Package ‘alr3’

January 2, 2012

Version 2.0.5

Date 2011/10/27

Title Data to accompany Applied Linear Regression 3rd edition

Author Sanford Weisberg <sandy@umn.edu>

Maintainer Sanford Weisberg <sandy@umn.edu>

Depends R (>= 2.1.1), car

LazyData yes

Description This package is a companion to the textbook S. Weisberg (2005),‘‘Applied Linear Regression,’’ 3rd edition, Wiley. It includes all the data sets discussed in the book (except one), and a few functions that are tailored to the methods discussed in the book. As of version 2.0.0,this package depends on the car package. Many functions formerly in alr3 have been renamed and now reside in car. Data files have beeen lightly modified to make some data columns row labels.

License GPL (>= 2)

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

Repository CRAN

Date/Publication 2011-10-27 19:35:19

R topics documented:

alr3-package 3
ais 4
allshoots 5
alr3-deprecated 5
alrWeb 7
baeskel 8
banknote 8

BGSall 9
BigMac2003 10
blowdown 11
brains 11
cakes 12
cathedral 13
caution 13
challeng 14
chloride 15
cloud 15
domedata 16
donner 17
downer 17
drugcost 18
dwaste 19
florida 20
forbes 21
ftcollinssnow 21
fuel2001 22
galapagos 23
galtonpeas 24
heights 24
highway 25
hooker 26
htwt 27
jevons 27
lakemary 28
lakes 29
landrent 30
lathe1 30
mantel 31
mile 32
Mitchell 32
MWwords 33
npdata 34
oldfaith 34
physics 35
pipeline 36
pod 36
prodscore 40
pureErrorAnova 41
randomLinComb 42
rat 43
salary 44
salarygov 44
segreg 45
shocks 46
sleep1 46

snake 47
sniffer 48
snowgeese 49
stopping 49
swan96 50
titanic 51
transact 51
turk0 52
turkey 53
twins 53
ufc 54
UN1 55
UN2 56
UN3 56
walleye 57
water 58
wblake 59
wm1 60
wm2 60
wm3 61
wm4 62
wool 63

Index **64**

`alr3-package` *Companion to Applied Linear Regression*

Description

This package accompanies S. Weisberg, Applied Regression Regression, Third Edition, Wiley, 2005. Most of the functions that were previously part of alr3 have been renamed and improved, and are now part of the car package.

Details

Package:	alr3
Version:	2.0.0
Date:	2010/03/15
Depends:	R (>= 2.1.1), car
License:	GPL (>= 2)
URL:	http://www.r-project.org, http://www.stat.umn.edu/alr

Author(s)

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

`ais` *Australian institute of sport data*

Description

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

Format

This data frame contains the following columns:

Sex (0 = male or 1 = female)

Ht height (cm)

Wt weight (kg)

LBM lean body mass

RCC red cell count

WCC white cell count

Hc Hematocrit

Hg Hemoglobin

Ferr plasma ferritin concentration

BMI body mass index, weight/(height)**2

SSF sum of skin folds

Bfat Percent body fat

Label Case Labels

Sport Sport

Source

Ross Cunningham and Richard Telford

References

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

Examples

```
head(ais)
```

allshoots — *Apple shoots data*

Description

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

Format

This data frame contains the following columns:

Day days from dormancy

n number of shoots sampled

ybar average number of stem units

SD within-day standard deviation

Type 1 if long shoots, 0 if shortshoots.

Source

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

References

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

Examples

```
head(longshoots)
```

alr3-deprecated — *Deprecated Functions in the alr3 Package*

Description

These functions are provided for compatibility with older versions of alr3 only, and may be removed eventually.

Usage

```
sigma.hat(...)

conf.intervals(...)

boot.case(...)

pure.error.anova(...)

delta.method(...)

powtran(...)

inv.tran.plot(...)

inv.tran.estimate(...)

inverse.response.plot(...)

inv.res.plot(...)

bctrans(...)

bctrans1(...)

lrt.bctrans(...)

resid.curv.test(...)

tukey.nonadd.test(...)

resplot(...)

residual.plots(...)

marginal.model.plot(...)

marginal.model.plots(...)

inf.index(...)

outlier.t.test(...)
```

Arguments

`...` pass arguments down.

Details

random.lin.comb pure.error.anova are now synonyms for sigmaHat, randomLinComb and pureErrorAnova, all of which are still part of the alr3 package.

conf.intervals is a synonym for confint in the stats package.

All the remaining functions now shadow newer functions in the car package. In several cases, the arguments to the function have changed as well, and so there is no guarantee that any of these will work as expected. Please substitute the corresponding function from car.

