Package ‘forward’

July 25, 2018

Version 1.0.4
Date 2018-07-24
Title Robust Analysis using Forward Search
Author Originally written for S-Plus by: Kjell Konis
<kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>
Ken Beath <ken.beath@mq.edu.au>
Maintainer Ken Beath <ken.beath@mq.edu.au>
Depends R (>= 3.5), MASS
Description Robust analysis using forward search in linear and
generalized linear regression models, as described in Atkinson, A.C. and Riani, M. (2000), Ro-
License GPL-2
NeedsCompilation no
Repository CRAN
Date/Publication 2018-07-25 07:20:08 UTC

R topics documented:

ar .............................................................. 2
bliss ........................................................... 3
calcium ....................................................... 3
carinsuk .................................................... 4
carr .......................................................... 4
cellular ....................................................... 5
chapman ..................................................... 6
derailme ...................................................... 6
dialectric ..................................................... 7
fwd.combn .................................................. 8
fwdglm ....................................................... 9
fwdlm ....................................................... 11
fwdscuo .................................................... 13
The ar data frame has 60 rows and 4 columns.

Usage

data(ar)

Format

This data frame contains the following columns:

- **x1** a numeric vector
- **x2** a numeric vector
- **x3** a numeric vector
- **y** a numeric vector
References


---

**bliss**

*Bliiss data*

Description

The bliss data frame has 8 rows and 4 columns.

Usage

data(bliss)

Format

This data frame contains the following columns:

- **Dose**: a numeric vector
- **Killed**: a numeric vector
- **Total**: a numeric vector
- **y**: a numeric vector

References


---

**calcium**

*Calcium data*

Description

Calcium uptake of cells suspended in a solution of radioactive calcium.
The calcium data frame has 27 rows and 2 columns.

Usage

data(calcium)

Format

This data frame contains the following columns:

- **Time**: a numeric vector
- **y**: a numeric vector
References

---
carinsuk

*Car insurance data*

**Description**
The *carinsuk* data frame has 128 rows and 5 columns.

**Usage**
```r
data(carinsuk)
```

**Format**
This data frame contains the following columns:

- **OwnerAge** a factor with levels: 17–20, 21–24, 25–29, 30–34, 35–39, 40–49, 50–59, 60+
- **Model** a factor with levels: A, B, C, D
- **CarAge** a factor with levels: 0–3, 10+, 4–7, 8–9
- **NClaims** a numeric vector
- **AvCost** a numeric vector

**References**

---
carr

*n-Pentane data*

**Description**
Reaction rate for Catalytic Isomerization of n-Pentane to Isopentane
The *carr* data frame has 24 rows and 4 columns.

**Usage**
```r
data(carr)
```


**cellular**

**Format**

This data frame contains the following columns:

- **hydrogen** partial pressure of hydrogen
- **npentane** partial pressure of n-pentane
- **isopentane** partial pressure of iso-pentane
- **rate** rate of disappearance of n-pentane

**References**


---

**cellular**

*Cellular differentiation data*

---

**Description**

The cellular data frame has 16 rows and 3 columns.

**Usage**

data(cellular)

**Format**

This data frame contains the following columns:

- **TNF** Dose of TNF (U/ml)
- **IFN** Dose of IFN (U/ml)
- **y** Number of cells differentiating

**References**

chapman  

**Chapman data**

**Description**

The chapman data frame has 200 rows and 7 columns.

**Usage**

```r
data(chapman)
```

**Format**

This data frame contains the following columns:

- **age** a numeric vector
- **highbp** a numeric vector
- **lowbp** a numeric vector
- **chol** a numeric vector
- **height** a numeric vector
- **weight** a numeric vector
- **y** a numeric vector

**References**


---

derailme  

**British Train Accidents.**

**Description**

These data are obtained from Atkinson and Riani (2000), which is a simplified version of the data in Evans (2000). The outcome is the number of deaths that occurred in a train accident with a categorical covariate describing the type of rolling stock, and an exposure variable giving the annual distance travelled by trains in that year, and was originally analysed using a Poisson model. As the data does not include observations with zero deaths, it will be analysed here as a zero-truncated Poisson with an offset of log of the train distance. The derailme data frame has 67 rows and 5 columns.

