(4), 555-574.
- Robert Gilchrist. (1982). *Glim 82: proceedings of the international conference on generalised linear models*. *Lecture Notes in Statistics*, 14(9), 3008-11.
- Smyth, G. K. . (2003). Pearson's goodness of fit statistic as a score test statistic. *Lecture Notes-Monograph Series*, 40, 115-126.

## Examples

```
set.seed(1)
yd <- data.frame(sample(c(0,1),30,replace=TRUE))
colnames(yd) <- "remiss"
set.seed(4)
xd <- data.frame(matrix(c(round(rnorm(100,0,2),2),round(rnorm(140,2,4),2),
sample(c(1,0),30,replace=TRUE),sample(1:80,30,replace=TRUE)),30,10))
colnames(xd) <- c(paste("X",1:8,sep=""),"gender","age")
yx <- cbind(yd,xd)
y <- "remiss"
comVar <- paste("X",1:3,sep="")
fm <- paste(y,"~",paste0(comVar,collapse = "+"),sep="")
model <- glm(fm,yx,family="binomial")
scoretest(model, as.matrix(yx[,4]))
```

---

stepOne

*Choose the best model in one step*

---

## Description

Choose the best model with minimal information criteria statistics in forward selection or maximal ones in backward elimination

## Usage

```
stepOne(findIn, p, n, sigma, tolerance, Ftrace, criteria, Y,X1, X0, k, SST)
```

## Arguments

findIn	Logical value, if FALSE then add independent variable to regression model, otherwise remove independent variable from regression model
p	The number of independent variable entered in regression
n	The sample size
sigma	Pure error variance from full regressoin model for Bayesian information criterion(BIC)
tolerance	Tolerance value for multicollinearity



Ftrace	Statistic of multivariate regression including Wilks' lambda, Pillai trace and Hotelling-lawley trace
criteria	Information criterion including AIC, AICc, BIC, SBC, HQ, HQc and SL
Y	Data set for dependent variable
X1	Data set for independent variables not in regression model
X0	Data set for independent variables entered in regression model
k	Forces the first k effects entered in regression model, and the selection methods are performed on the other effects in the data set
SST	Total sum of squares corrected for the mean for the dependent variable

### Details

This function can compute probability value or information criteria statistics with multivariate and univariate regression using least square method

### Value

PIC	P value or Information Criteria statistic value
SEQ	Pointer for independent variable enter or eliminate
SSE	Maximum or minimum of SSE
RkCh	Rank changed or not

### Author(s)

Junhui Li

### References

- Alsubaihi, A. A., Leeuw, J. D., and Zeileis, A. (2002). Variable selection in multivariable regression using sas/iml. , 07(i12).
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69(3), 161.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society*, 41(2), 190-195.
- Harold Hotelling. (1992). *The Generalization of Student's Ratio. Breakthroughs in Statistics.* Springer New York.
- Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. *Biometrika*, 76(2), 297-307.
- Judge, & GeorgeG. (1985). *The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /.* Wiley.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. *Mathematical Gazette*, 37(1), 123-131.
- Mckeon, J. J. (1974). F approximations to the distribution of hotelling's t20. *Biometrika*, 61(2), 381-383.

Mcquarrie, A. D. R., & Tsai, C. L. (1998). Regression and Time Series Model Selection. Regression and time series model selection /. World Scientific.

Pillai, K. C. S. (2006). Pillai's Trace. Encyclopedia of Statistical Sciences. John Wiley & Sons, Inc.

R.S. Sparks, W. Zucchini, & D. Coutsourides. (1985). On variable selection in multivariate regression. Communication in Statistics- Theory and Methods, 14(7), 1569-1587.

Sawa, T. (1978). Information criteria for discriminating among alternative regression models. Econometrica, 46(6), 1273-1291.

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics, 6(2), pages. 15-18.

---

stepwise

*Stepwise Regression*

---

### Description

Stepwise regression analysis can be performed with univariate and multivariate based on information criteria specified, which includes 'forward', 'backward' and 'bidirection' direction model selection method. Also continuous variables nested within class effect and weighted stepwise are considered.

