Package: forward (via r-universe)

September 14, 2024

Version 1.0.6

Date 2022-05-08

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@kjbeath.com.au>

Maintainer Ken Beath <ken@kjbeath.com.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), Robust Diagnostic Regression
Analysis, First Edition. New York: Springer.

License GPL-2

NeedsCompilation no

Date/Publication 2022-05-09 07:00:10 UTC

Repository https://kjbeath.r-universe.dev

RemoteUrl https://github.com/cran/forward

RemoteRef HEAD

RemoteSha b4db72ebfc87b882055e1958f4a2d2a6e71664c0

Contents

---

ar	*ar data*

Description

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.2

---

bliss	*Bliss 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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.20

---

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.13

---

carinsuk	*Car insurance data*

Description

The `carinsuk` data frame has 128 rows and 5 columns.

Usage

```
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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.16

---

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

```
data(carr)
```

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.15

---

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.19

---

chapman   *Chapman data*

Description

The chapman data frame has 200 rows and 7 columns.

Usage

```
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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.24

---

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

```
data(derailme)
```

Format

This data frame contains the following columns:

Month Month of accident

Year Year of accident

Type Type of rolling stock 1=Mark 1 train, 2=Post-Mark 1 train, 3=Non-passenger

TrainKm Amount of traffic on the railway system (billions of train km)

y Number of deaths that occurred in the train accident

Source

Atkinson and Riani (2000)

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.18

Evans, A. W. (2000). Fatal train accidents on Britain's mainline railways. Journal Royal Statistical Society A, 163(1), 99-119.

---

dialectric   *Dialectric data*

Description

The `dialectric` data frame has 128 rows and 3 columns.

Usage

```
data(dialectric)
```

Format

This data frame contains the following columns:

time Time (weeks)

temp Temperature (degrees Celsius)

y dialectric breakdown strength in kilovolts

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.17

---

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

Nijenhuis, A. and Wilf, H.S. (1978) *Combinatorial Computers and Calculators*. NY: Academic 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(1,1,1,1,2,2,2,3,3,4), 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(1,1,1,1,2,2,2,3,3,4), 3, tabulate, nbins=4)))
```

---

fwdglm	*Forward Search in Generalized Linear Models*

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.
nsamp	the initial subset for the forward search in generalized linear models is found by the function [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 nsamp good subsets or a maximum of 2*nsamp subsets.
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 "fwdglm" with the following components:

call	the matched call.
Residuals	a $(nx(n - p + 1))$ 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 $((n - p + 1)xp)$ matrix of coefficients.
tStatistics	a $((n-p+1)xp)$ matrix of t statistics for the coefficients, i.e. coef.est/SE(coef.est).
Leverage	a $(nx(n - p + 1))$ matrix of leverage values.
MaxRes	a $((n - p)x2)$ matrix of max deviance residuals in the best subsets and m-th deviance residuals.

MinDelRes	a $((n-p-1)x2)$ matrix of minimum deviance residuals out of best subsets and $(m+1)$-th deviance residuals.
ScoreTest	a $((n-p)x1)$ matrix of score test statistics for a goodness of link test.
Likelihood	a $((n-p)x4)$ matrix with columns containing: deviance, residual deviance, psuedo R^2 (computed as $1 - deviance/null.deviance$), dispersion parameter (computed as $\sum(pearson.residuals^2)/(m-p)$).
CookDist	a $((n-p)x1)$ matrix of forward Cook's distances.
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.
Weights	a $(nx(n-p))$ matrix of weights used at each step of the forward search.
inibsb	a vector giving the best starting subset chosen by lmsglm.
binary.response	
	logical, equal to TRUE if binary response.

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 6.

See Also

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

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapters 2-3.

See Also

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

Examples

