

Package ‘alr3’

September 29, 2009

Version 1.1.12

Date 2009/09/27

Title Methods and data to accompany Applied Linear Regression 3rd edition

Author Sanford Weisberg <sandy@stat.umn.edu>

Maintainer Sanford Weisberg <sandy@stat.umn.edu>

LazyData yes

LazyLoad no

Description This library is a companion to the textbook S. Weisberg (2005), “Applied Linear Regression,” 3rd edition, Wiley. It includes all the data sets discussed in the book (except one), and several few functions that are tailored to the methods discussed in the book. Ver. 1.0.3 corrects bugs in (1) weights for POD models; (2) use of ‘subset’ arguments with boot.case. Ver. 1.0.4 corrects a few bugs with overparameterized models. Ver. 1.0.5 changes one default argument. Ver 1.0.6 replaces the pure.error.anova function so it prints nicer and works with interactions. Ver 1.0.7 fixes a bug in residual.plots and adds an additional argument and adds a new function alrweb to access the website for the book in a browser. Ver 1.0.8 fixed a bug in delta.method for generalized linear models. Ver 1.0.9 fixes a bug with Yeo Johnson transformations. Ver 1.1.0 drops the conf.intervals function in favor of the confint function in stat, includes several minor bug fixes, and makes the package fully compatible with version 2.6.0 of R. 1.1.2 corrects a bug introduced in boot.case. Version 1.1.7 adds a namespace. Ver 1.1.8 fixes more bugs. Ver. 1.1.9 updated the delta.method to work with a user-supplied estimate of the variance matrix. 1.1.10 removed a conflict with another package. Data files lightly modified to make some data columns row labels.

License GPL (>= 2)

URL <http://www.r-project.org>, <http://www.stat.umn.edu/alr>

Repository CRAN

Date/Publication 2009-09-29 19:12:16

R topics documented:

ais	4
allshoots	5
alrweb	5
baeskel	6
banknote	7
bctrans	7
BGSall	10
BigMac2003	11
blowdown	12
brains	12
cakes	13
cathedral	14
caution	14
challeng	15
chloride	16
cloud	16
conf.intervals	17
delta.method	18
domedata	20
donner	21
downer	22
drugcost	23
dwaste	24
florida	25
forbes	25
ftcollinssnow	26
fuel2001	27
galapagos	28
galtonpeas	29
heights	29
highway	30
hooker	31
htwt	32
inf.index	32
inv.res.plot	33
inv.tran.plot	35
jevons	36
lakemary	37
lakes	38
landrent	39
lathe1	39
mantel	40
mile	41
Mitchell	42
mmps	42
MWwords	44

npdata	45
oldfaith	46
outlier.t.test	46
physics	47
pipeline	48
plot.bctrans	49
pod	50
powtran	54
prodscore	55
pure.error.anova	56
random.lin.comb	57
rat	58
residual.plots	59
salary	61
salarygov	61
segreg	62
shocks	63
sigma.hat	64
sleep1	64
snake	65
sniffer	66
snowgeese	67
stopping	67
swan96	68
titanic	69
transact	69
turk0	70
turkey	71
twins	71
ufc	72
UN1	73
UN2	74
UN3	75
walleye	76
water	76
wblake	77
wm1	78
wm2	79
wm3	80
wm4	80
wool	81

ais

Australian institute of sport data

Description

Data on 102 male and 100 female athletes collected at the Australian Institute of Sport.

Format

This data frame contains the following columns:

Sex (0 = male or 1 = female)

Ht height (cm)

Wt weight (kg)

LBM lean body mass

RCC red cell count

WCC white cell count

Hc Hematocrit

Hg Hemoglobin

Ferr plasma ferritin concentration

BMI body mass index, $\text{weight}/(\text{height})^{**2}$

SSF sum of skin folds

Bfat Percent body fat

Label Case Labels

Sport Sport

Source

Ross Cunningham and Richard Telford

References

S. Weisberg (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 6.4

Examples

```
data(ais)
```

allshoots	<i>Apple shoots data</i>
-----------	--------------------------

Description

Bland's Apple Shoot data. allshoots includes all the data, shortshoots just the short shoot data, and longshoots includes long shoots only.

Format

This data frame contains the following columns:

Day days from dormancy

n number of shoots sampled

ybar average number of stem units

SD within-day standard deviation

Type 1 if long shoots, 0 if shortshoots.

Source

Bland, J. (1978). A comparison of certain aspects of ontogeny in the long and short shoots of McIntosh apple during one annual growth cycle. Unpublished Ph. D. dissertation, University of Minnesota, St. Paul, Minnesota.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 5.3.

Examples

```
data(longshoots)
```

alrweb	<i>Access to the Applied Linear Regression website</i>
--------	--

Description

This function will access the website for Applied Linear Regression

Usage

```
alrweb(what = c("webpage", "errata", "primer"))
```

```
alrprimer()
```

```
alrerrata()
```

Arguments

what A character string indicating what page to open. The default "webpage" will open the main webpage, "errata" displays the Errata sheet for the book, and "primer" fetches and displays the primer for R.

Value

Either a webpage or a pdf document is displayed. `alrprimer()` and `alrerrata()` call `alrweb()` with the appropriate arguments.

Author(s)

Sanford Weisberg, based on the function `UsingR` in the `UsingR` package by John Verzani

Examples

```
## Not run: alrweb()
```

baeskel

Surface tension

Description

The data in the file were collected in a study of the effect of dissolved sulfur on the surface tension of liquid copper (Baes and Kellogg, 1953)

Format

This data frame contains the following columns:

Sulfur Weight percent sulfur

Tension Decrease in surface tension, dynes/cm

Source

Baes, C. and Kellogg, H. (1953). Effect of dissolved sulphur on the surface tension of liquid copper. *J. Metals*, 5, 643-648.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 7.4

Examples

```
data(baeskel)
```

`banknote`*Swiss banknote data*

Description

Six measurements made on 100 genuine Swiss banknotes and 100 counterfeit ones.

Format

This data frame contains the following columns:

Length Length of bill, mm

Left Width of left edge, mm

Right Width of right edge, mm

Bottom Bottom margin width, mm

Top Top margin width, mm

Diagonal Length of image diagonal, mm

Y 0 = genuine, 1 = counterfeit

Source

Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: A practical approach*. London: Chapman & Hall.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 12.5.

Examples

```
data(banknote)
```

`bctrans`*Univariate and multivariate transformations to normality*

Description

Estimates multivariate power transformations to multinormality by a maximum likelihood-like method. The univariate case is obtained when only one variable is specified.

Usage

```

bctrans(formula, data = NULL, subset, na.action = na.omit, ...)

## Use when you have a matrix or data.frame:
bctrans1(X, Y = NULL, start = NULL, family = "box.cox", call=NULL, ...)

## S3 methods for class 'bctrans'
lrt.bctrans(object, lrt=NULL, ones=TRUE, zeroes=TRUE)

```

Arguments

formula	A formula, giving the variables to be transformed. The formula can be <i>one-sided</i> , of the form $\sim X_1+X_2+X_3$, or <i>two-sided</i> , of the form $Y\sim X_1+X_2+X_3$. In the latter case, the response is not used in the transformation, but it will be used in the <code>plot</code> method. If you have previously computed a linear model fit, say <code>m1</code> , then you can use <code>m1</code> in place of the formula.
data	a <code>data.frame</code> (or list) from which the variables in the formula should be taken.
subset	an optional vector specifying a subset of observations to be used.
na.action	If set to <code>na.omit</code> , the default, missing values are permitted. If set to <code>na.fail</code> , missing values are not permitted.
...	In <code>bctrans</code> , these are additional arguments passed to <code>bctrans1</code> , and described below. In <code>bctrans1</code> , these are additional arguments passed to the function maximizer <code>optim</code> .
X	A vector, matrix, or data frame whose columns are to be transformed.
Y	If present this vector will be part of the object created, and will be used in drawing plots. It is not used for finding transformations.
start	Starting values for the power transformation parameters; if <code>NULL</code> (the default), univariate transformations will be computed and used as the start values.
family	The family of transformations. The most common is <code>"box.cox"</code> for the Box-Cox transformation. The <code>"yeo.johnson"</code> transformations are used if some elements of <code>X</code> are negative or zero. The family <code>"power"</code> is used only in the <code>plot.bctrans</code> and <code>add.trans</code> functions to give basic power transformations, which can't be normalized to have Jacobian one. The argument <code>lambda</code> indexes the family of transformations $\psi(X, \lambda)$
object	In <code>lrt.bctrans</code> , the name of a 'bctrans' object.
lrt	In the <code>lrt.bctrans</code> command, a list of vectors each of length equal to the number of columns in <code>X</code> . A Likelihood ratio test that the transformation parameters equal each of these vectors will be performed.
ones	In <code>lrt.bctrans</code> , if <code>TRUE</code> test all the transformation parameters equal to one against a general alternative.
zeroes	In <code>lrt.bctrans</code> , if <code>TRUE</code> test all the transformation parameters equal to zero against a general alternative.
call	Not to be set by the user.

Details

Given a matrix X with columns X_1, \dots, X_p , this routine selects transformation parameters $\lambda_1, \dots, \lambda_p$ from a one-parameter family of transformations such that the transformed variables $\psi(X_1, \lambda_1), \dots, \psi(X_p, \lambda_p)$ are as close to multivariate normal as possible.

The function uses the family of transformation you specify. If you use the family `box.cox` to select a transformation, it is usual to use standard power transformations in further calculations.

Value

`bctrans` returns an object of class `bctrans`, which may be printed or summarized. It is a list with components `X` containing the input data, `family` the family used, `start` the starting values, and, `optim`, the results from a call to `optim`, the function optimizer used in the routine.

Author(s)

A substantial part of this code is borrowed from the function `box.cox.powers` in the `car` package, written by John Fox, and documented in Fox (2002). It is based on a similar function in `Arc`; see Cook and Weisberg (1999). It was modified by Sanford Weisberg, sandy@stat.umn.edu and renamed `bctrans`.

References

- Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26 211-46.
- Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. Wiley.
- Fox, J. (2002). *R and S-Plus Companion to Applied Regression*. Sage.
- Velilla, S. (1993). A note on the multivariate Box-Cox transformation to normality. *Statistics and Probability Letters*, 17, 259-263.
- Weisberg, S. (2005) *Applied Linear Regression*, third edition. Wiley.
- Yeo, I. and Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87, 954-959.

See Also

[powtran](#), [optim](#), [pairs](#), [inv.res.plot](#), [plot.bctrans](#)

Examples

```
data(highway)
b <- highway[,c(8,1,2,10,5)] # select interesting columns
summary(ans <- bctrans1(b,family="yeo.johnson")) # zeros ==> use yeo.johnson
# or, compute using a formula and get the same answer.
summary(ans2 <-
bctrans(~Len+ADT+Trks+Shld+Sigs,data=highway,family="yeo.johnson"))
# or, first fit an lm, and extract the formula
m1 <- lm(Rate~Len+ADT+Trks+Shld+Sigs,data=highway)
summary(ans3 <- bctrans(m1,data=highway,family="yeo.johnson"))
```

```
# work with the response
b$Sigs <- (round(b$Sigs*b$Len)+1)/b$Len # redefine so no zeroes
summary(ans <- bctrans1(b)) # fit with box.cox
lrt.bctrans(ans,lrt=list(c(0,0,-1,1,0)))
plot(ans,family="power") # plot, but use ordinary powers
b <- cbind(b,powtran(ans)) # add transformed variables to data frame
```

BGSall

Berkeley guidance study

Description

Data from the Berkeley guidance study of children born in 1928-29 in Berkeley, CA. BGSall contains all the data, BGSboys the boys only, and BGSgirls the girls only.

