Package: lmreg (via r-universe)

August 30, 2024

Type Package

Title Data and Functions Used in Linear Models and Regression with R: An Integrated Approach

Version 1.2

Date 2019-04-08

Author Debasis Sengupta <shairiksengupta@gmail.com>, S. Rao Jammalamadaka <rao@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

Maintainer Jinwen Qiu <qjwsnow_ctw@hotmail.com>

Description Data files and a few functions used in the book 'Linear Models and Regression with R: An Integrated Approach' by Debasis Sengupta and Sreenivas Rao Jammalamadaka (2019).

License GPL (>= 2)

NeedsCompilation no

Depends MASS

Date/Publication 2019-04-09 04:40:03 UTC

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

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

RemoteRef HEAD

RemoteSha 8011a935c7e5a62b26ebfa423fede6fa9a866144

Contents

`airspeed` *Air speed experiment data*

Description

Air speed data, which is part of a larger data set from a designed experiment (Wilkie, 1962).

Usage

```
data(airspeed)
```

Format

A data frame with 18 observations on the following 3 variables.

`Posmaxspeed` The position of highest speed of air blown down the space between a roughened rod and a smoothed pipe surrounding it. The position is defined as the distance (in inches) from the center of the rod, in excess of 1.4 inches

`Reynolds` Reynolds number of air flow (dimensionless)

`Ribht` Height of ribs on the roughened rod (in inches)

Source

Wilkie, D. (1962) A method of analysis of mixed level factorial experiments. Applied Statistics, pp.184-195.

Examples

```
data(airspeed)
head(airspeed)
```

`anscombeplus` *Six data sets with similar regression summary*

Description

Six synthetic data sets with similar regression summary, for illustrating the importance of regression diagnostics.

Usage

```
data(anscombeplus)
```

Format

A data frame with 20 observations on 8 synthetic real-valued variables, labelled as `x1`, `y1`, `y2`, `y3`, `y4`, `y5`, `x2`, `y6`.

- `x1` Explanatory variable of first five data sets
- `y1` Response variable of first data set
- `y2` Response variable of second data set
- `y3` Response variable of third data set
- `y4` Response variable of fourth data set
- `y5` Response variable of fifth data set
- `x2` Explanatory variable of sixth data set
- `y6` Response variable of sixth data set

Details

This data set is presented by Sengupta and Jammalamadaka (2019), after expanding on the ideas of Anscombe (1973)

Source

Anscombe, F.J. (1973), Graphs in statistical analysis, American Statistician, vol.27, pp.17-21.

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach, World Scientific Publishing Co., Table 5.1.

Examples

```
data(anscombeplus)
head(anscombeplus)
```

`appletree` *Apple yield with cropping under tree*

Description

Apple crop volume under various ground covers underneath tree (Pearce, 1983)

Usage

```
data(appletree)
```

Format

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

`Weight` Total weight (in pounds) of apple produced in a plot in four years, post-treatment

`Treatment` Five types of permanent cropping under the apple tree (coded as 1 to 5), or no cropping at all (0)

`Block` Blocks coded as 1 to 4

`Volume` Total crop volume (in bushels) in four years, pre-treatment

Source

Pearce, S.C. (1983) The Agricultural Field Experiment, Wiley, Chechester, p.284.

Examples

```
data(appletree)
head(appletree)
```

`basis` *Basis of column space of a matrix*

Description

Computes an orthonormal basis of the column space of a given matrix.

Usage

```
basis(M, tol=sqrt(.Machine$double.eps))
```

Arguments

`M`	Matrix for which basis of the column space is needed.
`tol`	A relative tolerance to determine rank through qr decomposition (default = sqrt(.Machine$double.eps)).

Value

Returns a semi-orthogonal matrix with columns forming an orthonormal basis of the column space of M.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
basis(matrix(c(2,1,3,4,2,3,2,6,4,2,6,8),4,3))
```

---

binaries *Convert categorical variable to several binary variables*

Description

Stacks up in columns the values of all the binary variables that can be associated with different levels of a categorical variable.

Usage

```
binaries(x)
```

Arguments

x A categorical variable (either numeric or character).

Details

The name of each new variable is of the type v.x, where x is the level of the categorical variable for which this binary variable is equal to 1.

Value

A set of binary vectors, each having the value 1 for a unique level of x.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
x <- c(1,2,2,3,1,1,2,3,3,2,1)
binaries(x)
binaries(as.factor(x))
```

cisimult — *Simultaneous confidence intervals in a linear model*

Description

Produces two-sided Bonferroni and Scheffe simultaneous confidence intervals, together with corresponding single confidence intervals, for any vector of estimable functions A.beta in a linear model.

