Package 'R330'

February 19, 2015

Type Package

Title An R package for Stats 330

Version 1.0

Date 2012-01-18

Author Alan Lee, Blair Robertson

Maintainer Alan Lee <aj.lee@auckland.ac.nz>

Description This is a collection of useful functions and data for Stats 330

Depends s20x, leaps, rgl, lattice

License GPL-2

LazyLoad yes

Repository CRAN

Date/Publication 2012-06-07 06:08:12

NeedsCompilation no

R topics documented:

`R330-package` *Useful functions for Stats 330*

Description

Contains useful functions and data for the Stats 330 course at the University of Auckland

Details

Package:	R330
Type:	Package
Version:	1.0
Date:	2012-01-13
License:	GPL-2
LazyLoad:	yes

Author(s)

Alan Lee, Blair Robertson

Maintainer: Alan Lee <aj.lee@auckland.ac.nz >

`acc.df` *Data from the Auckland City Council*

Description

Data from the Auckland City Council where the aim was to predict the capital value from the rental value

Usage

```
data(acc.df)
```

Format

A data frame with observations on 96 properties having variables

`capital` the capital value of the property

`rental` the rental value of the property

Examples

```
data(acc.df)
acc.lm<-lm(capital~rental,data=acc.df)
```

acet.df *Acetylene Data*

Description

Percentage conversions of n-heptane to acetylene

Usage

```
data(acet.df)
```

Format

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

`percent.conv` percentage conversaions of n-heptane to acetylene

`temp` temperature in degrees centergrade

`ratio` ratio of H2 to n-heptane

`contact` contact time in seconds

Source

DW Marquardt and RD Snee (1975). Ridge regression in practice. American Statistian, 28, 3-20

Examples

```
data(acet.df)
summary(lm(percent.conv~temp+ratio+contact, data=acet.df))
```

ad.df *Advertising data*

Description

A data set which looks at the relationship between the sales and the expenditure on sales over 36 months

Usage

```
data(ad.df)
```

Format

A data frame with 35 observations on the following 3 variables:

`sales` monthly sales

`spend` amount spent on advertising this month

`prev.spend` amount spent on advertising in the previous month

Details

We lose one observation when prev.spend was created

Examples

```
data(ad.df)
advert.lm<-lm(sales~spend+prev.spend,data=ad.df)
```

---

added.variable.plots *Draws an added variable plot for each independent variable*

Description

Draws an added variable plot for each independent variable

Usage

```
added.variable.plots(f, intercept = TRUE,...)
## S3 method for class 'lm'
added.variable.plots(f, intercept = TRUE,...)
## S3 method for class 'formula'
added.variable.plots(f, intercept = TRUE, data, subset, weights,
 na.action, method = "qr", model = TRUE, x = FALSE, y = FALSE,
 qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)
```

Arguments

`f`	An lm object or model formula
`intercept`	If TRUE (default) an intercept is fitted
`data`	A data frame, list or environment containing the variables in the model.
`subset`	an optional vector specifying a subset of observations to be used in the fitting process.
`weights`	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.

na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y, qr, model	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
singular.ok	logical. If FALSE (the default in S but not in R) a singular fit is an error.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

Returns no value but draws an added variable plot for each variable.

Note

This function redirects to other functions based on the type of object. eg added.variable.plots.lm , added.variable.plots.formula

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(rubber.df)
rubber.lm<-lm(abloss~hardness+tensile,data=rubber.df)
par(mfrow=c(1,2))
added.variable.plots(rubber.lm)
```

`allpossregs` *Calculates all possible regressions*

Description

Calculates all possible regressions for subset selection

Usage

```
allpossregs(f, best = 1, Cp.plot = TRUE, text.cex = 0.8, dp = 3,
       cv.rep = 50, nvmax = 20, ...)
## S3 method for class 'lm'
allpossregs(f, best = 1, Cp.plot = TRUE, text.cex = 0.8, dp = 3,
       cv.rep = 50, nvmax = 20, ...)
## S3 method for class 'formula'
allpossregs(f, best = 1, Cp.plot = TRUE, text.cex = 0.8, dp = 3,
       cv.rep = 50, nvmax = 20, data, subset, weights, na.action,
       method = "qr", model = TRUE, x = FALSE, y = FALSE, qr = TRUE,
       singular.ok = TRUE, contrasts = NULL, offset, ...)
```

Arguments

`f`	an lm object or model formula
`best`	the number of models for each size (size=number of variables) to be printed
`Cp.plot`	print Cp plot? (TRUE=yes, FALSE=no)
`text.cex`	expansion factor for plot text
`dp`	number of decimal places
`cv.rep`	The number of random samplings when calculating the CV estimate of prediction error
`nvmax`	The maximum number of variables to be included in models.
`data`	A data frame, list or environment containing the variables in the model.
`subset`	an optional vector specifying a subset of observations to be used in the fitting process.
`weights`	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
`na.action`	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
`method`	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.

`x, y, qr, model`
: For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value.

  For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.

`singular.ok`
: logical. If FALSE (the default in S but not in R) a singular fit is an error.

`contrasts`
: an optional list. See the contrasts.arg of model.matrix.default.

`offset`
: this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.

`...`
: additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

A matrix with columns labeled:

`rssp`
: Residual Sum of Squares

`sigma2`
: low values indicate better model

`adjRsq`
: adjusted R squared for the model. Big values indicate good model

`Cp`
: Mallow's Cp measure of how well model predicts. Want small values

`AIC`
: Akaike Information Criterion, estimate of the difference between the fitted model and actualy model. Want small values

`BIC`
: Bayesian Information Criterion, estimate of the posterior probability that fitted model is correct one. Want small values

`CV`
: Cross-validation. Small values indicate good model.

`Variables`
: states which variables were included for the regression (val=1 means included)

Rows represent the number of variables in the model

Note

This function redirects to other functions based on the type of object. eg all.poss.regs.lm , all.poss.regs.formula

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(fatty.df)
allpossregs(ffa ~ age + skinfold + weight, data = fatty.df, Cp.plot=TRUE)
```

births.df *Risk factors for low birthweight*

Description

Data were collected at Baystate Medical Center, Springfield, Mass. during 1986, as part of a study to identify risk factors for low-birthweight babies.