**Usage**

```r
data(derailme)
```
\textit{dialectric}

\textbf{Format}

This data frame contains the following columns:

\begin{itemize}
  \item \textbf{Month}  Month of accident
  \item \textbf{Year}  Year of accident
  \item \textbf{Type}  Type of rolling stock 1=Mark 1 train, 2=Post-Mark 1 train, 3=Non-passenger
  \item \textbf{TrainKm}  Amount of traffic on the railway system (billions of train km)
  \item \textbf{y}  Number of deaths that occurred in the train accident
\end{itemize}

\textbf{Source}

Atkinson and Riani (2000)

\textbf{References}


\noindent
\begin{tabular}{ll}
\textbf{dialectric} & \textit{Dialectric data} \\
\end{tabular}

\textbf{Description}

The \textit{dialectric} data frame has 128 rows and 3 columns.

\textbf{Usage}

data(dialectric)

\textbf{Format}

This data frame contains the following columns:

\begin{itemize}
  \item \textbf{time}  Time (weeks)
  \item \textbf{temp}  Temperature (degrees Celsius)
  \item \textbf{y}  dialectric breakdown strength in kilovolts
\end{itemize}

\textbf{References}

fwd.combn

Generate all combinations of elements of x taken m at a time

Description

Generate all combinations of the elements of x taken m at a time. If x is a positive integer, returns
all combinations of the elements of seq(x) taken m at a time. If argument fun is not null, applies
a function given by the argument to each point. If simplify is FALSE, returns a list; else returns a
vector or an array. Optional arguments ... are passed unchanged to the function given by argument
fun, if any.

Usage

fwd.combn(x, m, fun = NULL, simplify = TRUE, ...)
fwd.nCm(n, m, tol = 1e-08)

Arguments

x a vector or a single value.
n a positive integer.
m a positive integer.
fun a function to be applied to each combination.
simplify logical, if TRUE returns a vector or an array, otherwise a list.
tol optional, tolerance value.
... optional arguments passed to fun.

Value

Returns a vector or an array if simplify = TRUE, otherwise a list.

Note

Renamed by Kjell Konis for inclusion in the Forward Library 11/2002

Author(s)

Scott Chasalow

References

Press.
Examples

fwd.combn(letters[1:4], 2)
fwd.combn(10, 5, min)  # minimum value in each combination
# Different way of encoding points:
fwd.combn(c(1,1,1,2,2,2,3,3,4), 3, tabulate, nbins = 4)
# Compute support points and (scaled) probabilities for a
# Multivariate-Hypergeometric(n = 3, N = c(4,3,2,1)) p.f.:
table(t(fwd.combn(c(1,1,1,2,2,2,3,3,4), 3, tabulate, nbins=4)))

Description

This function applies the forward search approach to robust analysis in generalized linear models.

Usage

fwdglm(formula, family, data, weights, na.action, contrasts = NULL, bsb = NULL,
        balanced = TRUE, maxit = 50, epsilon = 1e-06, nsamp = 100, trace = TRUE)

Arguments

formula a symbolic description of the model to be fit. The details of the model are the
         same as for glm.
family a description of the error distribution and link function to be used in the model.
        See family for details.
data an optional data frame containing the variables in the model. By default the
     variables are taken from the environment from which the function is called.
weights an optional vector of weights to be used in the fitting process.
na.action a function which indicates what should happen when the data contain NA's. The
         default is set by the na.action setting of options, and is na.fail if that is
         unset. The default is na.omit.
contrasts an optional list. See the contrasts.arg of model.matrix.default.
bsb an optional vector specifying a starting subset of observations to be used in
     the forward search. By default the "best" starting subset is chosen using the
     function lmsglm with control arguments provided by nsamp.
balanced logical, for a binary response if TRUE the proportion of successes on the full
     dataset is approximately balanced during the forward search algorithm.
maxit integer giving the maximal number of IWLS iterations. See glm.control for
details.
epsilon positive convergence tolerance epsilon. See glm.control for details.
the initial subset for the forward search in generalized linear models is found by the function \texttt{lmsglm}. This argument allows to control how many subsets are used in the robust fitting procedure. The choices are: the number of samples (100 by the default) or “all”. Note that the algorithm tries to find \texttt{nsamp} good subsets or a maximum of 2*\texttt{nsamp} subsets.

\texttt{trace} logical, if TRUE a message is printed for every ten iterations completed during the forward search.