Usage

```
cisimult(y, X, A, alpha, tol=sqrt(.Machine$double.eps))
```

Arguments

`y`	Responese vector in linear model.
`X`	Design/model matrix or matrix containing values of explanatory variables (generally including intercept).
`A`	Coefficient matrix (A.beta is the vector for which confidence interval is needed).
`alpha`	Collective non-coverage probability of confidence intervals.
`tol`	A relative tolerance to detect zero singular values while computing generalized inverse, in case X is rank deficient (default = sqrt(.Machine$double.eps)).

Details

Normal distribution of response (given explanatory variables and/or factors) is assumed.

Value

The three sets of confidence intervals listed as below:

`BFCB`	Two-sided Bonferroni simultaneous confidence intervals.
`SFCB`	Two-sided Scheffe simultaneous confidence intervals.
`SNCB`	The single confidence intervals.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(denim)
attach(denim)
X <- cbind(1, binaries(Denim), binaries(Laundry))
A <- rbind(c(0,1,-1,0,0,0,0), c(0,1,0,-1,0,0,0), c(0,0,1,-1,0,0,0))
cisimult(Abrasion, X, A, 0.05, tol = 1e-10)
detach(denim)
```

---

cisngl *Confidence interval for a linear parametric function in a linear model*

Description

Computes point estimate and confidence interval for a single linear parametric function in a linear model.

Usage

```
cisngl(y, X, p, alpha, type, tol=sqrt(.Machine$double.eps))
```

Arguments

y	Responese vector in linear model.
X	Design/model matrix or matrix containing values of explanatory variables (generally including intercept).
p	Coefficient vector of linear parametric function for which confidence interval is needed.
alpha	Non-coverage probability of confidence interval.
type	Type of confidence interval ("lower", "upper", "both").
tol	A relative tolerance to detect zero singular values while computing generalized inverse, in case X is rank deficient (default = sqrt(.Machine$double.eps)).

Details

Normal distribution of response (given explanatory variables and/or factors) is assumed.

Value

Returns a list of two objects:

estimate	Point estimate.
ci	Confidence interval.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
library(MASS)
data(birthwt)
attach(birthwt)
X <- cbind(1, smoke, binaries(race))
p <- c(0,1,0,0,0)
cisngl(bwt, X, p, 0.05, type = "upper", tol = 1e-10)
cisngl(bwt, X, p, 0.05, type = "both", tol = 1e-10)
detach(birthwt)
```

`cisv` *Table of condition indices and singular vectors*

Description

Computes the table of condition indices and model matrix singular vectors for a linear model.

Usage

```
cisv(lmobj)
```

Arguments

`lmobj`	An object produced by lm fitting.

Details

Columns containing different elements of a singular vector are labelled either as (Intercept) or by the variable name.

Value

Returns the table of condition indices and model matrix right singular vectors for the chosen model, with singular vectors appearing as rows next to the corresponding condition index. Columns containing different elements of a singular vector are labelled either as (Intercept) or by the variable name.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(imf2015)
lmimf <- lm(UNMP~CAB+DEBT+EXP+GDP+INFL+INV, data = imf2015)
cisv(lmimf)
```

---

compbasis — *Basis of orthogonal complement of column space of a matrix*

Description

Computes an orthonormal basis of the orthogonal complement of the column space of a given matrix.

Usage

```
compbasis(M, tol=sqrt(.Machine$double.eps))
```

Arguments

`M`	Matrix for which basis of the orthogonal complement of the column space is needed.
`tol`	A relative tolerance to determine rank through qr decomposition (default = sqrt(.Machine$double.eps)).

Value

Returns a semi-orthogonal matrix with columns forming an orthonormal basis of the orthogonal complement of the column space of M.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
compbasis(matrix(c(3,3,3,3),2,2))
```

`confelps` *Confidence ellipsiod for multiple parameters in a linear model.*

Description

Computes confidence ellipsiod for a vector of estimable functions in a linear model.

Usage

```
confelps(y, X, A, alpha, tol=sqrt(.Machine$double.eps))
```

Arguments

`y`	Responese vector in linear model.
`X`	Design/model matrix or matrix containing values of explanatory variables (generally including intercept).
`A`	Coefficient matrix (A.beta is the vector for which confidence interval is needed).
`alpha`	The non-coverage probability of confidence ellipsoid.
`tol`	A relative tolerance to detect zero singular values while computing generalized inverse, in case X is rank deficient (default = sqrt(.Machine$double.eps)).

Details

Normal distribution of response (given explanatory variables and/or factors) is assumed.

Value

Returns a list of three objects:

`CenterOfEllipse`
: Center of ellipsoid.

`MatrixOfEllipse`
: Matrix of ellipsoid, for describing quadratic form in terms of the vector of deviations from center of ellipsoid.