---

alrWeb *Access to the Applied Linear Regression website*

Description

This function will access the website for Applied Linear Regression

Usage

```
alrWeb(page = c("webpage", "errata", "primer"), script)
```

Arguments

page	A character string indicating what page to open. The default "webpage" will open the main webpage, "errata" displays the Errata sheet for the book, and "primer" fetches and displays the primer for R.
script	If set, this will open an R script for one of the chapters in the book in a browser window, or a script from one of the chapters in the R Primer. Possible values are "chapter1", "chapter2", ... "chapter12", "appendix" for chapters and appendix from the book, or "primer0", "primer1", ... "primer12", "primerappendix" for material corresponding to the R primer.

Value

Either a webpage or a pdf document is displayed. This function gives quick access to the website for the book and in particular to the R primer.

Author(s)

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

Examples

```
## Not run: alrWeb()
```

baeskel *Surface tension*

Description

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

Format

This data frame contains the following columns:

Sulfur Weight percent sulfur

Tension Decrease in surface tension, dynes/cm

Source

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

References

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

Examples

```
head(baeskel)
```

banknote *Swiss banknote data*

Description

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

Format

This data frame contains the following columns:

Length Length of bill, mm

Left Width of left edge, mm

Right Width of right edge, mm

Bottom Bottom margin width, mm

Top Top margin width, mm

Diagonal Length of image diagonal, mm

Y 0 = genuine, 1 = counterfeit

Source

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

References

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

Examples

```
head(banknote)
```

BGSall *Berkeley guidance study*

Description

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

Format

This data frame contains the following columns:

Sex 0 = males, 1 = females

WT2 Age 2 weight (kg)

HT2 Age 2 height (cm)

WT9 Age 9 weight (kg)

HT9 Age 9 height (cm)

LG9 Age 9 leg circumference (cm)

ST9 Age 9 strength (kg)

WT18 Age 18 weight (kg)

HT18 Age 18 height (cm)

LG18 Age 18 leg circumference (cm)

ST18 Age 18 strength (kg)

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

Source

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

References

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

Examples

```
head(BGSall)
head(BGSboys)
head(BGSgirls)
```

BigMac2003 *World cities data*

Description

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

Format

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

BigMac Minutes of labor to purchase a Big Mac

Bread Minutes of labor to purchase 1 kg of bread

Rice Minutes of labor to purchase 1 kg of rice

FoodIndex Food price index (Zurich=100)

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

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

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

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

TaxRate Tax rate paid by a primary teacher

TeachHours Primary teacher's hours of work per week:

Source

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

References

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

Examples

```
head(BigMac2003)
```

---

blowdown *Blowdown data*

Description

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

Format

This data frame contains the following columns:

D Tree diameter, in cm

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

y 1 of the tree died, 0 if it survived

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

Source

Roy Rich

References

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

Examples

```
head(blowBF)
```

---

brains *Mammal brain weights*

Description

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

Format

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

BrainWt Brain weight, grams

BodyWt Body weight, kg

Source

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

References

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

Examples

```
head(brains)
```

---

cakes — *Cakes data*

Description

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

Format

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

block a numeric vector

X1 Baking time, minutes

X2 Baking temperature, degrees F

Y Palatability score

Source

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

References

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

Examples

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

`cathedral` *Cathedrals*

Description

Heights and lengths of Gothic and Romanesque cathedrals.

Format

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

Type Romanesque or Gothic

Height Total height, feet

Length Total length, feet

Source

Stephen Jay Gould

References

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

Examples

```
head(cathedral)
```

`caution` *Caution data*

Description

Artificial data to illustrate problems with residual plots.

Format

This data frame contains the following columns:

x1 Artificial data item.

x2 Artificial data item.

y Artificial data item.

Source

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

References

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

Examples

```
head(caution)
```

`challeng` *Challenger data*

Description

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

Format

This data frame contains the following columns:

Temp Air Temp at launch (degrees F)

Pres Leak check pressure

Fail Number of O-rings that failed

n 6, number of O-rings in launch

Erosion Number of erosion incidents

BlowBy Number of blowby incidents

Damage Total Damage Index

Date date of flight

Source

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

References

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

Examples

```
head(challeng)
```

`chloride` *Chloride data*

Description

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

Format

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

Marsh March number

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

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

Cl Chloride concentration, mg/liter

Source

Stefanie Miklovic and Susan Galatowitsch

References

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

Examples

```
head(chloride)
```

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

Description

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

Format

This data frame contains the following columns:

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

D Day after June 16, 1975

S Suitability for seeding

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

P $10^7 m^3$ prewetness

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

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

Source

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

References

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

Examples

```
head(cloud)
```

`domedata` *Metrodome fan experiment*

Description

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

Format

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

Date a factor with levels `March-May`

Cond a factor with levels `Headwind`, `Tailwind`

Angle the actual angle

Velocity in feet per second

BallWt weight of ball in grams

BallDia diameter of ball in inches

Dist distance in feet of the flight of the ball

Source

Ivan Marusic

References

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

Examples

```
head(domedata1)
```

`donner` *Donner party*

Description

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

Format

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

Age Approximate age in 1846

Outcome 1 if survived, 0 if died

Sex Male or Female

Family.name Either a family name, hired or single

Status Family, single or hired

Source

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

References

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

Examples

```
head(donner)
```

`downer` *Downer data*

Description

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

Format

This data frame contains the following columns:

Calving 0 if measured before calving, 1 if after

Daysrec Days recumbent

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

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

Urea serum urea (mmol/l)

PCV Packed Cell Volume (Haemactocrit),

Inflamat inflamation 0=no, 1=yes

Myopathy Muscle disorder, 1 if present, 0 if absent

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

Source

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

References

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

Examples

```
head(downer)
```

`drugcost` *Drug cost.*

Description

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

Format

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

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

RXPM Number of prescriptions per member per year

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

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

COPAY Average Rx copayment

AGE Average age of member

F Percent female members

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

Source

Mark Siracuse

References

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

Examples

```
head(drugcost)
```

---

dwaste *Crock data.*

Description

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

Format

This data frame contains the following columns:

Day Day number

BOD Biological oxygen demand

TKN Total Kjeldahl nitrogen

TS Total Solids

TVS Total volatile solids

COD Chemical oxygen demand

O2UP Oxygen uptake

Source

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

References

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

Examples

```
head(dwaste)
```

`florida` *Florida presidential election*

Description

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

Format

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

Gore Vote for Gore

Bush Vote for Bush

Buchanan Vote for Buchanan

Source

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

References

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

Examples

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

forbes — *Forbes data*

Description

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

Format

This data frame contains the following columns. The first two columns of this data frame are also given in the data file forbes in the MASS package, with different column labels.

Temp Adjusted boiling point of water in degrees F.

Pressure Atmospheric pressure, in inches of Mercury

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

Source

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

References

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

Examples

```
head(forbes)
```

ftcollinssnow — *Ft. Collins snowfall*

Description

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

Format

This data frame contains the following columns:

YR1 Year corresponding to the September to December data

Early September to December snowfall, inches

Late January to June snowfall, inches

Source

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

References

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

Examples

```
head(ftcollinssnow)
```

---

fuel2001 *Fuel consumption*

Description

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

Format

This data frame contains the following columns. Row labels are the two-digit US Postal abbreviations for the US states.

Drivers Number of Licensed drivers in the state

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

Income Per capita personal income (year 2000)

Miles Miles of Federal-aid highway miles in the state

MPC Estimated miles driven per capita

Pop Population age 16 and over

Tax Gasoline state tax rate, cents per gallon

Source

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

References

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

Examples

```
head(fuel2001)
# Most of the examples in ALR3 that use these data first
# transform several of the columns
fuel2001 <- transform(fuel2001,
     Dlic=1000 * Drivers/Pop,
     Fuel=1000 * FuelC/Pop,
     Income=Income/1000)
pairs(Fuel~Tax + Dlic + Income + log2(Miles), data=fuel2001)
```

---

galapagos *Galapagos species data*

Description

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

Format

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

NS Number of Species

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

Area Surface area of island, hectares

Anear Area of closest island, hectares

Dist Distance to closest island, km

DistSC Distance from Santa Cruz Island, km

Elevation Elevation in m, missing values given as zero

EM 1 if elevation is observed, 0 if missing

Source

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

References

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

Examples

```
head(galapagos)
```

`galtonpeas` *Galton's peas*

Description

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

Format

This data frame contains the following columns:

Parent mean diameter of parent

Progeny mean diameter of offspring

SD offspring standard deviation

Source

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

References

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

Examples

```
head(galtonpeas)
```

`heights` *Pearson-Lee data*

Description

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

Format

This data frame contains the following columns:

Mheight Mother's ht, in.

Dheight Daughter's ht, in.

Source

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

References

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

Examples

```
head(heights)
```

`highway` *Highway accidents*

Description

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

Format

This data frame contains the following columns:

ADT average daily traffic count in thousands

Trks truck volume as a percent of the total volume

Lane total number of lanes of traffic

Acpt number of access points per mile

Sigs number of signalized interchanges per mile

Itg number of freeway-type interchanges per mile

Slim speed limit in 1973

Len length of the highway segment in miles

Lwid lane width, in feet

Shld width in feet of outer shoulder on the roadway

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

Rate 1973 accident rate per million vehicle miles

Source

Carl Hoffstedt

References

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

Examples

```
head(highway)
```

`hooker` *Hooker's data*

Description

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

Format

This data frame contains the following columns:

Temp Measured boiling temperature, degrees F.