Usage

```
data(births.df)
```

Format

A data frame with 189 observations on the following 11 variables:

`id` Identificatin code

`low` low birthweight (defined as less than 2500g) 0 = No, 1 = Yes

`age` Age of mother (years)

`lwt` weight of mother at last menstrual period (pounds)

`race` Race (1 = white, 2 = black, 3 = other)

`smoke` smoking status during pregnancy (0 = No, 1 = Yes)

`ptl` History of premature labour (0 = None, 1 = one, 2 = two, etc.

`ht` History of hypertension (0 = No, 1 = Yes)

`ui` Presence of Uterine Irritability (0 = No, 1 = Yes)

`ftv` Number of Physician Visits during the first trimester (0 = none, 1 = one, 2 = two, etc.)

`bwt` Birth weight (grams) (response)

Source

Hosmer & Lemeshow, Applied Logistic Regression. pp 25-26.

References

Hosmer, D.W. & Lemeshow, S.(2000), Applied Logistic Regression (2nd edition), John Wiley & Sons, New York.

Examples

```
data(births.df)
births.lm<-lm(bwt~age*race*smoke*ui*ht+lwt*race*smoke*ui*ht,data=births.df)
anova(births.lm)
```

boxcoxplot *Draws a Box-Cox plot*

Description

Draws a plot of the Box-Cox profile likelihood.

Usage

```
 boxcoxplot(f, p = seq(-2, 2, length = 20), ...)
 ## S3 method for class 'lm'
boxcoxplot(f, p = seq(-2, 2, length = 20), ...)
 ## S3 method for class 'formula'
boxcoxplot(f, p = seq(-2, 2, length = 20), data, subset,
    weights, na.action, method = "qr", model = TRUE, x = FALSE,
    y = FALSE, qr = TRUE, singular.ok = TRUE, contrasts = NULL,
    offset, ...)
```

Arguments

f	an lm object or a model formula
p	a vector of powers, representing plotting positions along the horizontal axis
data	A data frame, list or environment containing the variables in the model.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y, qr, model	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
singular.ok	logical. If FALSE (the default in S but not in R) a singular fit is an error.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.

offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
...	graphical arguments passed to plot function

Details

The function draws a graph of the negative of the profile likelihood as a function of the transformation power p. The regression coeficients and the standard deiatoin of the eroors have been profiled out. The indicated power is the value of p for which the function attains ist minimum. It may on rare occasions be necessary to adjust the range of p (default is (2,2)).

Value

Returns no value but draws a plot of the Box-Cox profile likelihood.

Note

This function redirects to other functions based on the type of object. eg boxcoxplot.formula, boxcoxplot.lm

Author(s)

Alan Lee, Blair Robertson

References

Box, GEP and Cox, DR. (1964). An analysis of transformations. Journal of the Royal Statistical Society, Series B 26 (2): pp 211-252

Examples

```
data(educ.df)
boxcoxplot(educ~urban + percap + under18, data=educ.df[-50,])

data(wine.df)
boxcoxplot(price~year+temp+h.rain+w.rain,data=wine.df)
```

---

budworm.df *Budworm data*

Description

Data from a small experiment on the toxicity to the tobacco budworm

Usage

```
data(budworm.df)
```

Format

A data frame with 12 observations on the following 4 variables:

`sex` the sex of the budworm

`dose` amount of cypermethrin exposed to

`s` number of budworms affected

`n` total number of budworms

Details

The data come from an experiment on the toxicity to the tobacco budworm Heliothis virescens of doses of the pyrethoid trans-cypermethrin to which the moths were beginning to show resistance. Batches of 20 moths of each sex were exposed for three days to the pyrethroid and the number in each batch that were dead or knocked down was recorded.

Source

Collette, D. (1991) Modelling Binary Data. Chapman and Hall, London. p 75

References

Venables, W.N and Ripley, B. (2002) Modern Applied Statitics with S, Springer, New York.

Examples

```
data(budworm.df)
bugs.glm<-glm(s/n~sex+dose,family=binomial,weights=n,data=budworm.df)
summary(bugs.glm)
```

---

cancer.df *Death Rates for Child Cancer*

Description

Data from study to investigate death rates for child cancer in the UK from 1951-1960 in Northumberland and Durham

Usage

```
data(cancer.df)
```

Format

A data frame with 8 observations on the following 5 variables:

`Cytology` a factor with levels `L` `M`

`Residence` type of residence either rural or urban

`Age` the age of the child classified as either 0-4 years or 5-14 years

`n` number of deaths

`pop` population

Examples

```
data(cancer.df)
cancer.glm<-glm(n ~ Cytology*Residence*Age, family=poisson,
offset=log(pop/100000), data=cancer.df)
anova(cancer.glm, test="Chisq")
```

---

chd.df *Coronary heart disease data*

Description

Shows the age of the subject and presence or absence of evidence of significant coronary heart disease.

Usage

```
data(chd.df)
```

Format

A data frame with 100 observations on the following 2 variables:

`age` age of subject in years

`chd` 0 indicates CHD absent, 1 indicates it is present

Source

Hosmer, and Lemeshow Applied Logistic Regression, pp 2-5

References

Hosmer, D. W., and Lemeshow, S. (2000). Applied Logistic Regression, Second edition, Wiley, New York.

Examples

```
data(chd.df)
chd.glm<-glm(chd~age,family=binomial,data=chd.df)
summary(chd.glm)
```

chem.df *Yield of Chemical Process*

Description

The data was collected to see if how the yields from a particular chemical process were associated with higher or lower flows and conversion percentages

Usage

```
data(chem.df)
```

Format

A data frame with 44 observations on the following 4 variables:

`yield` yield

`conversion` conversion as a percentage

`flow` flow

`ratio` ratio

Examples

```
data(chem.df)
chem.lm<-lm(yield~.,data=chem.df)
summary(chem.lm)
```

cherry.df *Girth, Height and Volume for Black Cherry Trees*

Description

This data set provides measurements of the girth, height and volume of timber in 31 felled black cherry trees. Note that girth is the diameter of the tree (in inches) measured at 4 ft 6 in above the ground.

Usage

```
data(cherry.df)
```

Format

A data frame with 31 observations on the following 3 variables:

`diameter` Tree diameter in inches

`height` Height in ft

`volume` Volume of timber in cubic ft

Source

Ryan, T. A., Joiner, B. L. and Ryan, B. F. (1976) The Minitab Student Handbook. Duxbury Press.