`threshold`
: Upper limit of quadratic form that completes specification of ellipsoid.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(denim)
attach(denim)
X <- cbind(1,binaries(Denim),binaries(Laundry))
A <- rbind(c(0,1,0,-1,0,0,0),c(0,0,1,-1,0,0,0))
confelps(Abrasion, X, A, 0.05,tol=1e-12)
detach(denim)
```

`denim` *Abrasion of denim jeans*

Description

Effects of Laundering Cycles and denim treatment on edge abrasion of denim jeans (Card et al., 2006). Data simulated to match means/SDs.

Usage

```
data(denim)
```

Format

A data frame with 90 observations on the following 3 variables.

`Laundry` Three levels of laundry cycles (1 = 0 cycle, 2 = 5 cycles, 3 = 25 cycles)

`Denim` Three types of denim treatments (1 = pre-washed, 2 = stone-washed, 3 = enzyme washed)

`Abrasion` abrasion score (lower score means higher damage)

Source

Card, A., Moore, M.A. and Ankeny, M. (2006) Garment washed jeans: Impact of launderings on physical properties. Int. J. Clothing Sc. Tech., 18, pp.43-52.

Examples

```
data(denim)
head(denim)
```

`drugprice` *Price of drugs under generic and brand names*

Description

Across-countries median of median price ratio (MPR) of some medicines available in the private market under the generic name and the brand name of the originator (Gelders et al., 2005).

Usage

```
data(drugprice)
```

Format

A data frame with 13 observations on the following 2 variables.

`Drug` Generic name of drug, a character vector

`Quantity` Unit for price computation, a character vector

`OriginatorMPR` Originator median price ratio, a numeric vector

`GenericMPR` Generic median price ratio, a numeric vector

Details

The data comes from a World Health Organization (WHO) commissioned study on variation of drug prices over a number of developing countries. For comparability, the price in a particular region is expressed as a ratio (called median price ratio or MPR) with respect to the organization's drug price indicator median values. The data reflect the across-country median of these ratios in respect of 13 medicines, most of which are in the WHO list of essential medicines.

Source

Gelders, S., Ewen, M., Noguchi, N. and Laing R. (2005). Price, Availability and Affordability: An International Comparison of Chronic Disease Medicines, Background report prepared for the WHO Planning Meeting on the Global Initiative for Treatment of Chronic Diseases, Cairo, December 2005.

Examples

```
data(drugprice)
head(drugprice)
```

frob — *Frobenius norm of a matrix*

Description

Computes the Frobenius norm of a given matrix.

Usage

```
frob(M)
```

Arguments

M	Matrix whose Frobenius norm is to be computed.

Value

A scalar value, describing the Frobenius norm (positive square root of sum of squared elements) of M.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
frob(matrix(2,3,2))
```

ganova — *ANOVA table for linear hypothesis in a linear model*

Description

Prepares Analysis of Variance table for testing a general linear hypothesis in a linear model

Usage

```
ganova(y, X, A, xi, tol=sqrt(.Machine$double.eps))
```

Arguments

- `y` — Responese vector in linear model.
- `X` — Design matrix or matrix containing values of explanatory variables (generally including intercept).
- `A` — Coefficient matrix (A.beta = xi is the null hypothesis to be tested).
- `xi` — A vector (A.beta = xi is the null hypothesis to be tested).
- `tol` — A relative tolerance to detect zero singular values while computing generalized inverse, in case the model matrix is rank deficient (default = sqrt(.Machine$double.eps)).

Value

Returns analysis of variance table for testing A.beta = xi in the linear model with response vector y and matrix of explanatory variables/factors X.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(denim)
attach(denim)
X <- cbind(1,binaries(Denim), binaries(Laundry))
A <- rbind(c(0,1,-1,0,0,0,0), c(0,1,0,-1,0,0,0))
xi <- c(0, 0)
ganova(Abrasion, X, A, xi)
detach(denim)
```

`girlgrowth` *Growth data for girls*

Description

Heights of some adolescent girls, aged 7 to 12, in the southern part of Kolkata, India around the year 2008.

Usage

```
data(girlgrowth)
```

Format

A data frame with 905 observations on the following 2 variables.

`Age` Age of girls (in years)

`Height` Height of girls (in cm)

Source

Dasgupta (2015), Physical Growth, Body Composition and Nutritional Status of Bengali School aged Children, Adolescents and Young adults of Calcutta, India: Effects of Socioeconomic Factors on Secular Trends, Report 158, Ney-van Hoogstraten Foundation, The Netherlands.

Examples

```
data(girlgrowth)
head(girlgrowth)
```

---

hanova — *ANOVA table for adequacy of a subset in a linear model)*

Description

Prepares the Analysis of Variance table for testing adequacy of a subset model within a linear model.

Usage

```
hanova(lm1, lm2)
```

Arguments

`lm1`	An lm object describing full model.
`lm2`	An lm object describing subset model.

Details

Normal distribution of response (given explanatory variables and/or factors) is assumed. The program simply reformats the output of the `anova` function.

Value

Returns analysis of variance table for testing adequacy of lm2 within lm1.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(birthwt)
lmbw <- lm(bwt ~ smoke+factor(race), data = birthwt)
lm1 <- lm(bwt ~ smoke, data = birthwt)
hanova(lm1,lmbw)
```

`hiv` *HIV data*

Description

Light absorbance for positive control samples in an ELISA test for HIV (Hoaglin et al., 1991).