\textbf{Value}

The function returns an object of class "\texttt{fwdglm}" with the following components:

- \texttt{call} the matched call.
- \texttt{Residuals} a (\(nx(n - p + 1)\)) matrix of residuals.
- \texttt{Unit} a matrix of units added (to a maximum of 5 units) at each step.
- \texttt{included} a list with each element containing a vector of units included at each step of the forward search.
- \texttt{Coefficients} a ((\(n - p + 1\))xp) matrix of coefficients.
- \texttt{tStatistics} a ((\(n-p+1\))xp) matrix of t statistics for the coefficients, i.e. coef.est/SE(coef.est).
- \texttt{Leverage} a (\(nx(n - p + 1)\)) matrix of leverage values.
- \texttt{MaxRes} a (\((n - p)x2\)) matrix of max deviance residuals in the best subsets and \(m\)-th deviance residuals.
- \texttt{MinDelRes} a (\((n - p - 1)x2\)) matrix of minimum deviance residuals out of best subsets and \((m + 1)\)-th deviance residuals.
- \texttt{ScoreTest} a ((\(n - p\))x1) matrix of score test statistics for a goodness of link test.
- \texttt{Likelihood} a (\((n - p)\)x4) matrix with columns containing: deviance, residual deviance, psuedo \(R^2\) (computed as \(1 - \text{deviance}/\text{null.deviance}\)), dispersion parameter (computed as \(\sum(\text{pearson.residuals}^2)/(m - p)\)).
- \texttt{CookDist} a (\((n - p)x1\)) matrix of forward Cook’s distances.
- \texttt{ModCookDist} a (\((n - p)x5\)) matrix of forward modified Cook’s distances for the units (to a maximum of 5 units) included at each step.
- \texttt{Weights} a (\(nx(n - p)\)) matrix of weights used at each step of the forward search.
- \texttt{inibsb} a vector giving the best starting subset chosen by \texttt{lmsglm}.
- \texttt{binary.response} logical, equal to TRUE if binary response.

\textbf{Author(s)}

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

\textbf{References}

See Also

summary.fwdglm, plot.fwdglm, fwdlm, fwsco.

Examples

data(cellular)
cellular$TNF <- as.factor(cellular$TNF)
cellular$IFN <- as.factor(cellular$IFN)
mod <- fwdglm(y ~ TNF + IFN, data=cellular, family=poisson(log), nsamp=200)
summary(mod)
## Not run: plot(mod)
plot(mod, 1)
plot(mod, 5)
plot(mod, 6, ylim=c(-3, 20))
plot(mod, 7)
plot(mod, 8)

fwdlm

Forward Search in Linear Regression

Description

This function applies the forward search approach to robust analysis in linear regression models.

Usage

fwdlm(formula, data, nsamp = "best", x = NULL, y = NULL, intercept = TRUE,
na.action, trace = TRUE)

Arguments

formula a symbolic description of the model to be fit. The details of the model are the same as for lm.
data an optional data frame containing the variables in the model. By default the variables are taken from the environment from which the function is called.
nsamp the initial subset for the forward search in linear regression is found by fitting the regression model with the R function lmsreg. This argument allows to control how many subsets are used in the Least Median of Squares regression. The choices are: the number of samples or "best" (the default) or "exact" or "sample". For details see lmsreg.
x A matrix of predictors values (if no formula is provided).
y A vector of response values (if no formula is provided).
intercept Logical for the inclusion of the intercept (if no formula is provided).
na.action a function which indicates what should happen when the data contain NA’s. The default is set by the na.action setting of options, and is na.fail if that is unset. The default is na.omit.
trace logical, if TRUE a message is printed for every ten iterations completed during the forward search.

Value

The function returns an object of class "fwdlm" with the following components:

- **call**: the matched call.
- **Residuals**: a \((n \times (n - p + 1))\) matrix of residuals.
- **Unit**: a matrix of units added (to a maximum of 5 units) at each step.
- **included**: a list with each element containing a vector of units included at each step of the forward search.
- **Coefficients**: a \(((n - p + 1) \times p)\) matrix of coefficients.
- **tStatistics**: a \(((n - p + 1) \times p)\) matrix of t statistics for the coefficients.
- **CookDist**: a \((n - p) \times 1\) matrix of forward Cook's distances.
- **ModCookDist**: a \(((n - p) \times 5)\) matrix of forward modified Cook’s distances for the units (to a maximum of 5 units) included at each step.
- **Leverage**: a \((n \times (n - p + 1))\) matrix of leverage values.
- **S2**: a \(((n - p + 1) \times 2)\) matrix with 1st column containing \(S^2\) and the 2nd column \(R^2\).
- **MaxRes**: a \(((n - p) \times 1)\) matrix of max studentized residuals.
- **MinDelRes**: a \(((n - p - 1) \times 1)\) matrix of minimum deletion residuals.
- **StartingModel**: a "lqs" object providing the the Least Median of Squares regression fit used to select the starting subset.