Pressure Measured air pressure, inches of Mercury.

Source

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

References

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

Examples

```
head(hooker)
```

htwt — *Artificial height and weight data*

Description

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

Format

This data frame contains the following columns:

Ht Height (cm) at age 18

Wt Weight (kg) at age 18

Source

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

References

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

Examples

```
head(htwt)
```

jevons — *Jevon's gold coin data*

Description

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

Format

This data frame contains the following columns:

Age Age of coins, decades

n Number of coins

Weight Average weight, grams

SD Standard deviation.

Min Minimum weight

Max Maximum weight

Source

Stephen Stigler

References

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

Examples

```
head(jevons)
```

---

lakemary *Lake Mary bluegills*

Description

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

Format

This data frame contains the following columns:

Age Years

Length mm

Source

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

References

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

Examples

```
head(lakemary)
```

`lakes` *Lake zooplankton diversity*

Description

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

Format

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

Species Number of zooplankton species

MaxDepth Maximum lake depth, m

MeanDepth Mean lake depth, m

Cond Specific conductance, micro Siemans

Elev Elevation, m

Lat N latitude, degrees

Long W longitude, degrees

Dist distance to nearest lake, km

NLakes number of lakes within 20 km

Photo Rate of photosynthesis, mostly by the 14C method

Area Lake area, in hectares

Source

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

References

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

Examples

```
head(lakes)
```

`landrent` *Land rent*

Description

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

Format

This data frame contains the following columns:

X1 average rent for all tillable land

X2 density of dairy cows (number per square mile)

X3 proportion of farmland used for pasture

X4 1 if liming required to grow alfalfa; 0 otherwise

Y average rent per acre planted to alfalfa

Source

Douglas Tiffany

References

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

Examples

```
head(landrent)
```

`lathe1` *Lathe data*

Description

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

Format

This data frame contains the following columns:

Feed Coded feed rate, coded as (actual feed rate -13)/6. Feed is in thousandths of an inch per revolution.

Speed Coded speed, coded as (actual speed -900)/300. Speed is in feet per minute.

Life Life of tool until failure, minutes

Source

M. R. Delozier

References

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

Examples

```
head(lathe1)
```

`mantel` *Mantel's artifical data for stepwise regression*

Description

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

Format

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

Y the response

X1 predictor 1

X2 predictor 2

X3 predictor 3

Source

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

References

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

Examples

```
head(mantel)
```

`mile` *World records for the mile run*

Description

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

Format

A data frame with 46 observations:

Year Year in which the record was set

Time Running time, in seconds

Name Name of person setting the record

Country Country of residence of the record setter

Place Place the record was set

Gender Gender of the record holder

Source

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

References

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

Examples

```
head(mile)
```

`Mitchell` *Mitchell soil temperature*

Description

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

Format

This data frame contains the following columns:

Month Months beginning Jan, 1976

Temp Average soil temperature, degrees C

Source

Kenneth G. Hubbard

References

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

Examples

```
head(Mitchell)
```

`MWwords` *Word frequencies from Mosteller and Wallace*

Description

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

Format

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

Hamilton Hamilton frequency

HamiltonRank Hamilton rank

Madison Madison frequency

MadisonRank Madison rank

Jay Jay frequency

JayRank Jay rank

Ulysses Word frequency in *Ulysses*

UlyssesRank Word rank in *Ulysses*

Source

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

References

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

Examples

```
head(MWwords)
```

`npdata` *Northern pike catch per unit effort*

Description

Catch per unit effort data for 16 Minnesota lakes

Format

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

CPUE Estimated catch per unit effect

SECPUE Estimated standard error of CPUE

Density Estimated fish density

SEdens Estimated standard error of Density

Source

R. Pierce, Minnesota Dept. of Natural Resources

References

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

Examples

```
head(npdata)
```

`oldfaith` *Old Faithful Geyser data*

Description

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

Format

This data frame contains the following columns:

Duration Duration in seconds

Interval Time to next eruption

Source

R. Hutchinson

References

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

Examples

```
head(oldfaith)
```

`physics` *Physics data*

Description

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

Format

This data frame contains the following columns:

x Inverse total energy

y Scattering cross-section/sec

SD Standard deviation

Source

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

References

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

Examples

```
head(physics1)
```

pipeline *Alaska pipeline*

Description

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

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

Format

This data frame contains the following columns:

Field Number of defects measured in the field.

Lab Number of defects measured in the field.

Batch Batch number

Source

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

References

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

Examples

```
head(pipeline)
```

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

Description

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

Usage

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

pod (x, ...)

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

## S3 method for class 'formula'
pod(formula, data = sys.parent(), group, subset, weights,
    na.action, mean.function = c("pod", "common", "parallel",
        "general"), singular.ok = FALSE, contrasts = NULL, offset,
    control = nls.control(), ...)