References

Atkinson, A. C. (1985) Plots, Transformations and Regression. Oxford University Press.

Examples

```
data(cherry.df)
cherry.lm =lm(volume~diameter+height,data=cherry.df)
new.df = data.frame(diameter=c(11,12),
             height=c(85,90))
predict(cherry.lm,new.df)
```

---

chickwts.df *Chicken Weights Data*

Description

An experiment comparing 12 methods of feeding chickens was done independently in two replicaes arranged in different houses

Usage

```
data(chickwts.df)
```

Format

A data frame with 24 observations on the following 4 variables:

`chickweight` weight gain

`protein` form of protien, either groundnut or soyabean

`protlevel` level of protein either 0, 1 or 2

`fish` level of fish solubles, either high or low

Source

John, J.A. and Quenouille, M.H. (1977). Experiments: Design and Analysis, 2nd edn. London: Griffin.

References

Cox, D. R. & Snell, E. J. (1981). Applied Statistics: Principles and Examples. Chapman and Hall, London.

Examples

```
data(chickwts.df)
model1<-lm(chickweight~protein*protlevel*fish, data=chickwts.df)
summary(model1)
```

---

coag.df *Blood Coagulation Data*

Description

Experiment designed to see the effect of four different diets on a subject's blood coagulation time

Usage

```
data(coag.df)
```

Format

A data frame with 24 observations on each of the two variables

`coag` Blood coagulation time

`diet` Which diet they were on either A, B, C or D

Source

Box, G.E.P., Hunter, J.S. Hunter, W.G. Statistics for experimenters, pp 165-197.

References

Box, G.E.P., Hunter, J.S. Hunter, W.G. (1978). Statistics for experimenters, Wiley, New York.

Examples

```
data(coag.df)
coag.lm <- lm(coag ~ diet,data = coag.df)
anova(coag.lm)
```

cross.val *Calculates cross-validated estimates of prediction error*

Description

For a logistic model, calculates cross-validated estimates of specificity, sensitivity and percentage correctly classified. For a Gaussian model, calculates a cross-validated estimate of prediction error.

Usage

```
 cross.val(f,  nfold = 10, nrep = 20, ...)
 ## S3 method for class 'lm'
cross.val(f,  nfold = 10, nrep = 20, ...)
 ## S3 method for class 'glm'
cross.val(f,  nfold = 10, nrep = 20, ...)
 ## S3 method for class 'formula'
cross.val(f, nfold = 10, nrep = 20,  family = gaussian,
     data, weights, subset, na.action, start = NULL, etastart,
     mustart, offset, control = list(...), model = TRUE,
     method = "glm.fit", x = FALSE, y = TRUE, contrasts = NULL, ...)
```

Arguments

f	an lm object, a glm object or a model formula
nfold	number of parts data set divided into
nrep	number of random spilts
family	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
data	A data frame, list or environment containing the variables in the model.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
start	starting values for the parameters in the linear predictor.
etastart	starting values for the linear predictor.
mustart	starting values for the vector of means.

`offset`	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
`control`	a list of parameters for controlling the fitting process. For glm.fit this is passed to glm.control.
`model`	a logical value indicating whether model frame should be included as a component of the returned value.
`method`	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
`x, y`	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
`contrasts`	an optional list. See the contrasts.arg of model.matrix.default.
`...`	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Details

The object returned depends on the class of the input.

Value

For a logistic model:

`Mean Specificity`
: false negatives

`Mean Sensitivity`
: false positives

`Mean Correctly classified`
: proportion correctly classified

For a Gaussian model, an estimate of the root mean squared error is returned.

Note

This function redirects to other functions based on the type of object. eg cross.val.glm , cross.val.formula

Author(s)

Alan Lee, Blair Robertson

References

Bradley Efron and Robert Tibshirani (1993). An Introduction to the Bootstrap. Chapman and Hall, London.

Examples

```
data(fatty.df)
fatty.lm <- lm(ffa~age+weight+skinfold, data=fatty.df)
cross.val(fatty.lm)
#
data(drug.df)
cross.val(DFREE ~ NDRUGTX + factor(IVHX) + AGE + TREAT, family=binomial,
     data=drug.df)
```

---

cycles.df *Cycles to failure of worsted yarn*

Description

Data which looked at the number of cycles to failure of lengths of worsted yarm under cycles of repeated loading

Usage

```
data(cycles.df)
```

Format

A data frame with 27 observations on the following 4 variables:

`yarn.length` factor which takes lengths 250, 300 or 350mm, which were classified as low, med or high respectively

`amplitude` the cycle amplitudes were taken to be 8, 9 or 10mm, which were classified as low, med or high respectively

`load` loads taken to be 40, 45 or 50 grams, which were classified as low, med or high respectively

`cycles` number of cycles to failure

Source

Cox and Snell, Applied Statistics: Principles and Examples, pp 98-102

References

Cox, D. R. & Snell, E. J. (1981). Applied Statistics: Principles and Examples. Chapman and Hall, London.

Examples

```
data(cycles.df)
library(lattice)
dotplot(cycles~yarn.length|amplitude*load,xlab="Yarn length",
 ylab="Cycles to failure",data=cycles.df,
strip=function(...)strip.default(...,strip.names=TRUE))
```

---

diets.df *Weight gains of rats*

Description

Data from an experiment on the weight gain of rats based on the source of protein and the amount

Usage

```
data(diets.df)
```

Format

A data frame with 60 observations on the following 3 variables:

`gain` amount of weight gained

`source` source of protein either beef, pork or cereal

`level` amount of protein given, high or low

Source

Snedecor, G.W., and Cochran, W.G., (1989), Statistical Methods.

Examples

```
data(diets.df)
plot.design(diets.df)
```

drug.df *Drug addiction data*

Description

Data from University of Massachusetts AIDS Research Unit IMPACT study, a medical study performed in the US in the early 90' s. The study aimed to evaluate two different treatments for drug addiction.