Author(s)

Originally written for S-Plus by: Kjell Konis
<kkonis@insightful.com> and Marco Riani
<mriani@unipr.it>
Ported to R by Luca Scrucca
<luca@stat.unipg.it>

References


See Also

summary.fwdlm, plot.fwdlm, fwdsco, fwdglm, lmsreg.

Examples

```r
library(MASS)
data(forbes)
plot(forbes, xlab="Boiling point", ylab="Pressure")
mod <- fwdlm(108*log10(pres) ~ bp, data=forbes)
summary(mod)
## Not run: plot(mod)
```
plot(mod, 1)
plot(mod, 6, ylim=c(-3, 1000))

### fwdSCO

**Forward Search Transformation in Linear Regression**

**Description**

This function applies the forward search approach to the Box-Cox transformation of response in linear regression models.

**Usage**

```r
fwdSCO(formula, data, nsamp = "best", lambda = c(-1, -0.5, 0, 0.5, 1), 
        x = NULL, y = NULL, intercept = TRUE, na.action, trace = TRUE)
```

**Arguments**

- **formula**: a symbolic description of the model to be fit. The details of the model are the same as for `lm`.
- **data**: an optional data frame containing the variables in the model. By default the variables are taken from the environment from which the function is called.
- **nsamp**: the initial subset for the forward search in linear regression is found by fitting the regression model with the R function `lmsreg`. This argument allows to control how many subsets are used in the Least Median of Squares regression. The choices are: the number of samples or "best" (the default) or "exact" or "sample". For details see `lmsreg`.
- **lambda**: a vector (or a single numerical value) of lambda values for the response transformation.
- **x**: A matrix of predictors values (if no formula is provided).
- **y**: A vector of response values (if no formula is provided).
- **intercept**: Logical for the inclusion of the intercept (if no formula is provided).
- **na.action**: a function which indicates what should happen when the data contain NA’s. The default is set by the na.action setting of options, and is na.fail if that is unset. The default is na.omit.
- **trace**: logical, if TRUE a message is printed for every ten iterations completed during the forward search.

**Value**

The function returns an object of class "fwdSCO" with the following components:

- **call**: the matched call.
- **Likelihood**: a $((n - p + 1) \times n \cdot \text{lambda})$ matrix of likelihood values.
- **ScoreTest**: a $((n - p + 1) \times n \cdot \text{lambda})$ matrix of score test statistic values.
Unit
Input

x

y

Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

`summary.fwdsco`, `plot.fwdsco`, `fwdlm`, `fwdglm`.

Examples

data(wool)
mod <- fwdco(y ~ x1 + x2 + x3, data = wool)
summary(mod)
plot(mod, plot.mle=FALSE)
plot(mod, plot.Sco=FALSE, plot.Lik=TRUE)

hawkins

<table>
<thead>
<tr>
<th>hawkins</th>
<th>Hawkins' data</th>
</tr>
</thead>
</table>

Description

The hawkins data frame has 128 rows and 9 columns.

Usage

data(hawkins)

Format

This data frame contains the following columns:

- **x1** a numeric vector
- **x2** a numeric vector
- **x3** a numeric vector
- **x4** a numeric vector

- a list with an element for each lambda values. Each element provides a matrix of units added (to a maximum of 5 units) at each step of the forward search.
- a list with \( n, p \) and the vector of lambda values used.
- The design matrix.
- The vector for the response.
Kinetics data

Description

Kinetics data (from Becton-Dickenson)
The kinetics data frame has 19 rows and 5 columns.

Usage

data(kinetics)

Format

This data frame contains the following columns:

- **Substrate**  substrate indicator
- **I0**  Inhibitor concentration
- **I3**  Inhibitor concentration
- **I10**  Inhibitor concentration
- **I30**  Inhibitor concentration
- **y**  initial velocity

References

**lakes** | *Lakes data*
---|---

**Description**

The `lakes` data frame has 29 rows and 3 columns.