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

## S3 method for class 'lm'
 pod(x, group, mean.function, control, ...)

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

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

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

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

Arguments

`formula`	A linear regression formula, as used in lm, or the results of a call to pod or a lm fit. See details below.
`x`	The result of a lm fit from which the formula and the arguments data, subset, weights, na.action, singular.ok, contrasts and offset will be taken.
`data`	An optional data frame for the data to be used
`group`	The name of a grouping variable (not quoted) that defines the groups; see details below.

`mean.function`	Which mean function should be fit? The default is "pod", that fits the partial one-dimensional mean function. The other options are "common", which fits a linear model with no grouping effects; "parallel" fits a parallel within-group regression, and "general", available in pod but not plot.pod, fits a separate coefficeint for each term in the model for each level of the grouping variable.
`subset`	As in lm, the subset of cases to be used in the analysis
`weights`	Weights will be used in fitting non-pod models. Since this argument is not supported for nls models, weights are ignored for fitting pod models. If nls is ever updated, then pod models will correctly use the weights.
`na.action`	At present, only na.omit is likely to work.
`singular.ok`	The default equal to `FALSE` is recommended.
`contrasts`	Same as in lm
`offset`	Same as in lm
`control`	A pod model is fit using the nonlinear least squares routine nls. This routine is very sensitive to starting values and other parameters set in the algorithm. This routine selects starting values for you that will often work very well, but in some problems the user may want to change the defaults to nls program using the `nls.control` function.
`...`	In pod, other arguments passed to nls, such as control parameters. In pod.anova, there are two additional arguments
`object`	The result of a call to pod.
`scale`	Used for test in anova
`test`	Default is to compute F tests.
`colors`	Colors for groups in the pod plot
`pch`	Ploting symbol for the groups in the pod plot
`identify`	If TRUE, clicking the mouse on a graph will print the case name of the nearest point. This continues until turned off (by pushing the escape key, among other ways of doing this).
`key`	The default is `"topleft"`, in which case a legend is added to the top left corner of the plot; other choices include `"bottomright"`. If `key` is a vector of two coordinates, the legend is drawn at the coordinates specified. If `key` is FALSE, no key is drawn; if TRUE, you can place the key interactively by clicking on the plot.
`xlab`	Horizontal label, optional
`ylab`	Vertical label, optional

Details

Suppose we start with a linear mean function specified by `y ~ x1 + x2 + x3`, where the right-side variables can be any valid R variables, such as transformations, but NOT factors or interactions (if you want to include these, you need to create the dummy variables yourself). The right-hand side variables must also be linearly independent. We also specify a grouping variable z with, say, g levels. Let G_i be a dummy variable with values equal to 1 when $z = i$, and zero otherwise. The pod mean function is then has the nonlinear mean function

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

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

The pod mean function is easily fit using `nls`. For example, if z has two levels, a nonlinear mean function that would work is `y ~ b0 + b1*x1 + b2*x2 + b3*x3 + G2*(th02 + th12*(b1*x1 + b2*x2 + b3*x3))`. Starting values can be determined by fitting the parallel mean function, for this is a linear mean function and easily fit.

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

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

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

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

Value

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

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

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

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

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

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

See Also

See Also nls, lm, nls.control

Examples

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

prodscore *Soil productivity*

Description

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

Format

This data frame contains the following columns:

County Name of the county

Value Assessed value in dollars per acre.

P Productivity score, a number between 1 and 100.

Year Tax year, either 1981 or 1982.

Source

Douglas Tiffany

References

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

Examples

```
head(prodscore)
```

`pureErrorAnova` *Pure Error analysis of variance*

Description

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

Usage

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

pureErrorAnova(mod)

## S3 method for class 'lm'
pureErrorAnova(mod)

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

Arguments

`mod`	an object of type `lm`

Details

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

Value

Returns an analsis of variance table.

Author(s)

Sanford Weisberg, sandy@stat.umn.edu

References

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

See Also

`lm`

Examples

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

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

---

randomLinComb — *Compute a random linear combination of the columns of a matrix or data frame*

Description

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

Usage

```
randomLinComb(X, seed = NULL)

## S3 method for class 'lm'
 randomLinComb(X, seed=NULL)
```

Arguments

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

Details

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

Value

A vector of length n.