Usage

```
data(hiv)
```

Format

A data frame with 75 observations on the following 3 variables.

`Absorbance` Measurement of absorbance of light (dimensionless)

`Lot` Five levels of lot

`Run` Five levels of run

Source

Hoaglin, D.C., Mosteller, F. and Tukey, J.W. (1991) Fundamentals of Exploratory Analysis of Variance, Wiley, New York, p.107.

Examples

```
data(hiv)
head(hiv)
```

hoop *Hoop tree data*

Description

Compressive strength and moisture content of wood in hoop trees (Williams, 1959).

Usage

```
data(hoop)
```

Format

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

`Temp` Temperature (in Celsius)

`Tree` Hoop tree number

`Strength` Maximum compressive strength parallel to the grain (in MPa)

`Moisture` Moisture content (100 times water mass/dry wood mass)

Source

Williams, E.J. (1959) Regression Analysis, Wiley, New York.

Examples

```
data(hoop)
head(hoop)
```

hypsplit *Testable and untestable hypotheses in linear model*

Description

Reduces a general hypothesis in a linear model into a pair of completely testable and completely untestable hypotheses.

Usage

```
hypsplit(X, A, xi, tol=sqrt(.Machine$double.eps))
```

Arguments

X	Design/model matrix or matrix containing values of explanatory variables (generally including intercept).
A	Coefficient matrix (A.beta = xi is the null hypothesis to be split).
xi	A vector (A.beta = xi is the null hypothesis to be tested).
tol	A relative tolerance to detect zero singular values while computing generalized inverse, in case X is rank deficient (default = sqrt(.Machine$double.eps)).

Value

A list of two objects:

testable	Coefficient matrix and constant vector for testable part of hypotheses.
untestable	Coefficient matrix and constant vector for untestable part of hypotheses.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(denim)
attach(denim)
X <- cbind(1, binaries(Denim), binaries(Laundry))
A <- rbind(c(0,1,0,0,0,0,0), c(0,0,1,0,0,0,0), c(0,0,0,1,0,0,0))
xi <- c(0,0,0)
hypotheses <- hypsplit(X, A, xi, tol=1e-13)
hypotheses[[1]]  # testable
hypotheses[[2]]  # untestable
detach(denim)
```

---

hyptest *Test of a linear hypothesis in a linear model*

Description

Carries out test of a single linear hypothesis in a linear model.

Usage

```
hyptest(lmobj, p, xi = 0, type = "both")
```

Arguments

`lmobj`	An object produced by lm fitting.
`p`	A numeric vector containing coefficients of the linear combination of model parameters.
`xi`	A numeric variable containing hypothesized value of the linear combination of model parameters (default = 0).
`type`	A character variable indicating the type of alternative: "upper" (one-sided), "lower" (one-sided) or "both" (default, two-sided).

Details

It is assumed that all the model parameters are estimable and the linear model is homoscedastic and normal.

Value

Returns the estimated value of the linear combination of model parameters, its standard error, the t-statistic, the degrees of freedom and the p-value.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(lifelength)
lmlife <- lm(Lifelength~factor(Category), data = lifelength)
p <- c(0,0,0,1,-1,0,0,0)
hyptest(lmlife, p, xi = 1, type = "upper")
```

---

imf2015 *IMF unemployment data*

Description

The estimated or reported figures of a number of economic variables for a few countries in the year 2015, extracted from IMF World Economic Outlook (2017)

Usage

```
data(imf2015)
```

Format

A data frame with 33 observations on the following 8 variables.

`Country` Country name, a character vector

`CAB` Current account balance as % of GDP, a numeric vector

`DEBT` Governmentt gross debt as % of GDP, a numeric vector

`EXP` Government total expenditure as % of GDP, a numeric vector

`GDP` GDP per capita, current prices in '000 US$, a numeric vector

`INFL` Inflation, average consumer prices in %, a numeric vector

`INV` Total investment as % of GDP, a numeric vector

`UNMP` Unemployment as % of labor force, a numeric vector

Source

http://www.imf.org/external/pubs/ft/weo/2017/01/weodata/weoselgr.aspx.

Examples

```
data(imf2015)
head(imf2015)
```

`intsectbasis` *Basis of intersection of two column spaces*

Description

Computes an orthonormal basis of the intersection of column spaces of two given matrices.