**Usage**

```r
data(lakes)
```

**Format**

This data frame contains the following columns:

- **NIN** average influent nitrogenon concentration
- **TW** water retention time
- **TN** mean annual nitrogen concentration

**References**


---

**leafpine** | *Pine data*
---|---

**Description**

The `leafpine` data frame has 70 rows and 3 columns.

**Usage**

```r
data(leafpine)
```

**Format**

This data frame contains the following columns:

- **girth** girth
- **height** height
- **volume** volume

**References**

Description

This function computes the Least Median Square robust fit for generalized linear models using deviance residuals.

Usage

lmsglm(x, y, family, weights, offset, n.samples = 100, max.samples = 200, epsilon = 1e-04, maxit = 50, trace = FALSE)

Arguments

x a matrix or data frame containing the explanatory variables.
y the response: a vector of length the number of rows of x.
family a description of the error distribution and link function to be used in the model. See family for details.
weights an optional vector of weights to be used in the fitting process.
offset optional, a priori known component to be included in the linear predictor during fitting.
n.samples number of good subsets to fit. It can be a numeric value or "all".
max.samples maximal number of subsets to fit. By default is set to twice n.samples.
epsilon positive convergence tolerance epsilon. See glm.control for details.
maxit integer giving the maximal number of IWLS iterations. See glm.control for details.
trace logical, if TRUE a message is printed for every ten iterations completed during the search.

Details

This function is used by fwdglm to select the starting subset for the forward search. For this reason, users do not generally need to use it.

Value

The function returns a list with the following components:

bsb a vector giving the best subset found
dev.res a vector giving the deviance residuals for all the observations
message a short message about the status of the algorithm
model the model provided by glm.fit using the units in the best subset found
Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

fwdglm, fwdlm, lmsreg, fwdso.

mice

Mice data

Description

The mice data frame has 14 rows and 4 columns.

Usage

data(mice)

Format

This data frame contains the following columns:

- **dose** dose level
- **prep** factor preparation: 0 = Standard preparation, 1 = Test preparation
- **conv** number with convolution
- **total** Total

References

molar

<table>
<thead>
<tr>
<th>molar</th>
<th>Molar data</th>
</tr>
</thead>
</table>

**Description**
Radioactivity versus molar concentration of nifedipene
The molar data frame has 15 rows and 2 columns.

**Usage**
data(molar)

**Format**
This data frame contains the following columns:

- **x**: log10(NIF concentration)
- **y**: Total counts for $5 \times 10^{-10}$ Molar NTD additive

**References**

mussels

<table>
<thead>
<tr>
<th>mussels</th>
<th>Mussels data</th>
</tr>
</thead>
</table>

**Description**
The mussels data frame has 82 rows and 5 columns.

**Usage**
data(mussels)

**Format**
This data frame contains the following columns:

- **W**: width
- **H**: height
- **L**: length
- **S**: shell mass
- **M**: mass
References


---

**ozone**

**Ozone data**

---

**Description**

Ozone concentration at Upland, CA.
The ozone data frame has 80 rows and 9 columns.

**Usage**

data(ozone)

**Format**

This data frame contains the following columns:

- **x1** a numeric vector
- **x2** a numeric vector
- **x3** a numeric vector
- **x4** a numeric vector
- **x5** a numeric vector
- **x6** a numeric vector
- **x7** a numeric vector
- **x8** a numeric vector
- **y** Ozone concentration (ppm)

**References**

Generalized Linear Models

Description

This function plots the results of a forward search analysis in generalized linear models.

Usage

```r
## S3 method for class 'fwdglm'
plot(x, which.plots = 1:11, squared = FALSE, scaled = FALSE,
     ylim = NULL, xlim = NULL, th.Res = 4, th.Lev = 0.25,
     sig.Tst = 2.58, sig.score = 1.96, plot.pf = FALSE,
     labels.in.plot = TRUE, ...) # If TRUE plots squared deviance residuals.
```