Author(s)

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

See Also

pure.error.anova

Examples

```
randomLinComb(matrix(rnorm(9),ncol=3))
```

`rat` *Rat data*

Description

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

Format

This data frame contains the following columns:

BodyWt BodyWt of the rat

LiverWt LiverWt measured after sacrifice

Dose Dose, roughly proportional to body weight

y dose of drug recovered after sacrifice of the animal

Source

Dennis Cook

References

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

Examples

```
head(rat)
```

`salary` *Salary data*

Description

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

Format

This data frame contains the following columns:

Degree Degree, 1 if PhD, 0 if Masters

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

Sex 1 if female, 0 if male

Year Years in current rank

YSdeg Years since highest degree earned

Salary dollars per year

Source

Sanford Weisberg

References

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

Examples

```
head(salary)
```

`salarygov` *Government salary study*

Description

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

Format

This data frame contains the following columns:

JobClass Name of job class

NW Number of women employees

NE Total number of employees in a job class

Score Difficulty score for job class

MaxSalary Maximum salary for job class

Source

Sanford Weisberg

References

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

Examples

```
head(salarygov)
```

`segreg` *Energy consumption*

Description

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

Format

This data frame contains the following columns:

Temp Monthly mean temperature, degrees F.

C Electricty consumption in KWH/day

Source

Charles Ng

References

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

Examples

```
head(segreg)
```

`shocks` *Small electric shocks in dairy cows*

Description

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

Format

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

Intensity Shock level, milliamps

m Number of trials

Y Number of times a positive reaction was observed

Source

R. Norell

References

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

Examples

```
head(shocks)
```

`sleep1` *Sleep in mammals*

Description

Includes species averages for 62 mammals.

Format

This data frame uses spcies as row lable and contains the following columns:

SWS Slow wave nondreaming sleep, hrs/day

PS Paradoxical dreaming sleep, hrs/day

TS Total sleep, hrs/day

BodyWt Body weight in kg

BrainWt Brain weight in g

Life Maximum life span, years

GP Gestation time, days

P Predation index, 1=low,5=hi

SE Sleep exposure index 1=exposed, 5=protected

D Danger index, 1=least, 5=most

Source

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

References

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

Examples

```
head(sleep1)
```

---

snake *Snake river data*

Description

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

Format

This data frame contains the following columns:

X water content of snow

Y water yield from April to July

Source

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

References

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

Examples

```
head(snake)
```

`sniffer` *Sniffer data*

Description

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

Format

This data frame contains the following columns:

TankTemp Initial tank temperature (degrees F)

GasTemp Temperature of the dispensed gasoline (degrees F)

TankPres Initial vapor pressure in the tank (psi)

GasPres Vapor pressure of the dispensed gasoline (psi)

Y Hydrocarbons emitted (grams)

Source

John Rice

References

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

Examples

```
head(sniffer)
```

`snowgeese` *Snow geese*

Description

Counts of flocks of snow geese.

Format

This data frame contains the following columns:

photo Photo count

obs1 Observer 1 count (aerial)

obs2 Observer 2 count (aerial)

Source

Dennis Cook

References

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

Examples

```
head(snowgeese)
```

`stopping` *Stopping distances*

Description

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

Format

This data frame contains the following columns:

Speed Speed (mph)

Distance Stopping distance (in feet)

Source

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

References

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

Examples

```
head(stopping)
```

---

swan96 *Black crappie study on Swan Lake, Minnesota*

Description

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

Format

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

Day Number of days after June 16, 1996

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

Source

Minnesota Department of Natural Resources

References

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

Examples

```
head(swan96)
```

`titanic` *Titanic*

Description

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

Format

This data frame contains the following columns:

Surv Number of survivors

N Survivors + Deaths

Class Crew or passanger class

Age Adult or child

Sex Male or female

Source

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

References

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

Examples

```
head(titanic)
```

`transact` *Transaction data*

Description

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

Format

This data frame contains the following columns:

T1 number of type 1 transactions

T2 number of type 2 transactions

Time total transaction time, minutes

Source

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

References

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

Examples

```
head(transact)
```

---

turk0 *Turkey data, one source*

Description

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

Format

This data frame contains the following columns:

A Amount of methionine supplement (percent of diet)

Gain Pen weight increase (g)

Source

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

References

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

Examples

```
head(turk0)
```

`turkey` *Turkey data, all sources*

Description

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

Format

This data frame contains the following columns:

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

Gain Ave. weight gain, over all replications

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

m Number of replications or pens

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

Source

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

References

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

Examples

```
head(turkey)
```

`twins` *Burt's twin data*

Description

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

Format

This data frame contains the following columns:

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

IQb biological

IQf foster

Source

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

References

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

Examples

```
head(twins)
```

`ufc` *Height-diameter data for Upper Flat Creek, Idaho*

Description

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

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

Format

A data frame with the following 5 variables.

Plot Plot number

Tree Tree within plot

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

Dbh Diameter 137 cm perpendicular to the bole, mm

Height Height of the tree, in decimeters

Source

Andrew Robinson

References

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

Examples

```
head(ufcgf)
```

UN1 *National statistics from the United Nations*

Description

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

Format

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

Fertility Expected number of live births per female, 2000

PPgdp Per capita 2001 GDP, in US \$

Details

These data were collected at published by the UN from a variety of sources. See original source for additional footnotes concerning values for individual countries. Country names are given in the first column of the data file. The same data, plus additional variables, is given in the file UN3.