Format

This data frame contains the following columns:

Sex 0 = males, 1 = females

WT2 Age 2 weight (kg)

HT2 Age 2 height (cm)

WT9 Age 9 weight (kg)

HT9 Age 9 height (cm)

LG9 Age 9 leg circumference (cm)

ST9 Age 9 strength (kg)

WT18 Age 18 weight (kg)

HT18 Age 18 height (cm)

LG18 Age 18 leg circumference (cm)

ST18 Age 18 strength (kg)

Soma Somatotype, a 1 to 7 scale of body type.

Source

Tuddenham, R. D. and Snyder, M. M. (1954). Physical Growth of California Boys and Girls from Birth to Eighteen years. Univ. of Calif. Publications in Child Development, 1, 183-364.

References

S. Weisberg (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 3.1.

Examples

```
data(BGSall)
data(BGSboys)
data(BGSgirls)
```

BigMac2003

World cities data

Description

Prices in many world cities from a 2003 Union Bank of Switzerland report.

Format

This data frame uses the name of the city as row names, and contains the following columns:

BigMac Minutes of labor to purchase a Big Mac

Bread Minutes of labor to purchase 1 kg of bread

Rice Minutes of labor to purchase 1 kg of rice

FoodIndex Food price index (Zurich=100)

Bus Cost in US dollars for a one-way 10 km ticket

Apt Normal rent (US dollars) of a 3 room apartment

TeachGI Primary teacher's gross income, 1000s of US dollars

TeachNI Primary teacher's net income, 1000s of US dollars

TaxRate Tax rate paid by a primary teacher

TeachHours Primary teacher's hours of work per week:

Source

Union Bank of Switzerland report, *Prices and Earnings Around the Globe* (2003 version), from http://www.ubs.com/1/e/ubs_ch/wealth_mgmt_ch/research.html.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 7.5.

Examples

```
data(BigMac2003)
```

 blowdown

Blowdown data

Description

Data from the Boundary Waters Canoe Area Wilderness Blowdown. The data frame blowdown includes nine species of trees, blowAPB gives data for aspen and paper birch, while blowBF gives data for Balsam Fir only.

Format

This data frame contains the following columns:

D Tree diameter, in cm

S Proportion of basal area killed for the four species BF, C, PB, BS, a measure of local severity of the storm.

y 1 if the tree died, 0 if it survived

SPP BF= balsam fir, BS= black spruce, C= cedar, JP= jackpine, PB= paper birch, RP= red pine, RM= red maple, BA = black ash, A= aspen.

Source

Roy Rich

References

S. Weisberg (2005). *Applied Linear Regression*, third edition. New York: Wiley.

Examples

```
data(blowBF)
```

brains

Mammal brain weights

Description

The data provided gives the average body weight in kilograms and the average brain weight in grams for sixty-two species of mammals.

Format

This data frame uses species names as row labels and contains the following columns:

BrainWt Brain weight, grams

BodyWt Body weight, kg

Source

Allison, T. and Cicchetti, D. (1976). Sleep in mammals: Ecology and constitutional correlates. *Science*, 194, 732-734.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 7.1.

Examples

```
data(brains)
```

cakes

Cakes data

Description

Oehlert (2000, Example 19.3) provides data from a small experiment on baking packaged cake mixes.

Usage

```
data(cakes)
```

Format

A data frame with 14 observations on the following 4 variables.

block a numeric vector

X1 Baking time, minutes

X2 Baking temperature, degrees F

Y Palatability score

Source

Oehlert, G. W. (2000). *A First Course in Design and Analysis of Experiments*. New York: Freeman.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 6.1.1.

Examples

```
data(cakes)
lm(Y~X1+X2+I(X1^2)+I(X2^2)+X1:X2, data=cakes)
```

`cathedral`*Cathedrals*

Description

Heights and lengths of Gothic and Romanesque cathedrals.

Format

This data frame uses cathedral names as row label and contains the following columns:

Type Romanesque or Gothic

Height Total height, feet

Length Total length, feet

Source

Stephen Jay Gould

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.7.

Examples

```
data(cathedral)
```

`caution`*Caution data*

Description

Artificial data to illustrate problems with residual plots.

Format

This data frame contains the following columns:

x1 Artificial data item.

x2 Artificial data item.

y Artificial data item.

Source

R. D. Cook and S. Weisberg (1999), Graphs in statistical analysis: Is the medium the message? *American Statistician*, 53, 29-37.

References

Weirsbeg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 8.1.5.

Examples

data(caution)

challeng

Challenger data

Description

Contains data from the performance of O-rings in 23 U.S. space shuttle flights prior to the Challenger disaster of January 20, 1986.

Format

This data frame contains the following columns:

Temp Air Temp at launch (degrees F)

Pres Leak check pressure

Fail Number of O-rings that failed

n 6, number of O-rings in launch

Erosion Number of erosion incidents

BlowBy Number of blowby incidents

Damage Total Damage Index

Date date of flight

Source

Dalal, S, Fowlkes, E. B. and Hoadley, B. (1989), Risk analysis of the space shuttle: Pre-challenger prediction of failure, *Journal of the American Statistical Association*, 84, 945-957. See also Tufte, E. R. (1997), *Visual and statistical Thinking: Displays of evidence for making decisions*, Cheshire, CT: Graphics Press.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 12.6.

Examples

data(challeng)

`chloride`*Chloride data*

Description

Seasonal variation in chloride level in marshes close to a road or distant from the road, in Minnesota.

Usage

```
data(chloride)
```

Format

A data frame with 32 observations on the following 4 variables.

Marsh March number

Type a factor with levels `Isolated` and `Roadside`

Month Month (4 = May, ... 10 = October)

Cl Chloride concentration, mg/liter

Source

Stefanie Miklovic and Susan Galatowitsch

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 6.5.

Examples

```
data(chloride)
library(nlme) # Use the mixed-effects package
# Fit a random coefficients model
summary(m1 <- lme(Cl~Month+Type, data=chloride, random=~1|Marsh))
```

`cloud`*Florida area cumulus experiment, FACE I.*

Description

The data summarize the results of the first Florida Area Cumulus Experiment, or FACE-1, designed to study the effectiveness of cloud seeding to increase rainfall in a target area (Woodley, Simpson, Biondini, and Berkley, 1977).

Format

This data frame contains the following columns:

A Action, 1=seed, 0=do not seed

D Day after June 16, 1975

S Suitability for seeding

C percent cloud cover in experimental area, measured using radar in Coral Gables, Florida

P $10^7 m^3$ prewetness

E echo motion category, either 1 or 2, a measure for type of cloud

Rain $10^7 m^3$ in target area

Source

Woodley, W.L., Simpson, J., Biondini, R., and Berkley, J. (1977). Rainfall results 1970-75: Florida area cumulus experiment. *Science*, 195, 735-742.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 9.17.

Examples

```
data(cloud)
```

```
conf.intervals
```

Compute marginal confidence intervals for regression estimates

Description

Thus function computes confidence intervals coefficient estimates for a regression object.

Usage

```
conf.intervals(...)
```

Arguments

... Arguments are passed to `confint` in the stats package

Value

This function calls `confint` in the stats package to compute confidence intervals of coefficient estimates.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

See Also

See also [confint](#)

delta.method	<i>Estimate and standard error of a nonlinear function of estimated regression coefficients</i>
--------------	---

Description

delta.method is a generic function that uses the delta method to get a first-order approximate standard error for a nonlinear function of a vector of random variables with known or estimated covariance matrix.

Usage

```
delta.method(object, ...)

## Default S3 method:
delta.method(object, g, var, ...)
## S3 method for class 'lm':
delta.method (object, g, var=vcov, parameterPrefix="b", ...)
## S3 method for class 'nls':
delta.method(object, g, var=vcov, ...)
## S3 method for class 'lmList':
delta.method(object, g, var=vcov, parameterPrefix="b", ...)
## S3 method for class 'nlsList':
delta.method(object, g, var=vcov, ...)
## S3 method for class 'lme':
delta.method (object, g, var=vcov, parameterPrefix="b", ...)
## S3 method for class 'nlme':
delta.method (object, g, var=vcov, ...)
## S3 method for class 'multinom':
delta.method (object, g, var=vcov, parameterPrefix="b", ...)
## S3 method for class 'polr':
delta.method (object, g, var=vcov, parameterPrefix="b", ...)
```

Arguments

object	For the default method, object is a named vector of p elements. This means that the call names(object) would return a list of p character strings that are the names of the elements of object. For the other methods, object is a regression object for which coef(object) returns a vector of parameter estimates.
g	A quoted string that is the function of the parameter estimates to be evaluated; see the details below.

<code>var</code>	The (estimated) covariance matrix of the coefficient estimates. For the default method, this argument is required. For all other methods, this argument must either provide the estimated covariance matrix or a function that when applied to <code>object</code> returns a covariance matrix. The default is generally to use the function <code>vcov</code> .
<code>parameterPrefix</code>	Typically a single letter with default "b" giving the prefix of the names of the parameter names used in the argument <code>g</code> ; see details.
<code>...</code>	Additional arguments; not currently used.

Details

Suppose x is a random vector of length p that is at least approximately normally distributed with mean β and estimated covariance matrix C . Then any function $g(\beta)$ of β , is estimated by $g(x)$, which is in large samples normally distributed with mean $g(\beta)$ and estimated variance $h'Ch$, where h is the first derivative of $g(\beta)$ with respect to β evaluated at x . This function returns both $g(x)$ and its standard error, the square root of the estimated variance.

The default method requires that you provide x in the argument `object`, C in the argument `var`, and a text expression in argument `g` that when evaluated gives the function g .

Since the delta method is often applied to functions of regression parameter estimates, the argument `object` may be the name of a regression object from which the vector x will be taken from `coef(object)`, and C will be taken from `vcov(object)` unless you provide some other estimate of variance, for example, using a sandwich estimator. Methods have been provided for many common regression models.

For regression models for which methods are not provided, you must extract the named vector of coefficient estimates and an estimate of its covariance matrix and then apply the default `delta.method` function.

In the argument `g` you must provide a quoted character string that gives the function of interest, for example `g="b1/b2"`, where `b1` and `b2` are names of two of the coefficient estimates.

For nonlinear regression objects of type `nls`, the call `coef(object)` returns the estimated coefficient vectors with names corresponding to parameter names. For example, `m2 <- nls(y~theta/(1+gamma*x), start = list(theta=2, gamma=3))` will have parameters named `c("theta", "gamma")`.

In many other familiar regression methods, such as `lm` and `glm`, the names of the coefficient estimates are the corresponding variable names, not parameter names. For example, in `m2 <- lm(Y~X1+X2)`, `names(coef(m2))` returns `c("(Intercept)", "X1", "X2")`. For models where the coefficient names are variable names, `delta.method` will provide names for the parameter estimates, given by "b0", "b1", ..., "bp", if `parameterPrefix` is left at its default value of "b". In this case, "b0" is the intercept (if the model has no intercept, then the numbering of the parameters starts with 1, not 0), "b1" is the first estimated parameter after the intercept, and so on.

Special characters should be avoided in the names of the elements of x as these can cause problems, and the `D` function used to compute derivatives may get confused. However, embedded spaces or ":" are permitted.

Value

A `data.frame` with two components named `Estimate` for the estimate, `SE` for its standard error. The value of `g` is given as a row label.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Section 6.1.2

See Also

First derivatives of `g` are computed using symbolic differentiation by the function `D`.

Examples

```
# cakes is a data frame with response Y, predictors X1 X2
data(cakes, package="alr3")
m1 <- lm(Y ~ X2 + I(X2^2), data = cakes) # quadratic polynomial
delta.method(m1, "-b1/(2*b2)") # X2 that maximizes the quadratic
# second order polynomial in two predictors:
m2 <- lm(Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2, data=cakes)
# Find X1 to maximize Y when X2=350:
delta.method(m2, "(b1+b5*350)/(-2*b3)")
```

domedata

Metrodome fan experiment

Description

These files give the results of two experiments to see if manipulating the air conditioning fans in the Minneapolis metrodome can effect the distance travelled by a baseball. The data in `domedata` were collected in April 2003. The experiment was repeated in May 2003 and `domedata1` gives the combined data from the two experiments.