Usage

```
data(drug.df)
```

Format

A data frame with 575 observations on the following 9 variables:

`ID` Identification Code

`AGE` Age at enrollment

`BECK` Back depression score

`IVHX` IV drug use history at admission (1=never, 2=previous, 3=recent)

`NDRUGTX` number of prior treatments

`RACE` subjects race (0 = white, 1 = other)

`TREAT` treatment duration (0 = short, 1 = long)

`SITE` treatment site (0 = A, 1 = B)

`DFREE` Remained drug free (1 = Yes, 0 = No) (response)

Source

Hosmer and Lemeshow, Applied Logistic Regression (2nd Ed), p28

References

Hosmer, D.W. and Lemeshow, S. (2000), Applied Logistic Regression (2nd Ed), Wiley, New York.

Examples

```
data(drug.df)
cross.val(DFREE ~ NDRUGTX + factor(IVHX) + AGE + TREAT, data = drug.df)
```

educ.df *Educations expenditure data*

Description

Data set from 50 US states on education expenditure

Usage

```
data(educ.df)
```

Format

A data frame with 50 observations on the following 4 variables:

`urban` number of residents per 1000 in urban areas

`educ` per capita expenditure on education (response)

`percap` per capita income

`under18` number of residents per 1000 under 18

Examples

```
data(educ.df)
educ.lm = lm(educ~urban + percap + under18, data=educ.df)
summary(educ.lm)
```

err.boot *Calculates a bootstrap estimate of prediction error*

Description

Calculates the training set prediction error and a bootstrap estimate of prediction error for the model in specified by a formula, lm object or glm object

Usage

```
err.boot(f, B=500, ...)
## S3 method for class 'lm'
err.boot(f, B=500, ...)
## S3 method for class 'glm'
err.boot(f, B=500, ...)
## S3 method for class 'formula'
err.boot(f, B=500,  family = gaussian, data, weights,
   subset, na.action, start = NULL, etastart, mustart, offset,
   control = list(...), model = TRUE, method = "glm.fit", x = FALSE,
   y = TRUE, contrasts = NULL, ...)
```

Arguments

f	An lm object, a glm object or a model formula
B	number of bootstrap replications
family	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
data	A data frame, list or environment containing the variables in the model.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
start	starting values for the parameters in the linear predictor.
etastart	starting values for the linear predictor.
mustart	starting values for the vector of means.
offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
control	a list of parameters for controlling the fitting process. For glm.fit this is passed to glm.control.
model	a logical value indicating whether model frame should be included as a component of the returned value.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

$err	Training set estimate
$Err	Bootstrap estimate

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(drug.df)
err.boot(DFREE ~ NDRUGTX + factor(IVHX) + AGE + TREAT, data= drug.df)
```

`ethanol.df` *Engine exhaust fumes from burning ethanol*

Description

Ethanol fuel was burned in a single-cylinder engine. For various settings of the engine compression and equivalence ratio, the emissions of nitrogen oxides were recorded.

Usage

```
data(ethanol.df)
```

Format

A data frame with 88 observations on the following 3 variables:

`NOx` Concentration of nitrogen oxides, NO and NO2, in micrograms per Joule

`C` Compression ratio of the engine

`E` Equivalence ratio - a measure of the richness of the air and ethanol fuel mixture.

Source

Brinkman, N.D. (1981) Ethanol Fuel: A Single - Cylinder Engine Study of Efficiency and Exhaust Emissions. SAE transactions, 90, 1410 - 1424.

References

Cleveland, William S. (1993) Visualizing Data. Hobart Press, Summit, New Jersey.

Examples

```
data(ethanol.df)
pairs20x(ethanol.df)
```

evap.df *Moisture evaporation data*

Description

The purpose was to see if the amount of evaporation could be predicted by the temperature humidity and wind speed

Usage

```
data(evap.df)
```

Format

A data frame with 46 observations on the following 11 variables:

`avst` average soil temperature over 24 hour period (x10)

`minst` minimum soil temperature over 24 hour period (x10)

`maxst` maximum soil temperature over 24 hour period (x10)

`avat` average air temperature over 24 hour period (x10)

`minat` minimum air temperature over 24 hour period (x10)

`maxat` maximum air temperature over 24 hour period (x10)

`avh` average humidity over 24 hour period (x10)

`minh` minimum humidity over 24 hour period (x10)

`maxh` maximum humidity over 24 hour period (x10)

`wind` average wind speed over a 24 hour period (x100)

`evap` amount of evaporation over 24 hour period

Examples

```
data(evap.df)
evap.lm<-lm(evap~avat+avh+wind,data=evap.df)
summary(evap.lm)
```

fatty.df *Fatty acid data*

Description

Data was collected to use physical measures to model a biochemical parameter in overweight children

Usage

```
data(fatty.df)
```

Format

A data frame with 20 observations on the following 4 variables:

`ffa` free fatty acid level in blood (response)

`age` age (months)

`weight` weight (pounds)

`skinfold` skinfold thickness (inches)

Examples

```
data(fatty.df)
fatty.lm<-lm(ffa~age+weight+skinfold,data=fatty.df)
```

funnel *Plots for checking for unequal variances*

Description

Plots for checking for unequal variances

Usage

```
funnel(f,...)
## S3 method for class 'lm'
funnel(f,...)
## S3 method for class 'formula'
 funnel(f, data, subset, weights, na.action, method = "qr",
 model = TRUE, x = FALSE, y = FALSE, qr = TRUE, singular.ok = TRUE,
 contrasts = NULL, offset, ...)
```

Arguments

f	an lm object or a model formula
data	A data frame, list or environment containing the variables in the model.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y, qr, model	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
singular.ok	logical. If FALSE (the default in S but not in R) a singular fit is an error.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

Prints the slope of the line of best fit for log std.errors vs log means. Returns (invisibly) the estimated variances of the observations, and draws (i) a plot of log standard deviations versus log means, and (ii) a plot of the smoothed squared residuals.

Note

This function redirects to other functions based on the type of object. eg funnel.lm , funnel.formula

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(educ.df)
educ50.lm = lm(educ~urban + percap + under18, data=educ.df, subset=-50)
funnel(educ50.lm)
#
funnel(educ~urban + percap + under18, data=educ.df, subset=-50)
```

HLstat *Performs a Hosmer-Lemeshow test*

Description

Calculates and prints a chi^2 statistic and a p-value for the Hosmer-Lemeshow test.