Usage

```
intsectbasis(A, B, tol1=sqrt(.Machine$double.eps), tol2=sqrt(.Machine$double.eps))
```

Arguments

`A`	First matrix.
`B`	Second matrix with identical number of rows.
`tol1`	A relative tolerance to detect zero singular values while computing generalized inverse, in case the matrix concerned is rank deficient (default = sqrt(.Machine$double.eps)).
`tol2`	A tolerance to detect if there is any non-zero singular value of a 'parallel sum' matrix, without which the intersection space is null (default = sqrt(.Machine$double.eps)).

Value

Returns a semi-orthogonal matrix with columns forming an orthonormal basis of the intersection of the column spaces of A and B.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
A<-matrix(2,3,5)
B<-matrix(3,3,2)
intsectbasis(A,B, tol1=sqrt(.Machine$double.eps), tol2=1e-14)
```

Iris — *Fisher's Iris data*

Description

Measurements of four dimensions of flowers of three species of the plant Iris (Iris setosa, Iris versicolor, and Iris virginica).

Usage

```
data(Iris)
```

Format

A data frame with 150 observations on the following 6 variables.

`Species_No` Species number

`Petal_width` Petal width (in cm)

`Petal_length` Petal length (in cm)

`Sepal_width` Sepal width (in cm)

`Sepal_length` Sepal length (in cm)

`Species_name` Species names: `Setosa`, `Verginica` or `Versicolor`, a character vector

Source

Fisher, R.A. (1936) The use of multiple measurements in taxonomic problems. Ann. Eugenics, 7, pp.179-188.

Examples

```
data(Iris)
head(Iris)
```

is.included — *Whether one column space is contained in another*

Description

Checks whether column space of one matrix is a subset of the column space of another matrix.

Usage

```
is.included(B, A, tol1=sqrt(.Machine$double.eps), tol2=sqrt(.Machine$double.eps))
```

Arguments

`B`	The matrix whose column space is to be checked for being a subset.
`A`	The matrix whose column space is to be checked for being a superset.
`tol1`	A relative tolerance to detect zero singular values while computing generalized inverse, in case A is rank deficient (default = sqrt(.Machine$double.eps)).
`tol2`	A relative tolerance to detect whether there is sufficient closeness between B and A.ginv(A).B (default = sqrt(.Machine$double.eps)).

Value

A logical value (TRUE if the column space of B is contained in the column space of A).

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
A <- cbind(c(2,1,-2),c(3,1,-1))
I <- diag(1,3)
is.included(A, I, tol1=sqrt(.Machine$double.eps), tol2=1e-15)
is.included(I, A, tol1=1e-14, tol2=sqrt(.Machine$double.eps))
is.included(projector(A), A, tol1=1e-15, tol2=1e-14)
is.included(A, projector(A))
```

ivif — *Intercept augmented variance inflation factors*

Description

Computes the intercept augmented variance inflation factors for a linear model.

Usage

```
ivif(lmobj)
```

Arguments

`lmobj`	An object produced by lm fitting.

Value

Returns the intercept augmented variance inflation factors for the model, with each VIF labelled either as (Intercept) or by the variable name.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(imf2015)
lmimf <- lm(UNMP~CAB+DEBT+EXP+GDP+INFL+INV, data = imf2015)
ivif(lmimf)
```

kinks — *Kink bands in rocks*

Description

Measurements of an angular dimension (beta angle) found in kink bands of Daling phyllite in the Darjeeling-Sikkim Himalayas.

Usage

```
data(kinks)
```

Format

A data frame with 100 observations on the following 3 variables.

`beta` Beta angle in kink bands (in degrees)

`order` Fold order (1 = main fold, 2 = sub-fold, 3,4 = sub-folds of successively higher order)

`type` Type of kink band (1 = conjugate, 2 = dextral, 3 = sinistral)

Source

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach, World Scientific Publishing Co., Table 6.8.

Examples

```
data(kinks)
head(kinks)
```

`LAcrime` *LA crime and temperature data*

Description

Monthly total counts of homicides and rapes in the city of Los Angeles from January 1975 to December 1993.

Usage

```
data(LAcrime)
```

Format

A data frame with 228 observations on the following 7 variables.

`Year` Year of record

`Month` Month of record

`Population` Population of the city in the year of record

`TempCelsius` Monthly average temperature recorded at the Los Angeles International Airport (in Celsius)

`Fahrenheit` Monthly average temperature recorded at the Los Angeles International Airport (in Fahrenheit)

`Homicide` Total count of homicides in the month and year of record

`Rape` Total count of rapes in the month and year of record

Source

The crime data: Carlson, S.M. (1998), Uniform Crime Reports: Monthly Weapon-Specific Crime and Arrest Time Series, 1975-1993, ICPSR06792-v1, Interuniversity Consortium for Political and Social Research, Ann Arbor, MI (https://www.icpsr.umich.edu/icpsrweb/NACJD/studies/6792). Temperature data for LAX (WMO ID 72295): National Oceanic and Atmospheric Administration, USA (http://www.ncdc.noaa.gov/ghcnm/v2.php)

Examples