Usage

```
data(domedata1)
```

Format

A data frame with 96 observations on the following 7 variables.

Date a factor with levels March- May

Cond a factor with levels Headwind, Tailwind

Angle the actual angle

Velocity in feet per second

BallWt weight of ball in grams

BallDia diameter of ball in inches

Dist distance in feet of the flight of the ball

Source

Ivan Marusic

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.21.

Examples

```
data (domedata1)
```

donner

Donner party

Description

The Donner Party was the most famous tragedy in the history of the westward migration in the United States. In the winter of 1846-47, about ninety wagon train emigrants were unable to cross the Sierra Nevada Mountains of California before winter, and almost one-half starved to death. Perhaps because they were ordinary people – farmers, merchants, parents, children. These data include some information about each of the members of the party from Johnson (1996).

Format

This data frame uses the person's name as row labels and contains the following columns:

Age Approximate age in 1846

Outcome 1 if survived, 0 if died

Sex Male or Female

Family.name Either a family name, hired or single

Status Family, single or hired

Source

Johnson, K. (1996). *Unfortunate Emigrants: Narratives of the Donner Party*. Logan, UT: Utah State University Press, <http://www.metrogourmet.com/crossroads/KJhome.htm>.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 12.4.

Examples

```
data (donner)
```

downer

*Downer data***Description**

For unknown reasons, some dairy cows become recumbant—they lay down. This condition can be serious, and may lead to death of the cow. These data are from a study of blood samples of over 500 cows studied at the Ruakura (N.Z.) Animal Health Laboratory during 1983-84. A variety of blood tests were performed, and for many of the animals the outcome (survived, died, or animal was killed) was determined. The goal is to see if survival can be predicted from the blood measurements. Case numbers 12607 and 11630 were noted as having exceptional care—and they survived.

Format

This data frame contains the following columns:

Calving 0 if measured before calving, 1 if after

Daysrec Days recumbent

CK Serum creatine phosphokinase (U/l at 30C)

AST serum asparate amino transferase (U/l at 30C)

Urea serum urea (mmol/l)

PCV Packed Cell Volume (Haemactocrit),

Inflamat inflammation 0=no, 1=yes

Myopathy Muscle disorder, 1 if present, 0 if absent

Outcome outcome: 1 if survived, 0 if died or killed

Source

Clark, R. G., Henderson, H. V., Hoggard, G. K. Ellison, R. S. and Young, B. J. (1987). The abiltiy of biochemical and haematoligical tests to predict recovery in periparturient recumbent cows. *NZ Veterinary Journal*, 35, 126-133.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 12.1.

Examples

data(downer)

`drugcost`*Drug cost.*

Description

These data are to try to understand the effect of health plan characteristics on drug costs. Health plans vary in size, given as member months. Some plans use generic drugs more than others. All differ on copayments. Some have strong restrictions on which drugs can be dispensed value of RI=0 means that all drugs are dispensed, RI=100 means that only one per category is available. The goal is to determine the terms that are related to cost, and in particular to understand the role of GS and RI in determining cost.

Format

This data frame uses a short code name for the drug plan as row labels and contains the following columns:

COST Ave. cost to plan for 1 prescription for 1 day

RXPM Number of prescriptions per member per year

GS Percent generic substitution, number between 0 (no substitution) to 100 (always use generic substitute)

RI Restrictiveness index (0=none, 100=total)

COPAY Average Rx copayment

AGE Average age of member

F Percent female members

MM Member months, a measure of the size of the plan

Source

Mark Siracuse

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 9.15.

Examples

```
data(drugcost)
```

`dwaste`*Crock data.*

Description

An experiment was conducted to study the *O2UP*, oxygen uptake in milligrams of oxygen per minute, given five chemical measurements: biological oxygen demand (BOD), total Kjeldahl nitrogen (TKN), total solids (TS), total vital solids (TVS), which is a component of TS, and chemical oxygen demand (COD), each measured in milligrams per liter (Moore, 1975).

Format

This data frame contains the following columns:

Day Day number

BOD Biological oxygen demand

TKN Total Kjeldahl nitrogen

TS Total Solids

TVS Total volatile solids

COD Chemical oxygen demand

O2UP Oxygen uptake

Source

Moore, J. (1975). Total Biomedical Oxygen Demand of Animal Manures. Unpublished Ph. D. dissertation, University of Minnesota.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 10.6

Examples

```
data(dwaste)
```

`florida`*Florida presidential election*

Description

County-by-county vote for president in Florida in 2000 for Bush, Gore and Buchanan.

Usage

```
data(florida)
```

Format

A data frame three variables for each of Florida's 67 counties.

Gore Vote for Gore

Bush Vote for Bush

Buchanan Vote for Buchanan

Source

http://www.abcnews.go.com.sections/politics/2000vote/general/FL_county.html

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 9.10.

Examples

```
data(florida)
## maybe str(florida) ; plot(florida) ...
```

`forbes`*Forbes data*

Description

The data consists of 17 pairs of numbers corresponding to observed boiling point and corrected barometric pressure, at locations in the Alps.

Format

This data frame contains the following columns:

Temp Adjusted boiling point of water in degrees F.

Pressure Atmospheric pressure, in inches of Mercury

Lpres 100 times log(Pressure,10), rounded to two digits beyond the decimal point

Source

Forbes, J. (1857). Further experiments and remarks on the measurement of heights and boiling point of water. *Transactions of the Royal Society of Edinburgh*, 21, 235-243.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 1.1.

Examples

```
data(forbes)
```

```
ftcollinssnow
```

```
Ft. Collins snowfall
```

Description

Monthly snowfall data for Fort Collins, CO, 1900-01 to 1992-93

Format

This data frame contains the following columns:

YR1 Year corresponding to the September to December data

Early September to December snowfall, inches

Late January to June snowfall, inches

Source

<http://www.ulysses.atmos.colostate.edu>

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 1.1.

Examples

```
data(ftcollinssnow)
```

`fuel2001`*Fuel consumption*

Description

Data on motor fuel consumption and related variables, for the year 2001. The unit is a state in the United States or the District of Columbia. Data are for 2001, unless noted.

Format

This data frame contains the following columns:

Drivers Number of Licensed drivers in the state

FuelC Gasoline sold for road use (1000s of gal.)

Income Per capita personal income (year 2000)

Miles Miles of Federal-aid highway miles in the state

MPC Estimated miles driven per capita

Pop Population age 16 and over

Tax Gasoline state tax rate, cents per gallon

State State name

Source

<http://www.fhwa.dot.gov/ohim/hs01/index.htm>

References

Weisberg, S. (2005). *Applied Linear Regression*, third edition. New York: Wiley.

Examples

```
data(fuel2001)
fuel2001$Dlic <- 1000*fuel2001$Drivers/fuel2001$Pop
fuel2001$Fuel <- 1000*fuel2001$FuelC/fuel2001$Pop
fuel2001$Income <- fuel2001$Income/1000
fuel2001$logMiles <- log(fuel2001$Miles,2)
pairs(Fuel~Tax+Dlic+Income+logMiles,data=fuel2001)
```

`galapagos`*Galapagos species data*

Description

Johnson and Raven (1973) have presented data giving the number of species and related variables for 29 different islands in the Galapagos Archipelago.

Format

This data frame uses the island name as row labels and contains the following columns:

NS Number of Species

ES Number of endemic species (occur only on that island)

Area Surface area of island, hectares

Anear Area of closest island, hectares

Dist Distance to closest island, km

DistSC Distance from Santa Cruz Island, km

Elevation Elevation in m, missing values given as zero

EM 1 if elevation is observed, 0 if missing

Source

Johnson, M.P., and Raven, P.H. (1973). Species number and endemism: The Galapagos Archipelago revisited. *Science*, 179, 893-895.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 10.8.

Examples

```
data(galapagos)
```

`galtonpeas`*Galton's peas*

Description

In a paper presented to the Royal Institute on February 9, 1877, Sir Francis Galton discussed his experiments on sweet peas in which he compared the sweet peas produced by parent plants to those produced by offspring plants. In these experiments he could observe inheritance from one generation to the next. Galton categorized the parent plants according to the typical diameter of the peas they produced.

Format

This data frame contains the following columns:

Parent mean diameter of parent

Progeny mean diameter of offspring

SD offspring standard deviation

Source

Pearson, K. (1930). *Life and Letters and Labours of Francis Galton*, Vol IIIa. Cambridge: Cambridge University Press.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 5.1.

Examples

```
data(galtonpeas)
```

`heights`*Pearson-Lee data*

Description

Karl Pearson organized the collection of data on over 1100 families in England in the period 1893 to 1898. This particular data set gives the heights in inches of mothers and their daughters, with up to two daughters per mother. All daughters are at least age 18, and all mothers are younger than 65. Data were given in the source as a frequency table to the nearest inch. Rounding error has been added to remove discreteness from graph.

Format

This data frame contains the following columns:

Mheight Mother's ht, in.

Dheight Daughter's ht, in.

Source

K. Pearson and A. Lee (1903), On the laws of inheritance in man, *Biometrika*, 2, 357–463, Table 31.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 1.1.

Examples

```
data(pearlee1)
```

highway

Highway accidents

Description

The data comes from a unpublished master's paper by Carl Hoffstedt. They relate the automobile accident rate, in accidents per million vehicle miles to several potential terms. The data include 39 sections of large highways in the state of Minnesota in 1973. The goal of this analysis was to understand the impact of design variables, Acpts, Slim, Sig, and Shld that are under the control of the highway department, on accidents.

Format

This data frame contains the following columns:

ADT average daily traffic count in thousands

Trks truck volume as a percent of the total volume

Lane total number of lanes of traffic

Acpt number of access points per mile

Sigs number of signalized interchanges per mile

Itg number of freeway-type interchanges per mile

Slim speed limit in 1973

Len length of the highway segment in miles

Lwid lane width, in feet

Shld width in feet of outer shoulder on the roadway

Hwy An indicator of the type of roadway or the source of funding for the road; 0 if MC, 1 if FAI, 2 if PA, 3 if MA

Rate 1973 accident rate per million vehicle miles

Source

Carl Hoffstedt

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 7.2.

Examples

```
data (highway)
```

hooker

Hooker's data

Description

In his original paper, Forbes provided additional data collected by the botanist Dr. Joseph Hooker on temperatures and boiling points measured often at higher altitudes in the Himalaya Mountains.

Format

This data frame contains the following columns:

Temp Measured boiling temperature, degrees F.

Pressure Measured air pressure, inches of Mercury.

Source

Forbes, J. (1957). Further experiments and remarks on the measurement of heights by boiling point of water. *Transactions of the Royal Society of Edinburgh*, 21, 235-243.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 2.2.4.

Examples

```
data (hooker)
```

htwt

*Artificial height and weight data***Description**

The data for this table are a sample size of ten 18-year old girls taken from the study that was conducted by Tuddenham and Snyder (1954).

Format

This data frame contains the following columns:

Ht Height (cm) at age 18

Wt Weight (kg) at age 18

Source

Tuddenham, R., and Snyder, M. (1954). Physical growth of California boys and girls from birth to age 18. *California Publications on Child Development*, 1, 183-364.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 2.1.

Examples

```
data(htwt)
```

inf.index

*Influence index plot***Description**

Provides index plots of Cook's distances, leverages, Studentized residuals and outlier significance levels for a regression object.