Usage

```
HLstat(f,...)
## S3 method for class 'glm'
HLstat(f,...)
## S3 method for class 'formula'
 HLstat(f, family = binomial, data = data, weights, subset,
    na.action, start = NULL, etastart, mustart, offset,
    control = list(...), model = TRUE, method = "glm.fit",
    x = FALSE, y = TRUE, contrasts = NULL, ...)
```

Arguments

f	a glm object or model formula
family	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
data	A data frame, list or environment containing the variables in the model.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
start	starting values for the parameters in the linear predictor.
etastart	starting values for the linear predictor.
mustart	starting values for the vector of means.

offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
control	a list of parameters for controlling the fitting process. For glm.fit this is passed to glm.control.
model	a logical value indicating whether model frame should be included as a component of the returned value.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

HL stat: the value of the Hosmer-Lemeshow test P-value: an approximate p-value for the test

Note

This function redirects to other functions based on the type of object. eg HLstat.glm , HLstat.formula

Author(s)

Alan Lee, Blair Robertson

Source

Hosmer and Lemeshow, Applied Logistic Regression (2nd Ed), p 147.

References

Hosmer, D.W. and Lemeshow, S. (2000), Applied Logistic Regression (2nd Ed), Wiley, New York.

Examples

```
data(drug.df)
drug.glm<-glm(DFREE ~ . - IVHX - ID + factor(IVHX), family = binomial,
           data = drug.df)
HLstat(drug.glm)
```

housing.df *Housing conditions satisfaction*

Description

A data set investingating the satisfaction with housing conditions in Copenhagen

Usage

```
data(housing.df)
```

Format

A data frame with 18 observations on the following 4 variables:

`sat` Satisfaction of householders with their present housing circumstances, (High, Medium or Low, ordered factor).

`infl` Perceived degree of influence householders have on the management of the property (High, Medium, Low).

`cont` Contact residents are afforded with other residents, (Low, High).

`count` number in each category

Source

Madsen, M. (1976). Statistical analysis of multiple contingency tables. Two examples. Scand J. Statist., 3,97-106

References

Cox, D. R. & Snell, E. J. (1981). Applied Statistics: Principles and Examples. Chapman and Hall, London.

Examples

```
data(housing.df)
housing.glm<-glm(count~sat*infl*cont, family=poisson, data=housing.df)
anova(housing.glm, test="Chisq")
```

`influenceplots` *Draws plots of influence measures*

Description

Draws plots of influence measures based on the family.

Usage

```
influenceplots(f, cex.lab=0.7, ...)
## S3 method for class 'lm'
influenceplots(f, cex.lab=0.7, ...)
## S3 method for class 'glm'
influenceplots(f, cex.lab=0.7, ...)
## S3 method for class 'formula'
influenceplots(f, cex.lab=0.7,  family = gaussian, data,
      weights, subset, na.action, start = NULL, etastart, mustart,
      offset, control = list(...), model = TRUE, method = "glm.fit",
      x = FALSE, y = TRUE, contrasts = NULL, ...)
```

Arguments

`f`	a lm object, a glm object or a model formula
`cex.lab`	An expansion factor for plot labels
`family`	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
`data`	A data frame, list or environment containing the variables in the model.
`weights`	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
`subset`	an optional vector specifying a subset of observations to be used in the fitting process.
`na.action`	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
`start`	starting values for the parameters in the linear predictor.
`etastart`	starting values for the linear predictor.
`mustart`	starting values for the vector of means.
`offset`	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.

control	a list of parameters for controlling the fitting process. For glm.fit this is passed to glm.control.
model	a logical value indicating whether model frame should be included as a component of the returned value.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Details

For Gaussian models, the function plots the influence measures calculated by the R function `influence.measures`. These include the DFBETAS for each coefficient, DDFITS, COVRATIO, Cook's D, and the hat matrix diagonals. Set the plot layout accordingly with par. For logistic models, four plots are produced: index plots of the deviance and Pearson residuals, Cook's D and the leave-one-out deviance change.

Value

Draws the plots but returns no value.

Note

This function redirects to other functions based on the type of object. eg influence.plots.lm , influence.plots.formula , influence.plots.glm

Author(s)

Alan Lee, Blair Robertson

References

Chambers, J. M. (1992) Linear models. Chapter 4 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Examples

```
data(educ.df)
educ.lm = lm(educ~urban + percap + under18, data=educ.df)
par(mfrow=c(2,4))
influenceplots(educ.lm)
#
influenceplots(educ~urban + percap + under18, data=educ.df)
```

`ingots.df` *Unreadiness for rolling of metal ingots*

Description

An experiment testing metal ingots prepared with different soaking times and heats. The number which were not ready were counted

Usage

```
data(ingots.df)
```

Format

A data frame which shows a condensed table.

`heat` heating times for ingots either 7, 24, 27 or 51

`soak` soaking times for ingots either 1.0, 1.7, 2.2, 2.8 or 4.0

`notready` the number of ingots not ready for rolling

`total` the total number of ingots which were tested under that set of conditions

Source

D.R.Cox. (1970) The Analysis of Binary Data, p. 11

References

D.R.Cox. (1970) The Analysis of Binary Data, Chapman and Hall, London

Examples

```
data(ingots.df)
ingots.glm<-glm(cbind(notready, total-notready)~heat +
   soak, weight=total, family = binomial, data = ingots.df)
summary(ingots.glm)
```

kyphosis.df *Data on Children who have had Corrective Spinal Surgery*

Description

The kyphosis data frame has 81 rows and 4 columns, representing data on children who have had corrective spinal surgery.

Usage

```
data(kyphosis.df)
```

Format

A data frame with 81 observations on the following 4 variables:

`Kyphosis` a factor with levels absent present indicating if a kyphosis (a type of deformation) was present after the operation.

`Age` in months

`Number` the number of vertebrae involved

`Start` the number of the first (topmost) vertebra operated on

Source

John M. Chambers and Trevor J. Hastie, Statistical Models in S, p 200.

References

John M. Chambers and Trevor J. Hastie eds. (1992) Statistical Models in S, Wadsworth and Brooks/Cole, Pacific Grove, CA.