```
data(LAcrime)
head(LAcrime)
```

`leprosy` *Treatment of leprosy*

Description

Pre- and post-treatment scores on abundance of leprosy for patients receiving different treatments (Senedecor and Cochran, 1967).

Usage

```
data(leprosy)
```

Format

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

`treatment` Treatment type: `A`, `D` or `F` (placebo), a character vector

`pre` Pre-treatment score, a numerical vector

`post` Post-treatment score, a numerical vector

Source

Snedecor, G.W. and Cochran, W.G. (1967) Statistical Methods, Iowa State University, Ames, p.421.

Examples

```
data(leprosy)
head(leprosy)
```

lifelength — *Age at death*

Description

William Guy's nineteenth century data on the age at death of persons belonging to different professions.

Usage

```
data(lifelength)
```

Format

A data frame with 690 observations on the following 2 variables.

`Category` Code for profession: 1 = historian, 2 = poet, 3 = painter, 4 = musician, 5 = mathematician or astronomer, 6 = chemist or natural philosopher, 7 = naturalist, 8 = engineer, architect or surveyor

`Lifelength` Age (in years) of deceased

Source

Guy, W. (1859) On the duration of life as affected by the pursuits of literature, science and art. J. Statist. Soc. London, 22.

Examples

```
data(lifelength)
head(lifelength)
```

multcomp — *Multiple comparison tests*

Description

Produces p-values of Bonferroni and Scheffe multiple comparison tests of several testable linear hypotheses.

Usage

```
multcomp(y, X, A, xi, tol=sqrt(.Machine$double.eps))
```

Arguments

- `y` — Responese vector in linear model.
- `X` — Design/model matrix or matrix containing values of explanatory variables (generally including intercept).
- `A` — Coefficient matrix (A.beta=xi is the set of multiple hypotheses that has to be tested).
- `xi` — A vector of values (A.beta=xi is the set of multiple hypotheses that has to be tested).
- `tol` — A relative tolerance to detect zero singular values while computing generalized inverse, in case X is rank deficient (default = sqrt(.Machine$double.eps)).

Details

Normal distribution of response (given explanatory variables and/or factors) is assumed.

Value

Returns F statistics and p-values of Bonferroni and Scheffe multiple comparison tests of the set of linear hypotheses. A set of five vectors:

- `A` — Specified coefficient matrix.
- `xi` — Specified values of A.beta.
- `Fstat` — Set of F-ratios for each hypothesis.
- `Bonferroni.p` — Set of Bonferroni p-values for different hypotheses.
- `Scheffe.p` — Set of Scheffe p-values for different hypotheses.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(denim)
attach(denim)
X <- cbind(1,binaries(Denim),binaries(Laundry))
A <- rbind(c(0,1,-1,0,0,0,0),c(0,1,0,-1,0,0,0),c(0,0,1,-1,0,0,0))
xi <- c(0,0,0)
multcomp(Abrasion, X, A, xi, tol=1e-14)
detach(denim)
```

`olympic` *Olympic sprint finals data*

Description

Times recorded by winners of men's olympic sprint finals in different categories from 1900 to 1988 (Lunn and McNeil, 1991).

Usage

```
data(olympic)
```

Format

A data frame with 20 observations on the following 6 variables.

`Year` Olympic year

`X100m` Winner's time (in seconds) for 100 meters sprint

`X200m` Winner's time (in seconds) for 200 meters sprint

`X400m` Winner's time (in seconds) for 400 meters sprint

`X800m` Winner's time (in seconds) for 800 meters sprint

`X1500m` Winner's time (in seconds) for 1500 meters sprint

Details

There are three missing years in the data; 1916, 1940 and 1944, when world wars prevented the olympic games from being held.

Source

Lunn, A.D. and McNeil, D.R. (1991) Computer-Interactive Data Analysis, Wiley, Chichester.

Examples

```
data(olympic)
head(olympic)
```

poison *Survival times of poisoned animals*

Description

Survival times of animals exposed to poison and treatment (Box and Cox, 1964).

Usage

```
data(poison)
```

Format

A data frame with 48 observations on the following 3 variables.

`Survtime` Survival time (in 10 hour units)

`Treatment` Treatment type: 1 = treatment A, 2 = treatment B, 3 = treatment C, 4 = treatment D

`Poison` Poison type: 1 = Poison I, 2 = Poison II, 3 = Poison III

Source

Box, G.E.P. and Cox, D.R. (1964) An analysis of transformations. J. Roy. Statist. Soc. Ser. B, 26, pp.211-252.

Examples

```
data(poison)
head(poison)
```

projector *Orthogonal projector of a matrix*

Description

Computes the orthogonal projection matrix for the column space of a given matrix.