Arguments

- `x`: a "fwdglm" object.
- `which.plots`: select which plots to draw, by default all. Each graph is addressed by an integer:
  1. deviance residuals
  2. leverages
  3. maximum deviance residuals
  4. minimum deviance residuals
  5. coefficients
  6. t statistics, i.e. coef.est/SE(coef.est)
  7. likelihood matrix: deviance, deviance explained, pseudo R-squared, dispersion parameter
  8. score statistic for the goodness of link test
  9. forward Cook's distances
  10. modified forward Cook's distances
  11. weights used at each step of the forward search for the units included
- `squared`: logical, if TRUE plots squared deviance residuals.
- `scaled`: logical, if TRUE plots scaled coefficient estimates.
- `ylim`: a two component vector for the min and max of the y axis.
- `xlim`: a two component vector for the min and max of the x axis.
- `th.Res`: numerical, a threshold for labelling the residuals.
- `th.Lev`: numerical, a threshold for labelling the leverages.
- `sig.Tst`: numerical, a value used to draw the confidence interval on the plot of the t statistics.
- `sig.score`: numerical, a value used to draw the confidence interval on the plot of the score test statistic.
- `plot.pf`: logical, in case of binary response if TRUE graphs contain all the step of the forward search, otherwise only those in which there is no perfect fit.
- `labels.in.plot`: logical, if TRUE units are labelled in the plots when required.
- `...`: further arguments passed to or from other methods.
Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

fwdglm, fwdlm, fwdsco.

Examples

```r
## Not run:
data(cellular)
mod <- fwdglm(y ~ as.factor(TNF) + as.factor(IFN), data=cellular,
               family=poisson(log), nsamp=200)
summary(mod)
plot(mod)
## End(Not run)
```

plot.fwdlm

Forward Search in Linear Regression

Description

This function plots the results of a forward search analysis in linear regression models.

Usage

```r
## S3 method for class 'fwdlm'
plot(x, which.plot = 1:10, squared = FALSE, scaled = TRUE,
     ylim = NULL, xlim = NULL, th.Res = 2, th.Lev = 0.25, sig.Tst = 2.58,
     labels.in.plot = TRUE, ...)
```

Arguments

- `x`: a "fwdlm" object.
- `which.plot`: select which plots to draw, by default all. Each graph is addressed by an integer:
  1. scaled residuals
  2. leverages
  3. maximum studentized residuals
  4. minimum deletion residuals
plot.fwdlm

5. coefficients
6. statistics
7. forward Cook’s distances
8. modified forward Cook’s distances
9. \(S^2\) values
10. \(R^2\) values

\textbf{squared} logical, if TRUE plots squared residuals.

\textbf{scaled} logical, if TRUE plots scaled coefficient estimates.

\textbf{ylim} a two component vector for the min and max of the y axis.

\textbf{xlim} a two component vector for the min and max of the x axis.

\textbf{th.Res} numerical, a threshold for labelling the residuals.

\textbf{th.Lev} numerical, a threshold for labelling the leverages.

\textbf{sig.Tst} numerical, a value (on the scale of the t statistics) used to draw the confidence interval on the plot of the t statistics.

\textbf{labels.in.plot} logical, if TRUE units are labelled in the plots when required.

\ldots further arguments passed to or from other methods.

\textbf{Author(s)}

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

\textbf{References}


\textbf{See Also}

\texttt{fwdlm, fwdsco, fwdglm}.

\textbf{Examples}

\begin{verbatim}
library(MASS)
data(forbes)
plot(forbes)
mod <- fwdlm(100*log10(pres) ~ bp, data=forbes)
summary(mod)
## Not run: plot(mod)
\end{verbatim}
Description

This function plots the results of a forward search analysis for Box-Cox transformation of response in linear regression models.

Usage

```r
## S3 method for class 'fwdscoco'
plot(x, plot.Sco = TRUE, plot.Lik = FALSE, th.Sco = 2.58,
     plot.mle = TRUE, ylim = NULL, xlim = NULL, ...)
```

Arguments

- `x`: a "fwdscoco" object.
- `plot.Sco`: logical, if TRUE plots the score test statistic at each step of the forward search for each lambda value.
- `plot.Lik`: logical, if TRUE plots the likelihood value at each step of the forward search for each lambda value.
- `th.Sco`: numerical, a value used to draw the confidence interval on the plot of the score test statistic.
- `plot.mle`: logical, if TRUE adds a point at the maximum likelihood value for the transformation computed in the final step, i.e. on the full dataset.
- `ylim`: a two component vector for the min and max of the y axis.
- `xlim`: a two component vector for the min and max of the x axis.
- `...`: further arguments passed to or from other methods.

Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

`fwdscoco, fwdlm, fwdglm`. 

plot.fwdscoco

---

*Forward Search Transformation in Linear Regression*
Examples

```r
## Not run:
data(wool)
mod <- fwdscordinary(x ~ x1 + x2 + x3, data = wool)
plot(mod, plot.mle=FALSE)
plot(mod, plot.Sco=FALSE, plot.Lik=TRUE)
## End(Not run)
```

### poison

**Poison data**

**Description**

Box and Cox poison data. Survival times in 10 hour units of animals in a $3 \times 4$ factorial experiment. The poison data frame has 48 rows and 3 columns.

**Usage**

```r
data(poison)
```

**Format**

This data frame contains the following columns:

- **time**: a numeric vector
- **poison**: a factor
- **treat**: a factor with levels: A, B, C, D

**References**


### rainfall

**Rainfall data**

**Description**

Toxoplasmosis data. The rainfall data frame has 34 rows and 3 columns.

**Usage**

```r
data(rainfall)
```
**salinity**

**Format**

This data frame contains the following columns:

- **Rain**  mm of rain
- **Cases**  cases of toxoplasmosis
- **Total**  total

**References**


### Description

The salinity data frame has 28 rows and 4 columns.

### Usage

data(salinity)

**Format**

This data frame contains the following columns:

- **lagsalinity**  Lagged salinity
- **trend**  Trend
- **waterflow**  Water flow
- **salinity**  Salinity

**References**

scglm

Goodness of Link Test in GLM

Description

Computes the score test statistic for the goodness of link test in generalized linear models.

Usage

scglm(x, y, family, weights, beta, phi = 1, offset)

Arguments

x a matrix or data frame containing the explanatory variables.
y the response: a vector of length the number of rows of x.
family a description of the error distribution and link function to be used in the model. See family for details.
weights an optional vector of weights to be used in the fitting process.
beta a vector of coefficients estimates
phi the dispersion parameter
offset optional, a priori known component to be included in the linear predictor during fitting.

Details


Value

Return the value of the score test statistic.

Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

fwdglm, fwdlm, score.s.
Score test for the Box-Cox transformation of the response

Description

Computes the approximate score test statistic for the Box-Cox transformation

Usage

```r
score.s(x, y, la, tol = 1e-20)
lambda.mle(x, y, init = c(-2, 2), tol = 1e-04)
```

Arguments

- `x`: a matrix or data frame containing the explanatory variables.
- `y`: the response: a vector of length the number of rows of `x`.
- `la`: the value of the lambda parameter.
- `tol`: tolerance value used to check for full rank matrix.
- `init`: range of values to search for MLE.

Details


Value

Return a list with two components:

- `Score`: the value of the score test statistic
- `Likelihood`: the value of the likelihood

Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

`fwdsc`, `fwdlm`, `fwdglm`. 
stackloss

Description

Brownlee’s stack loss data.
The stackloss data frame has 21 rows and 4 columns.

Usage

data(stackloss)

Format

This data frame contains the following columns:

- **Air**: Air flow
- **Temp**: Cooling water inlet temperature
- **Conc**: Acid concentration
- **Loss**: Stack loss

References


summary.fwdglm

Summarying Fit of Forward Search in Generalized Linear Regression

Description

summary method for class "fwdglm".

Usage

```r
## S3 method for class 'fwdglm'
summary(object, steps = "auto", remove.perfect.fit = TRUE, ...)
```

Arguments

- **object**: an object of class "fwdglm".
- **steps**: the number of forward steps to show.
- **remove.perfect.fit**: logical, controlling if perfect fit steps should be removed (only apply to binary responses).
- **...**: further arguments passed to or from other methods.
Author(s)

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

References


See Also

`fwdglm`.

summary.fwdlm

---

**summarizing Fit of Forward Search in Linear Regression**

**description**

summary method for class "fwdlm".

**Usage**

```r
## S3 method for class 'fwdlm'
summary(object, steps = "auto", ...)  
```

**Arguments**

- `object` an object of class "fwdlm".
- `steps` the number of forward steps to show.
- `...` further arguments passed to or from other methods.

**Author(s)**

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

**References**


**See Also**

`fwdglm`. 
Summary}

**Description**

summary method for class "fwd sco".