Examples

```
data(kyphosis.df)
pairs20x(kyphosis.df)
```

lizard.df *Lizard data*

Description

Site preferences of two species of lizard, grahami and opalinus

Usage

```
data(lizard.df)
```

Format

A data frame with 12 observations on the following 5 variables:

`length` perch lenght (short, long)

`height` perch height (high, low)

`time` time of day (early, late, mid)

`r` number of grahami lizards

`n` total number of lizards

Source

Schoener, T. W. (1970) Nonsynchronous spatial overlap of lizards in patchy habitats. Ecology 51, 408-418.

References

McCullagh, P. and Nelder, J. A. (1989.) Generalized Linear Models (2nd Edition). Chapman and Hall, London.

Examples

```
data(lizard.df)
plot.design(lizard.df, y=log(lizard.df$r
/(lizard.df$n-lizard.df$r)), ylab="mean of logits")
```

metal.df *Metal Removal*

Description

Data from an experiment to measure the rate of metal removal in a machining process on a lathe

Usage

```
data(metal.df)
```

Format

A data frame with 15 observations on the following 3 variables:

`hardness` hardness of the material being machined

`setting` speed setting of the lathe (fast, medium or slow

`rate` rate of metal removal (response)

Examples

```
data(metal.df)
med <-ifelse(metal.df$setting=="medium", 1,0)
slow<-ifelse(metal.df$setting=="slow", 1,0)
summary(lm(rate~med + slow + hardness, data=metal.df))
```

mines.df *Mining accident data*

Description

A data set with the number of accidents per mine in a 3 month period in 44 coal mines in West Virginia

Usage

```
data(mines.df)
```

Format

A data frame with 44 observations on the following 5 variables:

`COUNT` number of accidents (response)

`INB` inner burden thickness

`EXTRP` percentage of coal extracted from mine

`AHS` the average height of the coal seam in the mine

`AGE` the age of the mine

Examples

```
data(mines.df)
mines.glm<-glm(COUNT ~ INB + EXTRP + AHS + AGE,
               family=poisson, data=mines.df)
```

`onions.df` *Onion growing data*

Description

From an onion growing experiment at the Hort Research station at Pukekohe. Conducted to compare the effects of different curing methods on the number of skins on individual onions at different stages of maturity

Usage

```
data(onions.df)
```

Format

A data frame with 300 observations on the following 5 variables:

`maturity` Maturity of the onion as a percentage (50,70,90,95,100)

`cure` method of curing (traditional, shears or partial)

`block` the area of land the onions were grown in (1,2,3,4)

`skins` the number of skins

`weight` a numeric vector

Source

C.M. Triggs, personal communication

Examples

```
data(onions.df)
onions.glm<-glm(skins ~ factor(block),
      family=poisson, weight=weight, data=onions.df)
```

plum.df *Plum tree data*

Description

A study was conducted on the reproduction of plum trees by taking cuttings from older trees. Half the cuttings were planted immediately while the other half were bedded in sand until spring when they were planted. Two lengths of cuttings were used: long (12 cm) and short (6cm). A total of 240 cuttings were taken for each of the 4 combinations of planting time and cutting length and the number of cuttings that survived in each situation was recorded.

Usage

```
data(plum.df)
```

Format

A data frame with 4 observations on the following 4 variables:

`length` cutting length (long,short)

`time` planting time (spring, autumn)

`s` number that survived

`n` total number planted

Examples

```
data(plum.df)
plum.glm<-glm(cbind(s,n-s)~length*time, family=binomial, data=plum.df)
summary(plum.glm)
```

rats.df *Rat growth data*

Description

Each rat in the data set was measured on 11 dates expressed as days from start of the experiment. The purpose was to see if the growth rate was the same for each group.

Usage

```
data(rats.df)
```

Format

A data frame with 176 observations on the following 5 variables:

`growth` a numeric vector

`group` a numeric vector

`rat` a numeric vector

`change` a numeric vector

`day` days since the start of experiment

Details

Hand and Crowder (1996) describe data on the body weights of rats measured over 64 days. These data also appear in Table 2.4 of Crowder and Hand (1990). The body weights of the rats (in grams) are measured on day 1 and every seven days thereafter until day 64, with an extra measurement on day 44. The experiment started several weeks before "day 1." There are three groups of rats, each on a different diet.

Source

Pinheiro, J. C. and Bates, D. M. (2000), Mixed-Effects Models in S and S-PLUS, Springer, New York. (Appendix A.3)

Crowder, M. and Hand, D. (1990), Analysis of Repeated Measures, Chapman and Hall, London.

Hand, D. and Crowder, M. (1996), Practical Longitudinal Data Analysis, Chapman and Hall, London.

Examples

```
data(rats.df)
group.vec<-as.numeric(rats.df$group)
# convert group from factor to vector
plot(growth~day,type="n",
data=rats.df,
xlab="days elapsed",
ylab="Weight (grams)",
main="Rat Growth rates")
for(i in (0:15)){
index<-(1:11) + i*11
lines(rats.df$day[index],rats.df$growth[index],
lty=group.vec[index[1]])
}
legend(45,400,paste("Group",1:3),lty=c(1,2,3))
```

reg3d *3d plot of data*

Description

plots 3 variables on x,y,z axes and draws fitted plane

Usage

```
reg3d(data, wire = FALSE)
```

Arguments

data	a data frame, with the first three columns containing the values of x, y and z
wire	If TRUE draws a grid as a reference plane, if FALSE returns a solid reference plane

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(fatty.df)
reg3d(fatty.df,wire=TRUE)

data(rubber.df)
reg3d(rubber.df)
```

ROC.curve *draws an ROC curve*

Description

Draws an ROC curve and calculates the area under the curve

Usage

```
ROC.curve(f, ...)
## S3 method for class 'glm'
ROC.curve(f, ...)
## S3 method for class 'formula'
ROC.curve(f, family = gaussian, data, weights, subset, na.action,
   start = NULL, etastart, mustart, offset, control = list(...),
   model = TRUE,  method = "glm.fit", x = FALSE, y = TRUE,
   contrasts = NULL, ...)
```

Arguments

f	a glm object or a model formula
family	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
data	A data frame, list or environment containing the variables in the model.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
start	starting values for the parameters in the linear predictor.
etastart	starting values for the linear predictor.
mustart	starting values for the vector of means.
offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
control	a list of parameters for controlling the fitting process. For glm.fit this is passed to glm.control.
model	a logical value indicating whether model frame should be included as a component of the returned value.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Details

If the formula version is used, an error will occur unless family=binomial

Value

Returns no value but prints the area under the ROC curve and draws the curve

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(drug.df)
ROC.curve(DFREE ~ NDRUGTX + factor(IVHX) + AGE + TREAT, family=binomial,
     data= drug.df)
```

---

rubber.df *Rubber Specimen Data*

Description

Thirty rubber specimens were rubbed with an abrasive metal.