Usage

```
inf.index(m, cooks.distance, rstudent, outlier.t.test, leverages, ...)

## S3 method for class 'lm':
inf.index(m, cooks.distance = TRUE, rstudent = TRUE,
          outlier.t.test = TRUE, leverages = TRUE, ...)
```

Arguments

<code>m</code>	A regression object
<code>cooks.distance</code>	If TRUE, plot Cook's distances
<code>rstudent</code>	If TRUE, plot the Studentized residuals
<code>outlier.t.test</code>	If TRUE, index plot $n*2*qt(-abs(rstandard),m$df-1)$, the two-tailed significance levels for the outlier test based on the Bonferroni inequality
<code>leverages</code>	If TRUE, plot the leverages
<code>...</code>	Arguments passed to <code>plot</code>

Value

Used for its side effect of producing a graph.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 8

See Also

[cooks.distance](#)

Examples

```
data(rat)
names(rat)
m1 <- lm(y~BodyWt+LiverWt+Dose, data=rat)
inf.index(m1)
```

`inv.res.plot` *Inverse response plots to transform the response*

Description

For a `lm` object, draws an `inverse.response` plot with the response Y on the vertical axis and the fitted values \hat{Y} on the horizontal axis. Uses `nls` to estimate λ in the function $\hat{Y} = b_0 + b_1 Y^\lambda$. Adds the fitted curve to the plot. `inv.res.plot` is an alias for `inverse.response.plot`.

Usage

```
inverse.response.plot(m, lambda=c(0,1), maxiter=100, xlab=NULL, ...)
```

Arguments

<code>m</code>	A lm regression object
<code>lambda</code>	A vector of values for lambda. A plot will be produced with curves corresponding to these lambdas and to the least squares estimate of lambda
<code>xlab</code>	The horizontal axis label. If NULL, it is constructed by the function.
<code>maxiter</code>	Passed to <code>nls</code> , the maximum number of iterations to be used to compute the optimal value of lambda
<code>...</code>	Other arguments passed to <code>inv.tran.plot</code> and then to <code>plot</code> .

Value

As a side effect, a plot is produced with the response on the horizontal axis and fitted values on the vertical axis. Several lines are added to be plot as the ols estimates of the regression of \hat{Y} on Y^λ , interpreting $\lambda = 0$ to be natural logarithms.

Numeric output is a list with elements

<code>lambda</code>	Estimate of transformation parameter for the response
<code>se</code>	Standard error of the estimate
<code>RSS</code>	The residual sum of squares at the minimum

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 7

See Also

[inv.tran.plot](#) and [inv.tran.estimate](#), for which this is just a convenient front-end, and [nls](#).

Examples

```
data(highway)
highway$Sigs1 <- (round(highway$Sigs*highway$Len)+1)/highway$Len
attach(highway)
d <- data.frame(Rate=Rate, logLen=logb(Len, 2),
                logADT=logb(ADT, 2), logTrks=logb(Trks, 2),
                Slim=Slim, Shld=Shld, logSigs1=logb(Sigs1, 2))
attach(d)
m2 <- lm(Rate~logLen+logADT+logTrks+Slim+Shld+logSigs1, d)
inv.res.plot(m2, key=c(6, 2))
```

inv.tran.plot *Choose a predictor transformation visually or numerically*

Description

inv.tran.plot draws a two-dimensional scatterplot of Y versus X , along with the OLS fit from the regression of Y on $(X^\lambda - 1)/\lambda$. inv.tran.estimate find the nonlinear least squares estimate of λ and its standard error.

Usage

```
inv.tran.plot(x, y, lambda=c(-1, 0, 1), lty=1:(1+length(lambda)),
             col=rainbow(length(lambda)+1), xlab=deparse(substitute(x)),
             ylab=deparse(substitute(y)), family="box.cox", optimal=TRUE,
             key="topleft", ...)

inv.tran.estimate(x, y, family="box.cox", ...)
```

Arguments

x	The predictor variable
y	The response variable
lambda	The powers used in the plot. The optimal power than minimizes the residual sum of squares is always added unless optimal is FALSE.
family	The transformation family to use, "box.cox", "yeo.johnson", or a user-defined family.
optimal	Include the optimal value of lambda?
lty	line types corresponding to the powers
col	color corresponding to the powers
key	The default is "topleft", in which case a legend is added to the top left corner of the plot; other choices include "bottomright". If key is a vector of two coordinates, the legend is drawn at the coordinates specified. If key is FALSE, no key is drawn; if TRUE, you can place the key interactively by clicking on the plot.
xlab	Label for the horizontal axis.
ylab	Label for the vertical axis.
...	additional arguments passed to other methods.

Value

inv.tran.plot returns a graph and a data.frame with λ in the first column, and the residual sum of squares from the regression for that λ in the second column.

inv.tran.estimate returns a list with elements lambda for the estimate, se for its standard error, and RSS, the minimum value of the residual sum of squares.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>

References

Weisberg, S. (2005). *Applied Linear Regression*, third edition. New York: Wiley.

See Also

[powtran](#), [inverse.response.plot](#), [optimize](#), [nls](#)

Examples

```
data(baeskel)
attach(baeskel)
inv.tran.plot(Sulfur, Tension, key=c(.6, 450))
ans <- inv.tran.estimate(Sulfur, Tension)
# redraw the plot, including the nls estimate
inv.tran.plot(Sulfur, Tension, lambda=c(ans$lambda, -1, 0, 1), key=c(.6, 450))
```

jevons

Jevon's gold coin data

Description

In a study of coinage, W. Stanley Jevons weighed 274 gold sovereigns that he had collected from circulation in Manchester, England. For each coin, he recorded the weight, after cleaning, to the nearest .001 gram, and the date of issue. The age classes are coded 1 to 5, roughly corresponding to the age of the coin in decades. The standard weight of a gold sovereign was suppose to be 7.9876 grams; minimum legal weight was 7.9379 grams.

Format

This data frame contains the following columns:

Age Age of coins, decades

n Number of coins

Weight Average weight, grams

SD Standard deviation.

Min Minimum weight

Max Maximum weight

Source

Stephen Stigler

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 5.6.

Examples

```
data(jevons)
```

lakemary

Lake Mary bluegills

Description

78 bluegills were captured from Lake Mary, Minnesota. On each fish, a key scale was removed. The age of a fish is determined by counting the number of annular rings on the scale. The goal is to relate length at capture to the radius of the scale.

Format

This data frame contains the following columns:

Age Years

Length mm

Source

Collected by Richard Frie, and discussed in S. Weisberg (1986), A linear model approach to the backcalculation of fish length, *J. Amer. Statist. Assoc.*, 81, 922-929.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 11.2.1.

Examples

```
data(lakemary)
```

lakes

Lake zooplankton diversity

Description

These data give the number of known crustacean zooplankton species for 69 world lakes. Also included are a number of characteristics of each lake. There are missing values.

Format

This data frame uses lake name as row label and contains the following columns:

Species Number of zooplankton species

MaxDepth Maximum lake depth, m

MeanDepth Mean lake depth, m

Cond Specific conductance, micro Siemens

Elev Elevation, m

Lat N latitude, degrees

Long W longitude, degrees

Dist distance to nearest lake, km

NLakes number of lakes within 20 km

Photo Rate of photosynthesis, mostly by the ¹⁴C method

Area Lake area, in hectares

Source

Dodson, S. (1992), Predicting crustacean zooplankton species richness, *Limnology and Oceanography*, 37, 848–856.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 8.12.

Examples

```
data(lakes)
```

landrent	<i>Land rent</i>
----------	------------------

Description

The data were collected by Douglas Tiffany to study the variation in rent paid in 1977 for agricultural land planted to alfalfa.

Format

This data frame contains the following columns:

- X1** average rent for all tillable land
- X2** density of dairy cows (number per square mile)
- X3** proportion of farmland used for pasture
- X4** 1 if liming required to grow alfalfa; 0 otherwise
- Y** average rent per acre planted to alfalfa

Source

Douglas Tiffany

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 9.12.

Examples

```
data(landrent)
```

lathel	<i>Lathe data</i>
--------	-------------------

Description

These data are the results of an experiment to study the performance of cutting-tool material in cutting steel on a lathe. The two factors are revolution speed and feed rate. The response is tool life in minutes.

Format

This data frame contains the following columns:

- Feed** Coded feed rate, coded as (actual feed rate -13)/6. Feed is in thousandths of an inch per revolution.
- Speed** Coded speed, coded as (actual speed -900)/300. Speed is in feet per minute.
- Life** Life of tool until failure, minutes

Source

M. R. Delozier

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.3.

Examples

```
data(lathel)
```

```
mantel
```

Mantel's artificial data for stepwise regression

Description

An artificial data set suggested by N. Mantel to illustrate stepwise regression methods.

Usage

```
data(mantel)
```

Format

A data frame with 5 observations on the following 4 variables.

Y the response

X1 predictor 1

X2 predictor 2

X3 predictor 3

Source

Mantel, N. (1970). Why stepdown procedures in variable selection? *Technometrics*, 12, 621–625.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 10.2.

Examples

```
data(mantel)
```

mile

World records for the mile run

Description

World record times for the mile run, 1861–2003.

Usage

```
data(mile)
```

Format

A data frame with 46 observations:

Year Year in which the record was set

Time Running time, in seconds

Name Name of person setting the record

Country Country of residence of the record setter

Place Place the record was set

Gender Gender of the record holder

Source

Data source: <http://www.saunalahti.fi/~sut/eng/>

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.20.

Examples

```
data(mile)
```

 Mitchell

Mitchell soil temperature

Description

Data collected by Kenneth G. Hubbard on soil temperature at 20 cm depth in Mitchell, Nebraska for 17 years (1976-1992) The variable month is the month number.

Format

This data frame contains the following columns:

Month Months beginning Jan, 1976

Temp Average soil temperature, degrees C

Source

Kenneth G. Hubbard

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 1.2.

Examples

```
data(Mitchell)
```

 mmps

Marginal model plot

Description

For a regression object, plots the response on the vertical axis versus a linear combination u of terms in the mean function on the horizontal axis. Added to the plot are a `loess` smooth for the graph, along with a loess smooth from the plot of the fitted values on u .

Usage

```
mmps(object, vars=~., fitted=TRUE, layout=NULL, ask, ...)

## S3 method for class 'lm':
mmp(object, u, mean = TRUE, sd = FALSE,
     xlab = deparse(substitute(u)), degree = 1, span = 2/3, key="topleft",
     lineColors = c("blue", "red"), ...)

## S3 method for class 'glm':
```

```
mmp(object, u, mean = TRUE, sd = FALSE,
     xlab = deparse(substitute(u)), degree = 1, span = 2/3, key="topleft",
     lineColors = c("blue", "red"), ...)
```

Arguments

object	A regression object, of type either <code>lm</code> or <code>glm</code> , for which there is a <code>predict</code> method defined.
vars	A one-sided formula. A marginal model plot will be drawn for each variable on the right-side of this formula that is not a factor. The default is <code>~.</code> , which specifies that all the terms in <code>formula(object)</code> will be used. See examples below.
fitted	If the default <code>TRUE</code> , then a marginal model plot in the direction of the fitted values or linear predictor of a generalized linear model will be drawn.
layout	A reasonable layout for the plots in the window is determined by the program. If you don't like the default you can set your own layout: <code>c(2, 3)</code> means two rows and three columns.
ask	If <code>TRUE</code> , ask before clearing the graph window to draw more plots.
...	Additional arguments passed from <code>mmps</code> to <code>mmp</code> and then to <code>plot</code> . Users should generally use <code>mmps</code> .
u	The quantity to be plotted on the horizontal axis. The default is the predicted values <code>predict(object)</code> . Can be any other vector of length equal to the number of observations in the object.
mean	If <code>TRUE</code> , compare mean smooths
sd	If <code>TRUE</code> , compare sd smooths
xlab	label for horizontal axis
degree	Degree of the local polynomial, passed to <code>loess</code> . The usual default for <code>loess</code> is 2, but the default here is 1.
span	Span, the smoothing parameter for <code>loess</code> .
key	If not <code>NULL</code> , this gives the location of the legend, chosen from the default "topleft", or "bottomleft", "bottomright" or "topright".
lineColors	colors for data and model smooth, respectively

Details

`mmp` draws one marginal model plot. `mmps` draws all marginal model plots, versus each of the terms in the model and versus fitted values. `mmps` should not be used if some of the terms in the model are factors or interactions.