**Usage**

```r
## S3 method for class 'fwd sco'
summary(object, steps = "auto", lambdaMLE = FALSE, ...)
```

**Arguments**

- `object`: an object of class "fwd sco".
- `steps`: the number of forward steps to show.
- `lambdaMLE`: logical, controlling if the MLE of lambda calculated on the full dataset must be shown.
- `...`: further arguments passed to or from other methods.

**Author(s)**

Originally written for S-Plus by: Kjell Konis <kkonis@insightful.com> and Marco Riani <mriani@unipr.it>
Ported to R by Luca Scrucca <luca@stat.unipg.it>

**References**


**See Also**

- `fwd sco`.

---

**vaso**

**Vaso data**

**Description**

Finney’s data on vaso-con traction in the skin of the digits.
The vaso data frame has 39 rows and 3 columns.

**Usage**

```r
data(vaso)
```
Format

This data frame contains the following columns:

- **volume**: volume
- **rate**: rate
- **y**: response: 0 = nonoccurrence, 1 = occurrence

References


---

wool

**Wool data**

Description

Number of cycles to failure of samples of worsted yarn in a 33 experiment. The wool data frame has 27 rows and 4 columns.

Usage

data(wool)

Format

This data frame contains the following columns:

- **x1**: factor levels: -1, 0, 1
- **x2**: factor levels: -1, 0, 1
- **x3**: factor levels: -1, 0, 1
- **y**: cycles to failure a numeric vector

References

Index

* Topic **datasets**
ar, 2
bliss, 3
calcium, 3
carinsuk, 4
carr, 4
cellular, 5
chapman, 6
derailme, 6
dialectric, 7
hawkins, 14
kinetics, 15
lakes, 16
leafpine, 16
mice, 18
molar, 19
mussels, 19
ozone, 20
poison, 25
rainfall, 25
salinity, 26
stackloss, 29
vaso, 31
wool, 32
* Topic **math**
fwd.combn, 8
* Topic **models**
fwdglm, 9
fwdlm, 11
fwdsco, 13
lmsglm, 17
plot.fwdglm, 21
plot.fwdlm, 22
plot.fwdsco, 24
scglm, 27
score.s, 28
summary.fwdglm, 29
summary.fwdlm, 30
summary.fwdsco, 31
* Topic **regression**
fwdglm, 9
fwdlm, 11
fwdsco, 13
lmsglm, 17
plot.fwdglm, 21
plot.fwdlm, 22
plot.fwdsco, 24
scglm, 27
score.s, 28
summary.fwdglm, 29
summary.fwdlm, 30
summary.fwdsco, 31
* Topic **robust**
fwdglm, 9
fwdlm, 11
fwdsco, 13
lmsglm, 17
plot.fwdglm, 21
plot.fwdlm, 22
plot.fwdsco, 24
scglm, 27
score.s, 28
summary.fwdglm, 29
summary.fwdlm, 30
summary.fwdsco, 31
ar, 2
bliss, 3
calcium, 3
carinsuk, 4
carr, 4
cellular, 5
chapman, 6
derailme, 6
dialectric, 7
family, 9, 27
INDEX

fwd.combn, 8
fwd.nCm (fwd.combn), 8
fwdglm, 9, 12, 14, 18, 22–24, 27, 28, 30
fwdlm, 11, 11, 14, 18, 22–24, 27, 28, 30
fwdsco, 11, 12, 13, 18, 22–24, 28, 31

hawkins, 14

kinetics, 15

lakes, 16
lambda.mle (score.s), 28
leafpine, 16
leafpine, 16
lmsglm, 9, 10, 17
lmsreg, 11–13, 18

mice, 18
molar, 19
mussels, 19

ozone, 20

plot.fwdglm, 11, 21
plot.fwdlm, 12, 22
plot.fwdsco, 14, 24
poison, 25
print.fwdglm (fwdglm), 9
print.fwdlm (fwdlm), 11
print.fwdsco (fwdsco), 13
print.summary.fwdglm (summary.fwdglm), 29
print.summary.fwdlm (summary.fwdlm), 30
print.summary.fwdsco (summary.fwdsco), 31

rainfall, 25

salinity, 26
scglm, 27
score.s, 27, 28
stackloss, 29
summary.fwdglm, 11, 29
summary.fwdlm, 12, 30
summary.fwdsco, 14, 31

vaso, 31

wool, 32