Usage

```
data(rubber.df)
```

Format

A data frame with 30 observations on the following 3 variables:

`hardness` Hardness in degrees of Shore

`tensile` strength in kilograms per square centimetre

`abloss` the amount of material rubbed off in grams per horsepower-hour

Source

GEP Box, WR Cousins, FR Hindsworth, H Heeny, M Milbourne, W Spendley and WL Stevens (1957). In OL Davies (Ed.) Statistical Methods in Research and Production, 3rd Ed. Oliver and Boyd, London.

References

Chambers, J. M. et al. (1983). Graphical Methods for Data Analysis. Duxbury Press: Boston.

Examples

```
data(rubber.df)
rubber.lm<-lm(abloss~hardness+tensile,data=rubber.df)
pred<-fitted.values(rubber.lm)
res<-residuals(rubber.lm)
plot(pred,res)
```

salary.df *Study of Supervisor Performance*

Description

A survey of the clerical employment of a large financial organisation included questions related to employee satisfaction with their supervisors, designed to determine the overall effectiveness of the supervisor

Usage

```
data(salary.df)
```

Format

A data frame with 31 observations on the following 6 variables:

`X1` Handles employee conplaints

`X2` Does not allow special privileges

`X3` Opportunity to learn new things

`X4` Raises based on performances

`X5` Too critical of poor performances

`Y` Overall rating of job being done by supervisor

Source

S. Chatterjee, A.S. Hadi and B. Price, Regression Analysis by Example, p56

References

S. Chatterjee, A.S. Hadi and B. Price, (2000). Regression Analysis by Example (3rd Ed), Wiley, New York.

Examples

```
data(salary.df)
salary.lm<-lm(Y~X1+X2+X3+X4+X5,data=salary.df)
resids<-residuals(salary.lm)
pred<-fitted.values(salary.lm)
plot(pred,resids,type="n")
ncases<-length(resids)
text(pred,resids,1:ncases)
```

sport.df *Australian Institute of Sport*

Description

Data on 102 male and 100 female athletes collected at the Australian Institute of Sport, courtesy of Richard Telford and Ross Cunningham.

Usage

```
data(sport.df)
```

Format

A data frame with 158 observations on the following 5 variables:

`ID` ID

`sex` male or female

`sport` Sport

`BMI` Body mass index = weight/height^2

`X.Bfat` percentage body fat

Source

Richard Telford and Ross Cunningham, Australian National University.

References

Cook, R. D., and Weisberg, S. (1994). An Introduction to Regression Graphics. Wiley, New York.

Examples

```
data(sport.df)
library(lattice)
xyplot(X.Bfat~BMI|sport*sex,xlab="BMI",ylab="X.Bfat",data=sport.df)
```

stamford.df *Maximum Daily Ozone Concertrations*

Description

Daily maximum ozone concertrations at Stamford, Connecticut and Yonkers, New York, during the period 1 May 1974 to 30 September 1974.

Usage

```
data(stamford.df)
```

Format

A data frame with 136 observations on the following 2 variables:

`days` denotes which day observation occured on

`ozone` ozone in parts per billion

Source

Chambers, J. M. et al. Graphical Methods for Data Analysis. p346

References

Chambers, J. M. et al. (1983). Graphical Methods for Data Analysis. Duxbury Press: Boston.

Examples

```
data(stamford.df)
plot(stamford.df$days, stamford.df$ozone, xlab="Days", ylab="Ozone")
loess.stuff=loess(ozone~days, data=stamford.df, span=0.75)
lines(loess.stuff$x, loess.stuff$fitted)
```

test.lc *tests hypothesis cc^Tb=c*

Description

Given a linear model with coefficient vector b tests the linear hypothesis cc^Tb=c

Usage

```
test.lc(f, cc, c, ...)
## S3 method for class 'lm'
test.lc(f, cc, c, ...)
## S3 method for class 'glm'
test.lc(f, cc,c, ...)
## S3 method for class 'formula'
test.lc(f, cc, c, family = gaussian, data, weights, subset,
      na.action, start = NULL, etastart, mustart, offset,
      control = list(...), model = TRUE, method = "glm.fit",
      x = FALSE, y = TRUE, contrasts = NULL, ...)
```

Arguments

f	a glm object or a model formula
cc	a vector containing the coefficients of the linear combination
c	hypothetical value of the combination
family	a description of the error distribution and link function to be used in the model. This can be a character string naming a family function, a family function or the result of a call to a family function. (See family for details of family functions.)
data	A data frame, list or environment containing the variables in the model.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
start	starting values for the parameters in the linear predictor.
etastart	starting values for the linear predictor.
mustart	starting values for the vector of means.
offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
control	a list of parameters for controlling the fitting process. For glm.fit this is passed to glm.control.
model	a logical value indicating whether model frame should be included as a component of the returned value.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.

x, y	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

$est	gives estimate for cc
$std.err	gives standard error for cc
$df	degrees of freedom
$t.stat	gives the t stat for the hypothesis test
$p.val	gives p-value for the hypothesis test

Note

Redirects based on the type of object

Author(s)

Alan Lee, Blair Robertson

Examples

```
data(cherry.df)
cherry.lm = lm(log(volume)~log(diameter)+log(height),data=cherry.df)
cc = c(0,1,1)
c = 3
test.lc(cherry.lm, cc, c)
```

---

traffic.df *Highway accdient rates*

Description

Data gathered in the course of studying the relationship between accident rates on US highways and various characteristics of the highway

Usage

```
data(traffic.df)
```

Format

A data frame with 39 observations on the following 15 variables:

`obs` observation number

`rate` The accident rate per million vehicle miles (response)

`len` The length of the segment of highway in miles

`adt` Average daily traffic count ('000)

`trks` The percentage of the traffic that are trucks

`slim` the speed limit in mph

`lwd` the lane width in feet

`shld` the shoulder width in feet

`itg` number of freeway interchanges per mile

`sigs` number of entrances controlled by lights per mile

`acpt` number of access points per mile

`lane` number of lanes in each direction

`fai` dummy variable, equal to 1 if an interstate highway, zero otherwise

`pa` equal to 1 if a principal highway, 0 otherwise

`ma` equal to 1 if a major highway, 0 otherwise

Source

Carl Hoffstedt. This differs from the dataset highway in the alr3 package only by transformation of some of the columns.