Source

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

References

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

Examples

```
head(UN1)
```

UN2 *National statistics from the United Nations*

Description

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

Format

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

logFertility log Expected number of live births per female, 2000

logPPgdp Log Per capita 2001 GDP, in US \$

Purban Percent of population that is urban, 2001

Details

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

Source

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

References

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

Examples

```
head(UN2)
```

UN3 *National statistics from the United Nations*

Description

National health, welfare, and education statistics for 210 places, mostly UN members, but also other areas like Hong Kong that are not independent countries. Only 125 localities have data present for all the variables.

Format

This data frame uses locality name as row labels. It contains the following columns:

ModernC Percent of unmarried women using a modern method of contraception.

Change Annual population growth rate, percent.

PPgdp Per capita 2001 GDP, in US \$.

Frate Percent of females over age 15 economically active.

Pop Population, thousands.

Fertility Expected number of live births per female, 2000

Purban Percent of population that is urban, 2001

Details

These data were collected 2000-2003 and published by the UN from a variety of sources. See original source for additional footnotes concerning values for individual countries. Missing values in the data file indicated values for which no data is available, and will generally occur in less-developed localities.

Source

More recent values for these variables can be obtained from the following web pages: (1) `ModernC`: http://unstats.un.org/unsd/demographic/products/socind/contraceptive.htm; (2) `change`: http://unstats.un.org/unsd/demographic/products/socind/population.htm; (3) `PPgdp`: http://unstats.un.org/unsd/demographic/products/socind/inc-eco.htm; (4) `Frate`: http://unstats.un.org/unsd/demographic/products/socind/inc-eco.htm; (5) `Pop`: http://unstats.un.org/unsd/demographic/products/socind/population.htm; (6) `Fertility`: http://unstats.un.org/unsd/demographic/products/indwm/tab2c.htm; (7) `Purban`: http://unstats.un.org/unsd/demographic/products/socind/hum-sets.htm.

References

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

Examples

```
head(UN3)
```

`walleye` *Walleye length at age*

Description

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

Format

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

age Age of the fish, years

length Length, mm

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

Source

Michelle LeBeau

References

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

Examples

```
head(walleye)
```

---

water *California water*

Description

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

Format

This data frame contains the following columns:

Year collection year

APMAM Snowfall in inches measurement site

APSAB Snowfall in inches measurement site

APSLAKE Snowfall in inches measurement site

OPBPC Snowfall in inches measurement site

OPRC Snowfall in inches measurement site

OPSLAKE Snowfall in inches measurement site

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

Source

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

References

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

Examples

```
head(water)
```

`wblake` *West Bearskin Lake small mouth bass data.*

Description

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

Format

This data frame contains the following columns:

Age Age at capture (yrs)

Length Length at capture (mm)

Scale radius of a key scale, mm

Source

Minnesta Department of Natural Resources

References

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

Examples

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

`wm1` *Simple windmill data*

Description

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

Format

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

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

CSpd Windspeed in m/s at the candidate site

RSpd Windspeed for the reference site

Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

References

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

Examples

```
head(wm1)
```

`wm2` *Windmill data with direction information*

Description

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

Format

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

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

CSpd Windspeed in m/s at the candidate site

RSpd Windspeed for the reference site

RDir Wind direction, in degrees, at the reference site

Bin Wind direction binned into 16 equal width bins

Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

References

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

Examples

```
head(wm2)
```

---

wm3 *Binned wind speed data*

Description

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

Format

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

bin.count Number of observations in the bin

RSpd Average windspeed in the bin

References

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

Examples

```
head(wm3)
```

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

Description

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

Format

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

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

CSpd Wind speed (m/s) at candidate site

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

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

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

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

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

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

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

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

Spd1sin1 Spd times sin(direction) at site 1

Spd1cos1 Spd times cos(direction) at site 1

bin1 Wind direction bin for site 1

Dir1 Wind direction for site 1, in degrees

Source

Mark Ahlstrom and Rolf Miller, WindLogics, Inc.