### Usage

```
stepwise(data, y, exclude, include, Class, weights, selection,
         select, sle, sls, tolerance, Trace, Choose)
```

### Arguments

data	Data set including dependent and independent variables to be analyzed
y	Numeric or character vector for dependent variables
exclude	Numeric or character vector for independent variables removed from stepwise regression analysis
include	Forces 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
Class	Class effect variable
weights	The weights names numeric vector to provide a weight for each observation in the input data set. And 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.
selection	Model selection method including "forward", "backward" and "bidirection", 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

select	Specifies the criterion that uses to determine the order in which effects enter and/or leave at each step of the specified selection method including Akaike Information Criterion(AIC), the Corrected form of Akaike Information Criterion(AICc),Bayesian Information Criterion(BIC),Schwarz criterion(SBC),Hannan and Quinn Information Criterion(HQ), R-square statistic(Rsq), adjusted R-square statistic(adjRsqr), Mallows Cp statistic(CP) and Significant Levels(SL)
sle	Specifies the significance level for entry
sls	Specifies the significance level for staying in the model
tolerance	Tolerance value for multicollinearity, default is 1e-7
Trace	Statistic for multivariate regression analysis, including Wilks' lamda ("Wilks"), Pillai Trace ("Pillai") and Hotelling-Lawley's Trace ("Hotelling")
Choose	Chooses from the list of models at the steps of the selection process the model that yields the best value of the specified criterion. If the optimal value of the specified criterion occurs for models at more than one step, then the model with the smallest number of parameters is chosen. Choose method includes AIC, AICc, BIC, HQ, HQc, SBC,Rsq, adjRsqr, CP and NULL, if you do not specify the Choose option, then the model selected is the model at the final step in the selection process

### Details

Multivariate regression and univariate regression can be detected by parameter 'y', where numbers of elements in 'y' is more than 1, then multivariate regression is carried out otherwise univariate regression

### Author(s)

Junhui Li

### References

- Alsubaihi, A. A., Leeuw, J. D., and Zeileis, A. (2002). Variable selection in multivariable regression using sas/iml. , 07(i12).
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. Psychological Bulletin, 69(3), 161.
- Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. Journal of the Royal Statistical Society, 41(2), 190-195.
- Harold Hotelling. (1992). The Generalization of Student's Ratio. Breakthroughs in Statistics. Springer New York.
- Hocking, R. R. (1976). A biometrics invited paper. the analysis and selection of variables in linear regression. Biometrics, 32(1), 1-49.
- Hurvich, C. M., & Tsai, C. (1989). Regression and time series model selection in small samples. Biometrika, 76(2), 297-307.
- Judge, & GeorgeG. (1985). The Theory and practice of econometrics /-2nd ed. The Theory and practice of econometrics /. Wiley.

- Mallows, C. L. (1973). Some comments on cp. *Technometrics*, 15(4), 661-676.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate analysis. *Mathematical Gazette*, 37(1), 123-131.
- Mckeon, J. J. (1974). F approximations to the distribution of hotelling's t20. *Biometrika*, 61(2), 381-383.
- Mcquarrie, A. D. R., & Tsai, C. L. (1998). *Regression and Time Series Model Selection. Regression and time series model selection /*. World Scientific.
- Pillai, K. C. S. (2006). Pillai's Trace. *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc.
- R.S. Sparks, W. Zucchini, & D. Coutourides. (1985). On variable selection in multivariate regression. *Communication in Statistics- Theory and Methods*, 14(7), 1569-1587.
- Sawa, T. (1978). Information criteria for discriminating among alternative regression models. *Econometrica*, 46(6), 1273-1291.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), pags. 15-18.