Value

Used for its side effect of producing plots.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 8

See Also

[loess](#), [plot](#)

Examples

```
data(ufcwc)
c1 <- lm(Height ~ Dbh, ufcwc)
mmp(c1, ufcwc$Dbh, xlab="Diameter, Dbh", lineColors=c("black", "black"))
mmps(c1)
```

MWwords

Word frequencies from Mosteller and Wallace

Description

The data give the frequencies of words in works from four different sources: the political writings of eighteenth century American political figures Alexander Hamilton, James Madison, and John Jay, and the book *Ulysses* by twentieth century Irish writer James Joyce.

Format

This data frame uses the word as row labels and contains the following columns:

Hamilton Hamilton frequency

HamiltonRank Hamilton rank

Madison Madison frequency

MadisonRank Madison rank

Jay Jay frequency

JayRank Jay rank

Ulysses Word frequency in *Ulysses*

UlyssesRank Word rank in *Ulysses*

Source

Mosteller, F. and Wallace, D. (1964). *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 2.10.

Examples

```
data(MWwords)
```

npdata

Northern pike catch per unit effort

Description

Catch per unit effort data for 16 Minnesota lakes

Usage

```
data(npdata)
```

Format

A data frame with 16 observations on the following 4 variables.

CPUE Estimated catch per unit effect

SECPUE Estimated standard error of CPUE

Density Estimated fish density

SEdens Estimated standard error of Density

Source

R. Pierce, Minnesota Dept. of Natural Resources

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 4.6.3.

Examples

```
data(npdata)
```

 oldfaith

Old Faithful Geyser data

Description

Data on eruptions of Old Faithful Geyser, October 1980. Variables are the duration in seconds of the current eruption, and the time in minutes to the next eruption. Collected by volunteers, and supplied by the Yellowstone National Park Geologist. Data was not collected between approximately midnight and 6 AM.

Format

This data frame contains the following columns:

Duration Duration in seconds

Interval Time to next eruption

Source

R. Hutchinson

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 1.4.

Examples

```
data(oldfaith)
```

 outlier.t.test

Bonferroni test for outliers in linear models

Description

Compute significance levels for the mean-shift outlier model using the Bonferroni inequality

Usage

```
outlier.t.test(m, order=TRUE, bound=1)
```

Arguments

m	A model of type lm.
order	If TRUE, order the cases according to the p-value. If FALSE, don't order.
bound	Ignore cases with p-value bigger or equal to this value.

Details

Returns $\text{length}(\text{res}) * 2 * (1 - \text{pt}(\text{abs}(\text{res}), \text{df}))$, where $\text{res} = \text{rstandard}(\text{m})$ is the vector of Studentized residuals. These are two-sided Bonferroni significance levels for testing a single outlier.

Value

A data frame with columns giving the value of the studentized residual and corresponding Bonferroni p-value, and one row for each case for which the bound is satisfied.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

Weisberg, S. (2005). *Applied Linear Regression*, third edition, Wiley.

Examples

```
data(cloud)
m1 <- lm(logb(Rain, 2) ~ S + logb(C, 2) + logb(P, 2) + E + A, data=cloud, subset=-2)
outlier.t.test(m1)
```

physics

Physics data

Description

The file physics contains results for π^+ meson as input and π^+ meson as output. physics1 is for π^- to π^- .

Format

This data frame contains the following columns:

x Inverse total energy
y Scattering cross-section/sec
SD Standard deviation

Source

Weisberg, H., Beier, H., Brody, H., Patton, R., Raychaudhari, K., Takeda, H., Thern, R. and Van Berg, R. (1978). s-dependence of proton fragmentation by hadrons. II. Incident laboratory momenta, 30–250 GeV/c. *Physics Review D*, 17, 2875–2887.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 5.1.1.

Examples

```
data(physics1)
```

```
pipeline
```

```
Alaska pipeline
```

Description

The Alaska pipeline data consists of in-field ultrasonic measurements of the depths of defects in the Alaska pipeline. The depth of the defects were then re-measured in the laboratory. These measurements were performed in six different batches. The data were analyzed to calibrate the bias of the field measurements relative to the laboratory measurements. In this analysis, the field measurement is the response variable and the laboratory measurement is the predictor variable.

These data were originally provided by Harry Berger, who was at the time a scientist for the Office of the Director of the Institute of Materials Research (now the Materials Science and Engineering Laboratory) of NIST. These data were used for a study conducted for the Materials Transportation Bureau of the U.S. Department of Transportation.

Format

This data frame contains the following columns:

Field Number of defects measured in the field.

Lab Number of defects measured in the field.

Batch Batch number

Source

<http://www.itl.nist.gov/div898/handbook/pmd/section6/pmd621.htm>

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 8.3.

Examples

```
data(pipeline)
```

plot.bctrans *Plot and summary helpers for bctrans objects*

Description

Extract, view, and plot useful summaries of a ‘bctrans’ object for transformations to normality.

Usage

```
## S3 method for class 'bctrans':
plot(x, y=x$Y, lambda=coef(x), family=x$family, plot=pairs, ...)

## S3 method for class 'bctrans':
summary(object, digits=4, ...)

## S3 method for class 'bctrans':
coef(object, ...)

## S3 method for class 'bctrans':
vcov(object, ...)
```

Arguments

x	An object created by bctrans.
y	If not NULL, add this variable to the scatterplot matrix
object	An object created by bctrans.
lambda	A vector of transformation parameters.
family	A family of transformations, probably "box.cox" or "yeo.johnson"
plot	Name of the plotting function to view the results, usually pairs or similar.
digits	Accuracy of summary output for ‘bctrans’ objects.
...	Arguments passed to the plotting method.

Value

The `plot` method is used for its side effect of producing a scatterplot matrix. The helper `summary` summarizes a bctrans object. This helper calls `lrt.bctrans`, but the latter can be called directly. The helpers `coef` and `vcov` return, respectively, the estimated value of lambda, and an estimate of its covariance matrix (based on a profile log-likelihood).

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>

References

Weisberg, S. (2005) *Applied Linear Regression*, third edition. Wiley.

See Also

[bctrans](#), [lrt.bctrans](#), [pairs](#), [inv.res.plot](#)

Examples

```
data(highway)
b <- highway[,c(8,1,2,10,5)] # select interesting columns
summary(ans <- bctrans(b,family="yeo.johnson")) # zeros ==> use yeo.johnson
b$Sigs <- (round(b$Sigs*b$Len)+1)/b$Len # redefine so no zeroes
summary(ans <- bctrans(b)) # fit with box.cox
b$Sigs <- (round(b$Sigs*b$Len)+1)/b$Len # redefine so no zeroes
summary(ans <- bctrans(b)) # fit with box.cox
lrt.bctrans(ans,lrt=list(c(0,-1,1,0)))
plot(ans,family="power") # plot, but use ordinary powers
b <- cbind(b,powtran(ans)) # add transformed variables to data frame
```

pod

Fit partial one-dimensional, or POD models, based on a linear model

Description

A partial one-dimensional model, or a POD model, provides a concise description of a regression model with many predictors and one grouping variable. It requires a nonlinear regression fit.

Usage

```
## This is a generic function with different arguments depending on the class of the
## first argument. The generic form is
```

```
pod (x, ...)
```

```
## If the first argument to pod is a formula suitable for defining a linear model,
```

```
## S3 method for class 'formula':
```

```
pod(formula, data = sys.parent(), group, subset, weights,
     na.action, mean.function = c("pod", "common", "parallel",
     "general"), singular.ok = FALSE, contrasts = NULL, offset,
     control = nls.control(), ...)
```

```
## If the first argument to pod is the result of a linear model fit, the following
## function call is used. All the arguments for pod.formula can also be passed to
## pod.lm.
```

```
## S3 method for class 'lm':
```

```

pod (x, group, mean.function, control, ...)

## The following related function require a pod fit as the first argument:

## S3 method for class 'pod':
anova(object, scale=0, test="F", ...)

## S3 method for class 'pod':
plot(x, colors=rainbow(nlevels(x$group)),
      pch=1:nlevels(x$group), key="topleft", identify=FALSE,
      xlab="Linear Predictor", ylab=as.character(c(formula(x)[[2]])),
      ...)

## S3 method for class 'pod.lm':
plot(x, colors=rainbow(nlevels(x$group)),
      pch=1:nlevels(x$group), key="topleft", identify=FALSE,
      xlab="Linear Predictor", ylab=as.character(c(formula(x)[[2]])),
      ...)

```

Arguments

formula	A linear regression formula, as used in <code>lm</code> , or the results of a call to <code>pod</code> . See details below.
x	The result of a <code>lm</code> fit from which the formula and the arguments data, subset, weights, <code>na.action</code> , <code>singular.ok</code> , <code>contrasts</code> and <code>offset</code> will be taken.
data	An optional data frame for the data to be used
group	The name of a grouping variable (not quoted) that defines the groups; see details below.
mean.function	Which mean function should be fit? The default is “pod”, that fits the partial one-dimensional mean function. The other options are “common”, which fits a linear model with no grouping effects; “parallel” fits a parallel within-group regression, and “general”, available in <code>pod</code> but not <code>plot.pod</code> , fits a separate coefficient for each term in the model for each level of the grouping variable.
subset	As in <code>lm</code> , the subset of cases to be used in the analysis
weights	Weights will be used in fitting non-pod models. Since this argument is not supported for <code>nls</code> models, weights are ignored for fitting <code>pod</code> models. If <code>nls</code> is ever updated, then <code>pod</code> models will correctly use the weights.
na.action	At present, only <code>na.omit</code> is likely to work.
singular.ok	The default equal to <code>FALSE</code> is recommended.
contrasts	Same as in <code>lm</code>
offset	Same as in <code>lm</code>
control	A <code>pod</code> model is fit using the nonlinear least squares routine <code>nls</code> . This routine is very sensitive to starting values and other parameters set in the algorithm. This routine selects starting values for you that will often work very well, but in some problems the user may want to change the defaults to <code>nls</code> program using the <code>nls.control</code> function.

...	In pod, other arguments passed to nls, such as control parameters. In pod.anova, there are two additional arguments
object	The result of a call to pod.
scale	Used for test in anova
test	Default is to compute F tests.
colors	Colors for groups in the pod plot
pch	Plotting symbol for the groups in the pod plot
identify	If TRUE, clicking the mouse on a graph will print the case name of the nearest point. This continues until turned off (by pushing the escape key, among other ways of doing this).
key	The default is "topleft", in which case a legend is added to the top left corner of the plot; other choices include "bottomright". If key is a vector of two coordinates, the legend is drawn at the coordinates specified. If key is FALSE, no key is drawn; if TRUE, you can place the key interactively by clicking on the plot.
xlab	Horizontal label, optional
ylab	Vertical label, optional

Details

Suppose we start with a linear mean function specified by $y \sim x_1 + x_2 + x_3$, where the right-side variables can be any valid R variables, such as transformations, but NOT factors or interactions (if you want to include these, you need to create the dummy variables yourself). The right-hand side variables must also be linearly independent. We also specify a grouping variable z with, say, g levels. Let G_i be a dummy variable with values equal to 1 when $z = i$, and zero otherwise. The pod mean function is then has the nonlinear mean function

$$E(y|x, z) = \beta_0 + \beta'x + \sum_{j=2}^g G_j(\theta_{0j} + \theta_{ij}\beta'x)$$

This is a nonlinear mean function that specifies that the response depends on the predictors only through one linear combination, that the dependence is linear, but the slope and intercept may be different for each group.