References

Fox, J. and Weisberg, S. (2011) An R Companion to Applied Regression, Second Edition, Sage.

Weisberg, S. (2005) Applied Linear Regression, Third Edition. Wiley, Section 7.2.

Examples

```
data(traffic.df)
traffic.lm<-lm(rate~.,data=traffic.df)
summary(traffic.lm)
```

vapour.df *Hydrocarbon data*

Description

When petrol is pumped into a tank, hydrocarbon vapours are forced into the atmosphere. To reduce this significant source of air pollution, devices are installed to capture the vapour. A laboratory experiment was conducted in which the amount of vapour given off was measured.

Usage

```
data(vapour.df)
```

Format

A data frame with 125 observations on the following 5 variables:

`t.temp` initial tank temperature (degrees F)

`p.temp` temperature of the dispensed petrol (degrees F)

`t.vp` initial vapour pressure in tank (psi)

`p.vp` vapour pressure of the dispensed petrol (psi)

`hc` emitted dispensed hydrocarbons (g)(response)

Examples

```
data(vapour.df)
vapour.lm<-lm(hc~ t.temp + p.temp + t.vp + p.vp, data=vapour.df)
summary(vapour.lm)
```

vaso.df *vaso-constriction data*

Description

Data from a study of reflex vaso-constriction (narrowing of the blood vessels) of the skin of the fingers

Usage

```
data(vaso.df)
```

Format

A data frame with 39 observations on the following 3 variables.

`Volume` volume of air breathed in

`Rate` rate of intake of breath

`Response` 1 = vaso-constriction occurs, 0 = doesn't occur

Source

Finney, D. J. (1947). The estimation from individual records of the relationship between dose and quantal response. Biometrika, 34, 320-334.

References

Pregibon, D. (1981) Logistic regression diagnostics. Annals of Statistics, 9,705-724.

Examples

```
data(vaso.df)
plot(vaso.df$Rate,vaso.df$Volume,type="n", cex=1.2)
text(vaso.df$Rate,vaso.df$Volume,1:39,
col=ifelse(vaso.df$Response==1, "red","blue"),cex=1.2)
text(2.3,3.5,"blue: no VS", col="blue",adj=0, cex=1.2)
text(2.3,3.0,"red: VS", col="red", adj=0, cex=1.2)
```

WB.test *Performs a Weisberg-Bingham test for normality*

Description

Calulates and prints a chi^2 statistic and a p-value for the Weisberg-Bingham test. The p-value is calculated by simulation.

Usage

```
WB.test(f, n.rep=1000, ...)
## S3 method for class 'lm'
WB.test(f, n.rep=1000, ...)
## S3 method for class 'formula'
WB.test(f, n.rep=1000, data, subset, weights, na.action,
 method = "qr", model = TRUE, x = FALSE, y = FALSE,
 qr = TRUE, singular.ok = TRUE, contrasts = NULL, offset, ...)
```

Arguments

f	an lm object or model formula
n.rep	the number of simulations desired to compute the p-value. The default of 1000 should be adequate
data	A data frame, list or environment containing the variables in the model.
subset	an optional vector specifying a subset of observations to be used in the fitting process.
weights	an optional vector of 'prior weights' to be used in the fitting process. Should be NULL or a numeric vector.
na.action	a function which indicates what should happen when the data contain NAs. The default is set by the na.action setting of options, and is na.fail if that is unset. The 'factory-fresh' default is na.omit. Another possible value is NULL, no action. Value na.exclude can be useful.
method	the method to be used in fitting the model. The default method "glm.fit" uses iteratively reweighted least squares (IWLS): the alternative "model.frame" returns the model frame and does no fitting.
x, y, qr, model	For glm: logical values indicating whether the response vector and model matrix used in the fitting process should be returned as components of the returned value. For glm.fit: x is a design matrix of dimension n * p, and y is a vector of observations of length n.
singular.ok	logical. If FALSE (the default in S but not in R) a singular fit is an error.
contrasts	an optional list. See the contrasts.arg of model.matrix.default.
offset	this can be used to specify an a priori known component to be included in the linear predictor during fitting. This should be NULL or a numeric vector of length equal to the number of cases. One or more offset terms can be included in the formula instead or as well, and if more than one is specified their sum is used. See model.offset.
...	additional arguments to be passed to the low level regression fitting functions see lm and glm help files

Value

Returns no value but prints the value of the Weisberg-Bingham test statistic and the p-value of the test

Note

This function redirects to other functions based on the type of object. eg WB.test.lm , WB.test.formula

Author(s)

Alan Lee, Blair Robertson

References

S Weisberg and C Bingham (1975). An approximate analysis of variance test for non-normality suitable for machine calculation, Technometrics 17, 133-134.

Examples

```
data(cherry.df)
cherry.lm =lm(volume~diameter+height,data=cherry.df)
WB.test(cherry.lm)
```

`wine.df` *Bordeaux Wine data*

Description

A data set which attempted to assess the quality of various Bordeaux vintages based upon certain variables

Usage

```
data(wine.df)
```

Format

A data frame with 27 observations on the following 5 variables:

`year` year (1952-1980)

`price` Price (in 1980 US dollars, converted to an index with 1961=100)

`temp` average temp during the growing season (degrees Celcius)

`h.rain` total rainfall during harvest period (mm)

`w.rain` total rainfall over preceding winter (mm)

Source

The data are available at http://www.liquidasset.com/winedata.html

References

An article by Orly Ashenfelder is at http://www.liquidasset.com/orley.htm See also Orley Ashenfelter, David Ashmore, and Robert Lalonde, Bordeaux wine vintage quality and the weather. Chance Magazine, Fall 1995, pp.7-14

Examples

```
data(wine.df)
boxcoxplot(price~temp+h.rain+w.rain+year, data=wine.df)
```

Index