Usage

```
projector(M, tol=sqrt(.Machine$double.eps))
```

Arguments

`M`	A matrix for which the orthogonal projection matrix is to be computed.
`tol`	A relative tolerance to detect zero singular values while computing generalized inverse, in case M is rank deficient (default = sqrt(.Machine$double.eps)).

Value

Returns the orthogonal projection matrix for the column space of M.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
projector(matrix(3,3,3))
```

`skulls` *Egyptian skull development*

Description

Measurements of male Egyptian skulls from time periods ranging from 4000 BC to 150 AD.

Usage

```
data(skulls)
```

Format

A data frame with 150 observations on the following 5 variables.

- `MB` Maximal breadth (in mm)
- `BH` Basibregmatic height (in mm)
- `BL` Basialveolar length (in mm)
- `NH` Nasal height (in mm)
- `Year` Approximate Year of Skull Formation (negative = B.C., positive = A.D.)

Source

Thomson, A. and Randall-Maciver, R. (1905) Ancient Races of the Thebaid, Oxford University Press, Oxford.

Examples

```
data(skulls)
head(skulls)
```

splett2 *Energy data*

Description

Energy absorbed by four machines for Charpy V-notch testing.

Usage

```
data(splett2)
```

Format

A data frame with 99 observations on the following 2 variables.

`Energy` Energy absorbed by machine (in foot-pounds)

`Machine` Machine type (1 = Tinius1, 2 = Tinius2, 3 = Satec, 4 = Tokyo)

Source

Dataplot webpage of the National Institute of Standards and Technology (NIST), USA (https://www.itl.nist.gov/div898/software/dataplot/data/SPLETT2.DAT).

Examples

```
data(splett2)
head(splett2)
```

stars1 *Stars data 1*

Description

Distance of galactic objects from Earth and their velocities (Hubble, 1929).

Usage

```
data(stars1)
```

Format

A data frame with 24 observations on the following 2 variables.

`Distance` Distance from Earth (in million parsec; 1 parsec = 3.26 light years)

`Velocity` Velocity of galaxy (in km/s)

Source

Hubble, E. (1929) A relation between distance and radial velocity among extra galactic nebulae. Proc. Nat. Acad. Sc. 15, pp.168-73.

Examples

```
data(stars1)
head(stars1)
```

`stars2` *Stars data 2*

Description

Distance of additional galactic objects from Earth and their velocities (Humason, 1936).

Usage

```
data(stars2)
```

Format

A data frame with 21 observations on the following 2 variables.

`Distance` Distance from Earth (in million parsec; 1 parsec = 3.26 light years)

`Velocity` Velocity of Galaxy (in km/s)

Details

The galactic objects in this data set are much further away from Earth than those in the data set `stars1.txt`. These became available within a few years of the publication of Hubble's original work, through rapid advancesment in technology. Although the new data cemented Hubble's hypothesis that distant objects have proportionately higher velocity (as they should in a universe expanding with constant acceleration), the constant of proportionality turned out to be somewhat different from Hubble's original estimate.

Source

Humason, M.L. (1936) The apparent radial velocities of 100 extra galactic nebula. Astrophys. J. 83, pp.10-22.

Examples

```
data(stars2)
head(stars2)
```

supplbasis	*Supplementary basis vectors for column space of a matrix*

Description

Computes a basis which, together with a basis of some columns of a matrix, constitute a basis of the column space of the entire matrix.

Usage

```
supplbasis(A, B, tol=sqrt(.Machine$double.eps))
```

Arguments

`A`	Sub-matrix containing some columns of a matrix.
`B`	Sub-matrix containing remaining columns of same matrix.
`tol`	A relative tolerance to detect rank deficiency during qr decomposition (default = sqrt(.Machine$double.eps)).

Value

Returns a semi-orthogonal matrix whose columns, together with a basis of the column space of A, constitute a basis of the column space of the entire matrix (A:B).

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
A <- cbind(c(2,1,-2),c(3,1,-1))
B <- diag(c(1,1,0))
supplbasis(A,B)
```

tr — *Trace of matrix*

Description

Computes the trace of a given matrix.

Usage

```
tr(M)
```

Arguments

`M`	A matrix whose trace is to be computed.

Value

A scalar value, describing the trace of M.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
tr(matrix(2,2,2))
```

trout — *Brown trout hemoglobin data*

Description

The measured hemoglobin content in the blood of brown trout that were randomly allocated to four troughs, where different concentrations of sulfamerazine in food were administered 35 days prior to measurement (Gutsell, 1951).

Usage

```
data(trout)
```

Format

A data frame with 40 observations on the following 2 variables.

`Sulfamerazine` Concentrations of sulfamerazine (in grams per 100 pounds of fish)

`Hemoglobin` Hemoglobin content (in grams per 100 ml of blood)

Source

Gutsell, James S. (1951) The effect of sulfamerazine on the erythrocyte and hemoglobin content of trout blood, Biometrics 7(2), pp.171-179.

Examples