```
library(MASS)
data(forbes)
plot(forbes, xlab="Boiling point", ylab="Pressure)")
mod <- fwdlm(100*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

```
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 areused 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 $((n-p+1) x n.lambda)$ matrix of likelihood values.
ScoreTest	a $((n-p+1) x n.lambda)$ matrix of score test statistic values.
Unit	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.
Input	a list with n, p and the vector of lambda values used.
x	The design matrix.
y	The vector for the response.

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 4.

See Also

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

Examples

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

---

hawkins   *Hawkins' data*

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

x5 a numeric vector

x6 a numeric vector

x7 a numeric vector

x8 a numeric vector

y a numeric vector

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.4

`kinetics` *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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.12

`lakes` *Lakes data*

Description

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

Usage

```
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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.14

---

leafpine	*Pine data*

Description

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

Usage

```
data(leafpine)
```

Format

This data frame contains the following columns:

girth girth

height height

volume volume

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.10

---

lmsglm	*Forward Search in Generalized Linear Models*

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 6.

See Also

fwdglm, fwdlm, lmsreg, fwdsco.

---

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 convultion

total Total

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.21

---

molar	*Molar data*

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 5×10^{-10} Molar NTD additive

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.1

---

mussels   *Mussels data*

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.9

---

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.7

---

plot.fwdglm   *Forward Search in Generalized Linear Models*

Description

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

Usage

```
## 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, ...)
```

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 6.

See Also

fwdglm, fwdlm, fwdsco.

Examples

```
## 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

```
## S3 method for class 'fwdlm'
plot(x, which.plots = 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.plots	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
    5. coefficients
    6. statistics
    7. forward Cook's distances
    8. modified forward Cook's distances
    9. S^2 values
    10. R^2 values

squared	logical, if TRUE plots squared 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 (on the scale of the t statistics) used to draw the confidence interval on the plot of the t statistics.
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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapters 2-3.

See Also

[fwdlm](), [fwdsco](), [fwdglm]().

Examples

```
library(MASS)
data(forbes)
plot(forbes)
mod <- fwdlm(100*log10(pres) ~ bp, data=forbes)
summary(mod)
## Not run: plot(mod)
```

---

plot.fwdsco   *Forward Search Transformation in Linear Regression*

Description

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

Usage

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

Arguments

x	a "fwdsco" 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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapters 2-3.

See Also

fwdsco, fwdlm, fwdglm.

Examples

```
## Not run:
data(wool)
mod <- fwdsco(y ~ 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 3×4 factorial experiment. The poison data frame has 48 rows and 3 columns.

Usage

```
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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.8

---

rainfall   *Rainfall data*

Description

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

Usage

```
data(rainfall)
```

Format

This data frame contains the following columns:

Rain mm of rain

Cases cases of toxoplasmosis

Total total

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.22

---

salinity   *Salinity data*

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.6

---

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

See pag. 200–201 of Atkinson and Riani (2000).

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 6.

See Also

fwdglm, fwdlm, score.s.

---

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

```
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

See pag. 82–86 of Atkinson and Riani (2000).

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 4.

See Also

fwdsco, fwdlm, fwdglm.

---

stackloss   *Stackloss data*

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.5

---

summary.fwdglm   *Summarizing Fit of Forward Search in Generalized Linear Regression*

Description

summary method for class `"fwdglm"`.

Usage

```
## 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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 6.

See Also

[fwdglm](#).

---

summary.fwdlm	*Summarizing Fit of Forward Search in Linear Regression*

Description

summary method for class `"fwdlm"`.

Usage

```
## 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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapters 2-3.

See Also

[fwdlm](#).

---

summary.fwdsco	*Summarizing Fit of Forward Search Transformation in Linear Regression*

Description

summary method for class `"fwdsco"`.

Usage

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

Arguments

object	an object of class `"fwdsco"`.
steps	the number of forward steps to show.
lambdaMLE	logical, controlling if the MLE of lambda calculated on the full dataset must be 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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Chapter 4.

See Also

fwdsco.

---

vaso	*Vaso data*

Description

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

Usage

```
data(vaso)
```

Format

This data frame contains the following columns:

volume volume

rate rate

y response: 0= nonoccurrence, 1= occurrence

References

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.23

---

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

Atkinson, A.C. and Riani, M. (2000), *Robust Diagnostic Regression Analysis*, First Edition. New York: Springer, Table A.3

Index