## Examples

```

set.seed(4)
dfY <- data.frame(matrix(c(rnorm(20,0,2),c(rep(1,10),rep(2,10))),rnorm(20,2,3)),20,3))
colnames(dfY) <- paste("Y",1:3,sep="")
dfX <- data.frame(matrix(c(rnorm(100,0,2),rnorm(100,2,1)),20,10))
colnames(dfX) <- paste("X",1:10,sep="")
yx <- cbind(dfY,dfX)

tol <- 1e-7
Trace <- "Pillai"
sle <- 0.15
sls <- 0.15
# weights vector
w0 <- c(rep(0.5,2),rep(1,18))
w2 <- c(rep(0.5,3),rep(1,14),0.5,1,0.5)

#univariate regression with select = 'SBC' & choose = 'AIC' and select = 'CP' & choose = NULL
#without forced effect and continuous variable nested in class effect

stepwise(yx, y="Y1", exclude="Y3", include=NULL, Class=NULL,w0,
selection="backward", select="SBC", sle, sls, tol, Trace, Choose='AIC')
stepwise(yx, y="Y1",exclude="Y3", include=NULL, Class=NULL, w0,
selection="bidirection", select="CP", sle, sls, tol, Trace, NULL)

#univariate regression with select='AICc' & choose='HQc' and select='BIC' & choose = NULL
#with forced effect and continuous variable nested in class effect
stepwise(yx, y="Y1", exclude="Y3", include="Y2", Class="Y2", w2,
selection="forward", select='AICc', sle, sls, tol, Trace, 'HQc')
stepwise(yx, y="Y1", exclude="Y3", include="Y2", Class="Y2", w2,
selection="bidirection", 'BIC', sle, sls, tol, Trace, NULL)

#multivariate regression with select='HQ' & choose='BIC'
#with forced effect and continuous variable nested in class effect

```

```
stepwise(yx, y=c("Y1","Y3"), exclude=NULL, include="Y2", Class="Y2", w2,
selection="bidirection", select='HQ', sle, sls, tol, Trace, 'BIC')
```

---

stepwiselogit

*Stepwise Logistic Regression*


---

## Description

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

```
stepwiselogit(data, y, exclude = NULL, include = NULL, selection = "bidirection",
select = "SL", sle = 0.15, sls = 0.15)
```

## Arguments

data	Data set including dependent and independent variables to be analyzed
y	Numeric or character vector for dependent variables
exclude	Numeric or character vector for independent variables removed from stepwise regression analysis
include	Forces 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 uses information criteria to find a specified number of best models containing one, two, or three variables, and so on, up to the single model containing all of the explanatory variables.
select	Specifies the criterion that uses to determine the order in which effects enter and leave at each step of the specified selection method including Akaike Information Criterion(AIC), the Corrected form of Akaike Information Criterion(AICc), Schwarz criterion(SBC), and Significant Levels(SL)
sle	Specifies the significance level for entry, default is 0.15
sls	Specifies the significance level for staying in the model, default is 0.15

## Value

RegressionModelsSelectedbyInformationCriterion  
summary of regression models selected by information criterion

SummaryOfSelection  
summary of selection process

AnalysisOfMaximumLikelihoodEstimate  
analysis of maximum likelihood estimate for the selected model  
GoodnessOfTeste  
Hosmer and Lemeshow goodness of fit (GOF) test

**Author(s)**

Junhui Li

**References**

Agresti, A. (1984), Analysis of Ordinal Categorical Data, New York: John Wiley & Sons.  
Agresti, A. (2014), Categorical Data Analysis, New York: John Wiley & Sons.  
Hurvich, C. M. and Tsai, C.-L. (1993), A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection, Journal of Time Series Analysis, 14, 271-279  
Hosmer, D. W., Jr. and Lemeshow, S. (2000), Applied Logistic Regression, 2nd Edition, New York: John Wiley & Sons.

**Examples**

```
set.seed(1)
yd <- data.frame(sample(c(0,1),30,replace=TRUE))
colnames(yd) <- "remiss"
set.seed(4)
xd <- data.frame(matrix(c(round(rnorm(100,0,2),2),round(rnorm(140,2,4),2),
sample(c(1,0),30,replace=TRUE),sample(1:80,30,replace=TRUE)),30,10))
colnames(xd) <- c(paste("X",1:8,sep=""),"gender","age")
yx <- cbind(yd,xd)
y <- "remiss"
stepwiselogit(yx,y,selection="bidirection",select="IC(3/2)")
```

# Index

- \*Topic **best subset selection**
    - bestsubset, 4
  - \*Topic **model fit statistics**
    - ModelFitStat, 6
  - \*Topic **package**
    - StepReg-package, 2
  - \*Topic **score test**
    - scoretest, 7
  - \*Topic **stepwise logistic regression**
    - stepwiselogit, 13
  - \*Topic **stepwise regression**
    - stepOne, 8
    - stepwise, 10
- bestsubset, 4
- ModelFitStat, 6
- scoretest, 7
- stepOne, 8
- StepReg (StepReg-package), 2
- StepReg-package, 2
- stepwise, 10
- stepwiselogit, 13