```
data(trout)
head(trout)
```

`waist` *Waist circumference and adipose tissue data*

Description

Waist circumference and adipose tissue data (Daniel and Cross, 2013).

Usage

```
data(waist)
```

Format

A data frame with 109 observations on the following 2 variables.

`Waist` Waist circumference (in centimeters)

`AT` Area of lower abdominal adipose tissue (in squared centimeters)

Source

Daniel, W.W. and Cross, C.L. (2013) Biostatistics: A Foundation for Analysis in the Health Sciences, tenth edition, Wiley, New York, Table 9.3.1.

Examples

```
data(waist)
head(waist)
```

`worldpop` *World population data*

Description

The midyear population of the world for the years 1981-2000.

Usage

```
data(worldpop)
```

Format

A data frame with 20 observations on the following 2 variables.

`Year` Calendar year

`Pop.billion` Population (in billion)

Source

U.S. Census Bureau, International Data Base (http://www.census.gov/ipc/www/idbnew.html)

Examples

```
data(worldpop)
head(worldpop)
```

`worldrecord` *World record running times data*

Description

Men's and women's world record times for various out-door running distances, recognized by the International Association of Athletics Federations (IAAF) as of 17 November, 2017.

Usage

```
data(worldrecord)
```

Format

A data frame with 10 observations on the following 3 variables.

`Distance` Running distance (in meters)

`MenRecord` Men's record time (in seconds)

`WomenRecord` Women's record time (in seconds)

Source

International Association of Athletics Federations (https://www.iaaf.org/records/by-category/world-records).

Examples

```
data(worldrecord)
head(worldrecord)
```

`Wright` *Wright brothers' wind tunnel data*

Description

Wright brothers' 1901 wind tunnel data on pressure over different types of wings at different angles.

Usage

```
data(Wright)
```

Format

A data frame with 222 observations on the following 3 variables.

`Pressure` Air pressure (in psi)

`Angle` Angle of wing (in degrees)

`Wing` Wing type

Source

Dataplot webpage of the National Institute of Standards and Technology (NIST), USA (https://www.itl.nist.gov/div898/software/dataplot/data/WRIGHT11.DAT)

Examples

```
data(Wright)
head(Wright)
```

yX — *Prepare design matrix for two way layout with single oberservation per cell*

Description

Prepares design matrix for two way classified data with single observation per cell and response vector in corresponding order.

Usage

```
yX(response, treatments, blocks)
```

Arguments

`response`	Response vector as provided (numeric).
`treatments`	Vector of treatment levels as provided (either numeric or character).
`blocks`	Vector of block levels as provided (either numeric or character).

Value

Returns a list with following components.

`X`	A binary matrix with number of rows equal to length of response and number of columns equal to the total number of levels of treatments and blocks plus one. Each row has exactly three 1s: in the first position and in the two positions representing the treatment and block levels.
`y`	Numeric vector of response values, permuted to correspond with the rows of X.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(airspeed)
yX(airspeed$Posmaxspeed,airspeed$Reynolds,airspeed$Ribht)
```

yXm *Prepare design matrix for balanced two way layout*

Description

Prepares design matrix for balanced two way classified data and response vector in corresponding order.

Usage

```
yXm(response, treatments, blocks)
```

Arguments

`response`	Response vector as provided (numeric).
`treatments`	Vector of treatment levels as provided (either numeric or character).
`blocks`	Vector of block levels as provided (either numeric or character).

Value

Returns a list with following components.

`X`	A binary matrix with number of rows equal to length of response and number of columns equal to the total number of levels of treatments and blocks plus one. Each row has exactly three 1s: in the first position and in the two positions representing the treatment and block levels.
`y`	Numeric vector of response values, permuted to correspond with the rows of X.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(poison)
yXm(poison$Survtime,poison$Treatment,poison$Poison)
```

yXn *Prepare design matrix for nested model with groups and subgroups*

Description

Prepares design matrix for nested model with groups and subgroups and response vector in corresponding order.

Usage

```
yXn(response, group, subgroup)
```

Arguments

`response`	Response vector as provided (numeric).
`group`	Vector of group labels as provided (either numeric or character).
`subgroup`	Vector of subgroup labels as provided (either numeric or character).

Value

Returns a list with following components.

`X`	A binary matrix with number of rows equal to length of response and number of columns equal to the total number of levels of treatments and blocks plus one. Each row has exactly three 1s: in the first position and in the two positions representing the group and the subgroup.
`y`	Numeric vector of response values, permuted to correspond with the rows of X.

Author(s)

Debasis Sengupta <shairiksengupta@gmail.com>, Jinwen Qiu <qjwsnow_ctw@hotmail.com>

References

Sengupta and Jammalamadaka (2019), Linear Models and Regression with R: An Integrated Approach.

Examples

```
data(kinks)
yXn(kinks$beta,kinks$type,kinks$order)
```

Index