The pod mean function is easily fit using `nls`. For example, if z has two levels, a nonlinear mean function that would work is $y \sim b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + G_2 * (th_{02} + th_{12} * (b_1 * x_1 + b_2 * x_2 + b_3 * x_3))$. Starting values can be determined by fitting the parallel mean function, for this is a linear mean function and easily fit.

The function `pod` automates this process; you need only input the linear part of the mean function, and `pod` does the rest.

The function `anova.pod` is a little different from the generic `anova` function. It fits and compares all four mean functions that can be fit with the `pod` function: (1) no group effect; (2) parallel within group method; (3) pod mean function; (4) mean function with all main effects and two-factor interactions with the grouping variable.

The function `plot` will plot the response versus the fitted linear combination for either the common, parallel or pod models. There is no 2D plot for the general model.

Objects of class `pod` have methods defined for the generic functions `print`, `summary`, `coef`, `anova`, `deviance`, `vcov`, `resid`, `formula`, `fitted`, `df.residual` and `predict`. `pod$response` returns the values of the response.

Value

If `mean.function="pod"`, a list of class `pod`, with the following components

<code>nls.fit</code>	The usual output from a <code>nls</code> fit.
<code>linear.part</code>	The estimated linear predictor $\beta'x$.
<code>call</code>	The original call to <code>pod</code>
<code>group</code>	The grouping variable

If `mean.function` has any other value, an object of class `c("pod.lm", "lm")` is returned. Since the only method for `pod.lm` objects is a plotting method, except for plotting these objects behave like `lm` objects.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

Cook, R. D. and Weisberg, S. (2004). Partial-one dimensional models. *American Statistician*, 58, 110-116.

Weisberg, S. (2005) *Applied Linear Regression*, third edition. New York: Wiley.

See Also

See Also [nls](#), [lm](#), [nls.control](#)

Examples

```
data(ais)
m1 <- pod(LBM ~ Ht + Wt + RCC, data= ais, group= Sex)
anova(m1) # compare four models
plot(m1) # draw the plot
m2 <- update(m1, mean.function="parallel")
plot(m2, key=c(55, 95))
```

powtran

*Power transformations***Description**

powtran computes members of families of transformations indexed by one parameter, the Box-Cox power family, or the Yeo and Johnson (2000) family, or the basic power family, interpreting zero power as logarithmic. The family can be modified to have Jacobian one, or not, except for the basic power family.

Usage

```
powtran(U, lambda, family, modified)

## Default S3 method:
powtran(U, lambda, family="box.cox", modified=TRUE)

## S3 method for class 'bctrans':
powtran(U, lambda=coef(U), family=U$family, modified=FALSE)

## S3 method for class 'matrix':
powtran(U, lambda, family="power", modified=FALSE)

## S3 method for class 'data.frame':
powtran(U, lambda, family="power", modified=FALSE)
```

Arguments

U	U can be a vector, a matrix, a data.frame, or an object of type bctrans. Missing values are permitted. U must be strictly positive unless family="yeo.johnson".
lambda	The transformation parameter, typically a value between -2 and 2. For U of type matrix, data.frame, or bctrans, lambda must be a vector with the same number of elements as U has columns. lambda has no default, except for objects for type bctrans where the default is the value computed by bctrans.
family	The transformation family to be used. See the details below.
modified	TRUE to divide by the Jacobian, as need to transform the response, FALSE for no division. The default is TRUE except for objects of type bctrans, where the default is FALSE.

Details

The Box-Cox family of *scaled power transformations*, family="box.cox", equals $(U^\lambda - 1)/\lambda$ for $\lambda \neq$ zero, and $\log(U)$ if $\lambda = 0$.

If family="yeo.johnson" then the Yeo-Johnson transformations are used. This is Box-Cox transformation of $U + 1$ for nonnegative values, and of $|U| + 1$ with parameter $2 - \lambda$ for U negative.

The basic power transformation is simply U^λ if lambda not zero, and $\log(U)$ otherwise.

If `modified=TRUE`, then the scaled transformations are divided by the Jacobian, which is a function of the geometric mean.

Value

For `U` a vector, matrix or `data.frame`, returns an object of the same type as `U` of transformed values. For class `bctrans`, returns an object of the same class and shape as `U$X`.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

- Weisberg, S. (2005). *Applied Linear Regression*, third edition. New York: Wiley, Chapter 7.
- Yeo, In-Kwon and Johnson, Richard (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87, 954-959.

See Also

[inv.tran.plot](#), [bctrans](#), [inv.res.plot](#), [tran.family](#).

Examples

```
data(ufcwc)
attach(ufcwc)
powtran(Height, lambda=0)
inv.tran.plot(Dbh, Height, lam = c(-1, 0, 1), family="box.cox")
summary(ans <- bctrans1(cbind(Dbh, Height))) # bctrans1 is used for a matrix input
plot(ans)
```

prodscore

Soil productivity

Description

Soil productivity scores for farms in townships in four counties in the Minneapolis St. Paul metropolitan area, 1981-82. The goal is to see if the productivity score is a good predictor of the assessed value of the farmland. If so, then productivity score could be used to set assessed value for farms enrolled in the “green acres” program that requires that urban farmland be taxed at its agricultural value only without regard to development potential.

Format

This data frame contains the following columns:

County Name of the county

Value Assessed value in dollars per acre.

P Productivity score, a number between 1 and 100.

Year Tax year, either 1981 or 1982.

Source

Douglas Tiffany

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.9.

Examples

```
data(prodscore)
```

```
pure.error.anova
```

Pure Error analysis of variance

Description

For a linear model object, finds the sum of squares for lack of fit and the sum of squares for pure error. These are added to the standard anova table to give a test for lack of fit. If there is no pure error, then the regular anova table is returned.

Usage

```
### This is a generic function.

pure.error.anova(mod)

## S3 method for class 'lm':
pure.error.anova(mod)

### Methods for other than models for type lm have not been defined.
```

Arguments

mod an object of type lm

Details

For regression models with one predictor, say $y \sim x$, this method fits $y \sim x + \text{factor}(x)$ and prints the anova table. With more than one predictor, it computes a random linear combination L of the terms in the mean function and then gives the anova table for `update(mod, ~.+factor(L))`.

Value

Returns an analysis of variance table.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

Weisberg, S. (2005). *Applied Linear Regression*, third edition, New York: Wiley, Chapter 5.

See Also

[lm](#)

Examples

```
x <- c(1, 1, 1, 2, 3, 3, 4, 4, 4, 4)
y <- c(2.55, 2.75, 2.57, 2.40, 4.19, 4.70, 3.81, 4.87, 2.93, 4.52)
m1 <- lm(y~x)
anova(m1) # ignore pure error
pure.error.anova(m1) # include pure error

data(forbes)
m2 <- lm(Lpres~Temp, data=forbes)
pure.error.anova(m2) # function does nothing because there is no pure error
```

random.lin.comb	<i>Compute a random linear combination of the columns of a matrix or data frame</i>
-----------------	---

Description

Computes Xa where X is an $n \times p$ matrix, and a is a random vector of length p .

Usage

```
random.lin.comb(X, seed = NULL)

## S3 method for class 'lm':
random.lin.comb(X, seed=NULL)
```

Arguments

X	An $n \times p$ matrix or data frame. For the <code>lm</code> method, X is a linear regression model, and <code>random.lin.comb</code> is applied to <code>model.matrix(X)</code>
seed	If not NULL, use this to initialize the random number generator

Details

The matrix X is first scaled and centered. a is computed to have random uniform components on $(-1, 1)$.

Value

A vector of length n .

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

See Also

[pure.error.anova](#)

Examples

```
random.lin.comb(matrix(rnorm(9), ncol=3))
```

rat

Rat data

Description

Data collected in an experiment in which rats were injected with a dose of a drug approximately proportional to body weight. At the end of the experiment, the animal's liver was weighed, and the fraction of the drug recovered in the liver was recorded. The experimenter expected the response to be independent of the predictors.

Format

This data frame contains the following columns:

BodyWt BodyWt of the rat

LiverWt LiverWt measured after sacrifice

Dose Dose, roughly proportional to body weight

y dose of drug recovered after sacrifice of the animal

Source

Dennis Cook

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 9.2.3.

Examples

```
data(rat)
pairs(rat)
summary(m1 <- lm(y~BodyWt+LiverWt+Dose, data=rat))
inf.index(m1)
```

```
residual.plots      Residual plots and curvature tests for linear model fits
```

Description

Plots the residuals versus each term in a mean function and versus fitted values. Also computes a curvature test for each of the plots by adding a quadratic term and testing the quadratic to be zero. This is Tukey's test for nonadditivity when plotting against fitted values.

Usage

```
### This is a generic function with only one required argument:

residual.plots (m, ...)

### When the first argument is a linear model (of class lm), the form of the
### function is

## S3 method for class 'lm':
residual.plots(m, vars=~., fitted=TRUE, plot=TRUE,
              layout=NULL, ask, ...)

### The following are three related functions:

resplot(m, varname="tukey", type="pearson",
        plot=TRUE, add.quadratic=TRUE,
        ylab=paste(string.capitalize(type), "Residuals"), ...)

resid.curv.test(m, varname)

tukey.nonadd.test(m)
```

Arguments

m	lm regression object
vars	A one-sided formula that specifies a subset of the predictors. One residual plot is drawn for each column specified. The default <code>~.</code> is to plot against all predictors. For example, the specification <code>vars = ~.-X3</code> would plot against all predictors except for X3.
fitted	If TRUE, the default, plot against fitted values.
tukey	If TRUE, draw plot of residuals versus fitted values and compute Tukey's test of non-additivity.
layout	If set to a value like <code>c(1, 1)</code> or <code>c(4, 3)</code> , the layout of the graph will have this many rows and columns. If not set, the program will select an appropriate layout. If the number of graphs exceed nine, you must select the layout yourself, or you will get a maximum of nine per page.

ask	If TRUE, ask the user before drawing the next plot; FALSE if don't ask.
...	residual.plots passes these arguments to resplot. resplot passes them to plot.
varname	Quoted variable name for the horizontal axis, "tukey" by default for Tukey's test and the plot versus fitted values.
type	Type of residuals to be used. Pearson residuals are appropriate for lm objects since there are equivalent to ordinary residuals with ols and correctly weighted residuals with wls.
ylab	Label for the yaxis. The default is the residual type.
add.quadratic	if TRUE, fits the quadratic regression of the vertical axis on the horizontal axis.
plot	If TRUE, draw the plot(s).

Details

residual.plots draws all residuals plots, versus each term specified first-order term in the model (interactions are automatically skipped) and versus fitted values, and returns all the curvature tests. resplot, which is called by residual.plots, should be viewed as an internal function, and is included here to display its arguments, which can be used with residual.plots as well. resid.curv.test computes the curvature test only. For any factors a boxplot will be drawn.

Value

Returns a data.frame with one row for each plot drawn, one column for the curvature test statistic, and a second column for the corresponding p-value. This function is used primarily for its side effect of drawing residual plots.

Author(s)

Sanford Weisberg, <sandy@stat.umn.edu>

References

S. Weisberg (2005), *Applied Linear Regression*, third edition, Wiley, Chapter 8

See Also

See Also [lm](#)

Examples

```
data(highway)
highway$Sigs <- (round(highway$Sigs*highway$Len)+1)/highway$Len
attach(highway)
d <- data.frame(Rate=Rate, logLen=logb(Len, 2),
               logADT=logb(ADT, 2), logTrks=logb(Trks, 2),
               Slim=Slim, Shld=Shld, logSigs1=logb(Sigs, 2))
attach(d)
m2 <- lm(Rate~logLen+logADT+logTrks+Slim+Shld+logSigs1, d)
residual.plots(m2)
```

`salary`*Salary data*

Description

Salary of faculty in a small Midwestern college in the early 1980s.

Format

This data frame contains the following columns:

Degree Degree, 1 if PhD, 0 if Masters

Rank Rank (1 = Asst Prof, 2 Assoc Prof, 3 Prof)

Sex 1 if female, 0 if male

Year Years in current rank

YSdeg Years since highest degree earned

Salary dollars per year

Source

Sanford Weisberg

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.10.

Examples

```
data(salary)
```

`salarygov`*Government salary study*

Description

Data on non-unionized job classes in a US county in 1986. Included are the job class difficulty score, the number of employees in the class, number of female employees, and the name of the class.

Format

This data frame contains the following columns:

JobClass Name of job class
NW Number of women employees
NE Total number of employees in a job class
Score Difficulty score for job class
MaxSalary Maximum salary for job class

Source

Sanford Weisberg

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 7.3.

Examples

```
data(salarygov)
```

segreg

Energy consumption

Description

Data on electricity consumption (KWH) and mean temperature (degrees F) for one building on the University of Minnesota's Twin Cities campus. for 39 months in 1988-92. The goal is to model consumption as a function of temperature. Higher temperature causes the use of air conditioning, so high temperatures should mean high consumption. This building is steam heated, so electricity is not used for heating.

Format

This data frame contains the following columns:

Temp Monthly mean temperature, degrees F.
C Electricity consumption in KWH/day

Source

Charles Ng

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 11.3.

Examples

```
data (segreg)
```

```
shocks
```

Small electric shocks in dairy cows

Description

Results of a small experiment to learn about the effects of small electric shocks on dairy cows.

Usage

```
data (shocks)
```

Format

A data frame with 6 observations on the following 3 variables.

Intensity Shock level, milliamps

m Number of trials

Y Number of times a positive reaction was observed

Source

R. Norell

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 12.3.

Examples

```
data (shocks)
```

`sigma.hat` *Return the scale estimate for a regression model*

Description

This function provides a consistent method to return the estimated scale from a linear, generalized linear, nonlinear, or other model.

Usage

```
sigma.hat(object)
```

Arguments

`object` A regression object of type `lm`, `glm` or `nls`

Details

For an `lm` or `nls` object, the returned quantity is the square root of the estimate of σ . For a `glm` object, the returned quantity is the square root of the estimated dispersion parameter.

Value

A nonnegative number

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

Examples

```
data(forbes)
m1 <- lm(Lpres ~ Temp, data=forbes)
sigma.hat(m1)
```

`sleep1` *Sleep in mammals*

Description

Includes species averages for 62 mammals.

Format

This data frame uses species as row label and contains the following columns:

- SWS** Slow wave nondreaming sleep, hrs/day
- PS** Paradoxical dreaming sleep, hrs/day
- TS** Total sleep, hrs/day
- BodyWt** Body weight in kg
- BrainWt** Brain weight in g
- Life** Maximum life span, years
- GP** Gestation time, days
- P** Predation index, 1=low,5=hi
- SE** Sleep exposure index 1=exposed, 5=protected
- D** Danger index, 1=least, 5=most

Source

Allison, T. and Cicchetti, D. (1976). Sleep in Mammals: Ecological and Constitutional Correlates *Science*, vol. 194, pp. 732-734.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 6.2.

Examples

```
data(sleep1)
```

snake

Snake river data

Description

The data give the water content of snow and the water yield in inches in the Snake River watershed in Wyoming.

Format

This data frame contains the following columns:

- X** water content of snow
- Y** water yield from April to July

Source

Wilm, H. G. (1950). Statistical control in hydrologic forecasting. "Res. Notes", 61, Pacific Northwest Forest Range Experiment Station, Oregon.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 2.7.3.

Examples

```
data(snake)
```

```
sniffer
```

```
Sniffer data
```

Description

When gasoline is pumped into a tank, hydrocarbon vapors are forced out and into the atmosphere. To reduce this significant source of air pollution, devices are installed to capture the vapor. In testing these vapor recovery systems, a "sniffer" measures the amount recovered. John Rice provided the data for the file sniffer.txt.

Format

This data frame contains the following columns:

TankTemp Initial tank temperature (degrees F)

GasTemp Temperature of the dispensed gasoline (degrees F)

TankPres Initial vapor pressure in the tank (psi)

GasPres Vapor pressure of the dispensed gasoline (psi)

Y Hydrocarbons emitted (grams)

Source

John Rice

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 8.3.2.

Examples

```
data(sniffer)
```

`snowgeese`*Snow geese*

Description

Counts of flocks of snow geese.

Format

This data frame contains the following columns:

photo Photo count

obs1 Observer 1 count (aerial)

obs2 Observer 2 count (aerial)

Source

Dennis Cook

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 5.5.

Examples

```
data(snowgeese)
```

`stopping`*Stopping distances*

Description

Ezekiel and Fox (1959) data on auto stopping distances.

Format

This data frame contains the following columns:

Speed Speed (mph)

Distance Stopping distance (in feet)

Source

Ezekiel, M. and Fox, K. A. (1959). *Methods of Correlation Analysis, Linear and Curvilinear*, New York: Wiley.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.2.

Examples

```
data(stopping)
```

swan96

Black crappie study on Swan Lake, Minnesota

Description

Log catch per unit effort of 200 mm or longer black crappies was recorded 27 times over the course of 1996 on Swan Lake, Minnesota.

Usage

```
data(swan96)
```

Format

A data frame with 27 observations on the following 2 variables.

Day Number of days after June 16, 1996

LCPUE log of the catch of 200 mm or longer black crappies per unit effort (WHAT IS THE BASE?)

Source

Minnesota Department of Natural Resources

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 11.4.

Examples

```
data(swan96)
```

titanic	<i>Titanic</i>
---------	----------------

Description

For each person on board the fatal maiden voyage of the ocean liner Titanic, this dataset records sex, age (adult/child), economic status (first/second/third class, or crew) and whether or not that person survived.

Format

This data frame contains the following columns:

Surv Number of survivors
N Survivors + Deaths
Class Crew or passenger class
Age Adult or child
Sex Male or female

Source

Report on the Loss of the ‘Titanic’ (S.S.) (1990), *British Board of Trade Inquiry Report* (reprint), Gloucester, UK: Allan Sutton Publishing. Taken from the *Journal on Statistical Education Archive*, submitted by rdawson@husky1.stmarys.ca.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 12.2.4.

Examples

```
data(titanic)
```

transact	<i>Transaction data</i>
----------	-------------------------

Description

Data on transaction times in branch offices of a large Australian bank.

Format

This data frame contains the following columns:

T1 number of type 1 transactions
T2 number of type 2 transactions
Time total transaction time, minutes

Source

Cunningham, R. and Heathcote, C. (1989), Estimating a non-Gaussian regression model with multicollinearity. *Australian Journal of Statistics*, 31,12-17.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 4.6.1.

Examples

```
data(transact)
```

turk0	<i>Turkey data, one source</i>
-------	--------------------------------

Description

Turkey weight increase in an experiment in which the supplementation with methionine was varied.

Format

This data frame contains the following columns:

A Amount of methionine supplement (percent of diet)

Gain Pen weight increase (g)

Source

Cook, R. D. and Witmer, J. (1985). A note on the parameter-effects curvature. *Journal of the American Statistical Association*, 80, 872-878.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.13.

Examples

```
data(turk0)
```

`turkey`*Turkey data, all sources*

Description

Data from an experiment on the growth of turkeys. 60 pens of turkeys were grown with a similar diet, supplemented with a dose of methionine from one of three sources. The response is average pen weight. Recorded is dose, source, m, always 5 except for dose=0, average weight gain, and within group SS.

Format

This data frame contains the following columns:

A Dose: Amount of supplement as a percent of the total diet

Gain Ave. weight gain, over all replications

S A factor for the source of methionine, three levels numbers 1, 2 and 3.

m Number of replications or pens

SD SD of the m pens with the same values of S and A.

Source

R. D. Cook and J. Witmer (1985). A note on parameter-effects curvature. *Journal of the American Statistical Association*, 80, 872–878.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 1.1.

Examples

```
data(turkey)
```

`twins`*Burt's twin data*

Description

The given data are IQ scores from identical twins; one raised in a foster home, and the other raised by birth parents.

Format

This data frame contains the following columns:

C Social class, 1=high, 2=medium, 3=low

IQb biological

IQf foster

Source

Burt, C. (1966). The genetic estimation of differences in intelligence: A study of monozygotic twins reared together and apart. *Br. J. Psych.*, 57, 147-153.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.16.

Examples

```
data(twins)
```

ufc

Height-diameter data for Upper Flat Creek, Idaho

Description

These data are forest inventory measures from the Upper Flat Creek stand of the University of Idaho Experimental Forest, dated 1991.

The file `ufc` contains all the data. `ufcwc` contains only Western red cedar. `ufcgf` contains only grand fir.

Usage

```
data(ufcgf)
```

Format

A data frame with the following 5 variables.

Plot Plot number

Tree Tree within plot

Species a factor with levels `DF` = Douglas-fir, `GF` = Grand fir, `SF` = Subalpine fir, `WL` = Western larch, `WC` = Western red cedar, `WP` = White pine

Dbh Diameter 137 cm perpendicular to the bole, mm

Height Height of the tree, in decimeters

Source

Andrew Robinson

References

Weisberg, S. (2005). *Applied Linear Regression*, third edition. New York: Wiley.

Examples

```
data(ufcgf)
```

UN1

National statistics from the United Nations

Description

Demographic data for 193 places, mostly UN members, but also other areas like Hong Kong that are not independent countries.

Usage

```
data(UN1)
```

Format

This data frame uses the locality name as a row label. In some cases the geographic area is smaller than a country; for example Hong Kong. The file contains the following columns:

Fertility Expected number of live births per female, 2000

PPgdp Per capita 2001 GDP, in US \$

Details

These data were collected at published by the UN from a variety of sources. See original source for additional footnotes concerning values for individual countries. Country names are given in the first column of the data file.

Source

<http://unstats.un.org/unsd/demographic>

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 1.3.

Examples

```
data(UN1)
```

UN2

National statistics from the United Nations

Description

National health, welfare, and education statistics for 193 places, mostly UN members, but also other areas like Hong Kong that are not independent countries.

Usage

```
data (UN2)
```

Format

This data frame uses locality name as row labels. In some cases, the geographic area is smaller than a country; for example, Hong Kong. The file contains the following columns:

logFertility log Expected number of live births per female, 2000

logPPGdp Log Per capita 2001 GDP, in US \$

Purban Percent of population that is urban, 2001

Details

These data were collected and published by the UN from a variety of sources. See original source for additional footnotes concerning values for individual countries. Country names are given in the first column of the data file.

Source

<http://unstats.un.org/unsd/demographic>

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 3.1.

Examples

```
data (UN2)
```

Description

National health, welfare, and education statistics for 127 places, mostly UN members, but also other areas like Hong Kong that are not independent countries. Localities with missing data have been removed from this data set.

Format

This data frame uses locality name as row labels. In some cases, the geographic area is smaller than a country; for example, Hong Kong. It contains the following columns:

ModernC Percent of unmarried women using a modern method of contraception

Change Annual population growth rate, percent

PPgdp Per capita 2001 GDP, in US \$

Frate Percent of females over age 15 economically active

Pop Population, thousands

Fertility Expected number of live births per female, 2000

Purban Percent of population that is urban, 2001

Details

These data were collected and published by the UN from a variety of sources. See original source for additional footnotes concerning values for individual countries. Locality names are given in the first column of the data file.

Source

<http://unstats.un.org/unsd/demographic>

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 7.7.

Examples

```
data(UN3)
```

`walleye`*Walleye length at age*

Description

These data give length and age for over 3000 walleye (a type of fish) captured in Butternut Lake, Wisconsin, in three periods with different management method in place.

Usage

```
data(walleye)
```

Format

A data frame with 3198 observations on the following 3 variables.

age Age of the fish, years

length Length, mm

period 1 = pre 1990, 2 = 1991-1996, 3=1997-2000

Source

Michelle LeBeau

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 11.3.

Examples

```
data(walleye)
```

`water`*California water*

Description

Can Southern California's water supply in future years be predicted from past data? One factor affecting water availability is stream runoff. If runoff could be predicted, engineers, planners and policy makers could do their jobs more efficiently. Multiple linear regression models have been used in this regard. This dataset contains 43 years worth of precipitation measurements taken at six sites in the Owens Valley (labeled APMAM, APSAB, APSLAKE, OPBPC, OPRC, and OPSLAKE), and stream runoff volume at a site near Bishop, California.

Format

This data frame contains the following columns:

Year collection year

APMAM Snowfall in inches measurement site

APSAB Snowfall in inches measurement site

APSLAKE Snowfall in inches measurement site

OPBPC Snowfall in inches measurement site

OPRC Snowfall in inches measurement site

OPSLAKE Snowfall in inches measurement site

BSAAM Stream runoff near Bishop, CA, in acre-feet

Source

Source: <http://www.stat.ucla.edu>.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 1.5.

Examples

```
data(dwp)
```

wblake

West Bearskin Lake small mouth bass data.

Description

Data on samples of small mouth bass collected in West Bearskin Lake, Minnesota, in 1991. The file `wblake` includes only fish of ages 8 or younger, while `wblake2` adds a few older fish.

Format

This data frame contains the following columns:

Age Age at capture (yrs)

Length Length at capture (mm)

Scale radius of a key scale, mm

Source

Minnesota Department of Natural Resources

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 1.1.

Examples

```
data(wblake) # excludes fish age 9 or older
data(wblake2) # all fish
```

wml

Simple windmill data

Description

Windspeed data collected at a test site for a windmill, and also at a nearby long-term weather site, in Northern South Dakota. Data collected every six hours for all of 2002, except that all of the month of May and a few other observations are missing.

Usage

```
data(wml)
```

Format

A data frame with 1116 observations on the following 3 variables.

Date A text variable with values like "2002/1/2/6" meaning the reading at 6AM on January 2, 2002

CSpd Windspeed in m/s at the candidate site

RSpd Windspeed for the reference site

Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 2.13.

Examples

```
data(wml)
```

wm2

Windmill data with direction information

Description

Windspeed data collected at a test site for a windmill, and also at a nearby long-term weather site, in Northern South Dakota. Data collected every six hours for all of 2002, except that all of the month of May and a few other observations missing.

Usage

`data(wm2)`

Format

A data frame with 1116 observations on the following 5 variables.

Date A text variable with values like "2002/1/2/6" meaning the reading at 6AM on January 2, 2002

CSpd Windspeed in m/s at the candidate site

RSpd Windspeed for the reference site

RDir Wind direction, in degrees, at the reference site

Bin Wind direction binned into 16 equal width bins

Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.8.

Examples

`data(wm2)`

`wm3`*Binned wind speed data*

Description

For the windspeed data in the file `wm2`, this is the number of observations in each of the bins, and the average windspeed in that bin

Usage

```
data(wm3)
```

Format

A data frame with 16 observations on the following 2 variables.

bin.count Number of observations in the bin

RSpd Average windspeed in the bin

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Problem 6.8.2.

Examples

```
data(wm3)
```

`wm4`*Windmill data with direction information and four reference sites*

Description

Windspeed data collected at a test site for a windmill, and also at four nearby long-term weather site, in Northern South Dakota. Data collected every six hours for all of 2002, except that all of the month of May and a few other observations are missing.

Usage

```
data(wm4)
```

Format

A data frame with 1114 observations on the following 14 variables.

Date A text variable with values like "2002/1/2/6" meaning the reading at 6AM on January 2, 2002

CSpd Wind speed (m/s) at candidate site

Spd1 Wind speed (m/s) at reference site 1

Spd2 Wind speed (m/s) at reference site 2

Spd3 Wind speed (m/s) at reference site 3

Spd4 Wind speed (m/s) at reference site 4

Spd1Lag1 Lag 1 Wind speed (m/s) at reference site 1

Spd2Lag1 Lag 1 Wind speed (m/s) at reference site 2

Spd3Lag1 Lag 1 Wind speed (m/s) at reference site 3

Spd4Lag1 Lag 1 Wind speed (m/s) at reference site 4

Spd1sin1 Spd times sin(direction) at site 1

Spd1cos1 Spd times cos(direction) at site 1

bin1 Wind direction bin for site 1

Dir1 Wind direction for site 1, in degrees

Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 10.4.1.

Examples

```
data(wm4)
```

```
wool
```

```
Wool data
```

Description

This is a three-factor experiment with each factor at three levels, for a total of 27 runs. Samples of worsted yarn were with different levels of the three factors were given a cyclic load until the sample failed. The goal is to understand how cycles to failure depends on the factors.

Format

This data frame contains the following columns:

Len length of specimen (250, 300, 350 mm)

Amp amplitude of loading cycle (8, 9, 10 min)

Load load (40, 45, 50g)

Cycles number of cycles until failure

Source

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). *J. Royal Statist. Soc.*, B26, 211-46.

References

Weisberg, S. (2005). *Applied Linear Regression*, 3rd edition. New York: Wiley, Section 6.3.

Examples

```
data(wool)
```

Index

*Topic **datasets**

ais, 2
allshoots, 3
baeskel, 4
banknote, 5
BGSall, 8
BigMac2003, 9
blowdown, 10
brains, 10
cakes, 11
cathedral, 12
caution, 12
challeng, 13
chloride, 14
cloud, 14
domedata, 18
donner, 19
downer, 20
drugcost, 21
dwaste, 22
florida, 23
forbes, 23
ftcollinssnow, 24
fuel2001, 25
galapagos, 26
galtonpeas, 27
heights, 27
highway, 28
hooker, 29
htwt, 30
jevons, 34
lakemary, 35
lakes, 36
landrent, 37
lathel, 37
mantel, 38
mile, 39
Mitchell, 40
MWwords, 42

npdata, 43
oldfaith, 44
physics, 45
pipeline, 46
prodscore, 53
rat, 56
salary, 59
salarygov, 59
segreg, 60
shocks, 61
sleep1, 62
snake, 63
sniffer, 64
snowgeese, 65
stopping, 65
swan96, 66
titanic, 67
transact, 67
turk0, 68
turkey, 69
twins, 69
ufc, 70
UN1, 71
UN2, 72
UN3, 73
walleye, 74
water, 74
wblake, 75
wm1, 76
wm2, 77
wm3, 78
wm4, 78
wool, 79

*Topic **hplot**

inf.index, 30
inv.res.plot, 31
inv.tran.plot, 33
mmps, 40
pod, 48

- residual.plots, 57
- *Topic **interface**
 - alrweb, 3
- *Topic **manip**
 - random.lin.comb, 55
- *Topic **models**
 - bctrans, 5
 - delta.method, 16
 - plot.bctrans, 47
 - pod, 48
 - pure.error.anova, 54
- *Topic **regression**
 - bctrans, 5
 - conf.intervals, 15
 - delta.method, 16
 - inf.index, 30
 - inv.res.plot, 31
 - inv.tran.plot, 33
 - mmps, 40
 - outlier.t.test, 44
 - plot.bctrans, 47
 - pod, 48
 - powtran, 52
 - pure.error.anova, 54
 - residual.plots, 57
 - sigma.hat, 62
- ais, 2
- allshoots, 3
- alrerrata (*alrweb*), 3
- alrprimer (*alrweb*), 3
- alrweb, 3
- anova.pod (*pod*), 48
- baeskel, 4
- banknote, 5
- bctrans, 5, 48, 53
- bctrans1 (*bctrans*), 5
- BGSall, 8
- BGSboys (*BGSall*), 8
- BGSgirls (*BGSall*), 8
- BigMac2003, 9
- blowAPB (*blowdown*), 10
- blowBF (*blowdown*), 10
- blowdown, 10
- brains, 10
- cakes, 11
- cathedral, 12
- caution, 12
- challeng, 13
- chloride, 14
- cloud, 14
- coef.bctrans (*plot.bctrans*), 47
- coef.pod (*pod*), 48
- conf.intervals, 15
- confint, 16
- cooks.distance, 31
- D, 18
- delta.method, 16
- deviance.pod (*pod*), 48
- df.residual.pod (*pod*), 48
- domedata, 18
- domedata1 (*domedata*), 18
- donner, 19
- downer, 20
- drugcost, 21
- dwaste, 22
- fitted.pod (*pod*), 48
- florida, 23
- forbes, 23
- formula.pod (*pod*), 48
- ftcollinssnow, 24
- fuel2001, 25
- galapagos, 26
- galtonpeas, 27
- heights, 27
- highway, 28
- hooker, 29
- htwt, 30
- inf.index, 30
- inv.res.plot, 7, 31, 48, 53
- inv.tran.estimate, 32
- inv.tran.estimate
 - (*inv.tran.plot*), 33
- inv.tran.plot, 32, 33, 53
- inverse.response.plot, 34
- inverse.response.plot
 - (*inv.res.plot*), 31
- jevons, 34
- lakemary, 35
- lakes, 36

- landrent, 37
- lathel, 37
- lm, 51, 55, 58
- loess, 42
- longshoots (*allshoots*), 3
- lrt.bctrans, 47, 48
- lrt.bctrans (*bctrans*), 5

- mantel, 38
- marginal.model.plot (*mmps*), 40
- mile, 39
- Mitchell, 40
- mmp (*mmps*), 40
- mmps, 40
- MWwords, 42

- nls, 32, 34, 51
- nls.control, 49, 51
- npdata, 43

- oldfaith, 44
- optim, 7
- optimize, 34
- outlier.t.test, 44

- pairs, 7, 48
- physics, 45
- physics1 (*physics*), 45
- pipeline, 46
- plot, 42
- plot.bctrans, 7, 47
- plot.pod (*pod*), 48
- pod, 48
- podnls.fit (*pod*), 48
- podresponse (*pod*), 48
- powtran, 7, 34, 52
- predict.pod (*pod*), 48
- print.bctrans (*bctrans*), 5
- print.pod (*pod*), 48
- prodscore, 53
- pure.error.anova, 54, 56

- random.lin.comb, 55
- rat, 56
- resid.curv.test (*residual.plots*), 57
- resid.pod (*pod*), 48
- residual.plots, 57
- residuals.pod (*pod*), 48

- resplot (*residual.plots*), 57

- salary, 59
- salarygov, 59
- segreg, 60
- shocks, 61
- shortshoots (*allshoots*), 3
- sigma.hat, 62
- sleep1, 62
- snake, 63
- sniffer, 64
- snowgeese, 65
- stopping, 65
- summary.bctrans (*plot.bctrans*), 47
- summary.pod (*pod*), 48
- swan96, 66

- titanic, 67
- tran.family, 53
- transact, 67
- tukey.nonadd.test
 (*residual.plots*), 57
- turk0, 68
- turkey, 69
- twins, 69

- ufc, 70
- ufcdf (*ufc*), 70
- ufcgf (*ufc*), 70
- ufcwc (*ufc*), 70
- UN1, 71
- UN2, 72
- UN3, 73

- vcov.bctrans (*plot.bctrans*), 47
- vcov.pod (*pod*), 48

- walleye, 74
- water, 74
- wblake, 75
- wblake2 (*wblake*), 75
- wm1, 76
- wm2, 77
- wm3, 78
- wm4, 78
- wool, 79