References

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

Examples

```
head(wm4)
```

`wool` *Wool data*

Description

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

Format

This data frame contains the following columns:

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

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

Load load (40, 45, 50g)

Cycles number of cycles until failure

Source

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

References

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

Examples

```
head(wool)
```

Index

∗Topic **datasets**
ais, 4
allshoots, 5
baeskel, 8
banknote, 8
BGSall, 9
BigMac2003, 10
blowdown, 11
brains, 11
cakes, 12
cathedral, 13
caution, 13
challeng, 14
chloride, 15
cloud, 15
domedata, 16
donner, 17
downer, 17
drugcost, 18
dwaste, 19
florida, 20
forbes, 21
ftcollinssnow, 21
fuel2001, 22
galapagos, 23
galtonpeas, 24
heights, 24
highway, 25
hooker, 26
htwt, 27
jevons, 27
lakemary, 28
lakes, 29
landrent, 30
lathe1, 30
mantel, 31
mile, 32
Mitchell, 32
MWwords, 33
npdata, 34
oldfaith, 34
physics, 35
pipeline, 36
prodscore, 40
rat, 43
salary, 44
salarygov, 44
segreg, 45
shocks, 46
sleep1, 46
snake, 47
sniffer, 48
snowgeese, 49
stopping, 49
swan96, 50
titanic, 51
transact, 51
turk0, 52
turkey, 53
twins, 53
ufc, 54
UN1, 55
UN2, 56
UN3, 56
walleye, 57
water, 58
wblake, 59
wm1, 60
wm2, 60
wm3, 61
wm4, 62
wool, 63
∗Topic **hplot**
pod, 36
∗Topic **interface**
alrWeb, 7
∗Topic **manip**
randomLinComb, 42

∗Topic **models**
pod, 36
pureErrorAnova, 41
∗Topic **package**
alr3-package, 3
∗Topic **regression**
pod, 36
pureErrorAnova, 41

ais, 4
allshoots, 5
alr3 *(*alr3-package*)*, 3
alr3-deprecated, 5
alr3-package, 3
alrWeb, 7
anova.pod *(*pod*)*, 36

baeskel, 8
banknote, 8
bctrans *(*alr3-deprecated*)*, 5
bctrans1 *(*alr3-deprecated*)*, 5
BGSall, 9
BGSboys *(*BGSall*)*, 9
BGSgirls *(*BGSall*)*, 9
BigMac2003, 10
blowAPB *(*blowdown*)*, 11
blowBF *(*blowdown*)*, 11
blowdown, 11
boot.case *(*alr3-deprecated*)*, 5
brains, 11

cakes, 12
car, *7*
cathedral, 13
caution, 13
challeng, 14
chloride, 15
cloud, 15
coef.pod *(*pod*)*, 36
conf.intervals *(*alr3-deprecated*)*, 5

delta.method *(*alr3-deprecated*)*, 5
deviance.pod *(*pod*)*, 36
df.residual.pod *(*pod*)*, 36
domedata, 16
domedata1 *(*domedata*)*, 16
donner, 17
downer, 17
drugcost, 18
dwaste, 19

fitted.pod *(*pod*)*, 36
florida, 20
forbes, 21
formula.pod *(*pod*)*, 36
ftcollinssnow, 21
fuel2001, 22

galapagos, 23
galtonpeas, 24

heights, 24
highway, 25
hooker, 26
htwt, 27

inf.index *(*alr3-deprecated*)*, 5
inv.res.plot *(*alr3-deprecated*)*, 5
inv.tran.estimate *(*alr3-deprecated*)*, 5
inv.tran.plot *(*alr3-deprecated*)*, 5
inverse.response.plot
*(*alr3-deprecated*)*, 5

jevons, 27

lakemary, 28
lakes, 29
landrent, 30
lathe1, 30
lm, *40, 41*
longshoots *(*allshoots*)*, 5
lrt.bctrans *(*alr3-deprecated*)*, 5

mantel, 31
marginal.model.plot *(*alr3-deprecated*)*, 5
marginal.model.plots *(*alr3-deprecated*)*,
5
mile, 32
Mitchell, 32
MWwords, 33

nls, *40*
nls.control, *38*, *40*
npdata, 34

oldfaith, 34
outlier.t.test *(*alr3-deprecated*)*, 5

physics, 35

physics1 *(physics)*, 35
pipeline, 36
plot.pod *(pod)*, 36
pod, 36
podnls.fit *(pod)*, 36
podresponse *(pod)*, 36
powtran *(alr3-deprecated)*, 5
predict.pod *(pod)*, 36
print.pod *(pod)*, 36
prodscore, 40
pure.error.anova, *42*
pure.error.anova *(alr3-deprecated)*, 5
pureErrorAnova, 41

randomLinComb, 42
rat, 43
resid.curv.test *(alr3-deprecated)*, 5
resid.pod *(pod)*, 36
residual.plots *(alr3-deprecated)*, 5
residuals.pod *(pod)*, 36
resplot *(alr3-deprecated)*, 5

salary, 44
salarygov, 44
segreg, 45
shocks, 46
shortshoots *(allshoots)*, 5
sigma.hat *(alr3-deprecated)*, 5
sleep1, 46
snake, 47
sniffer, 48
snowgeese, 49
stopping, 49
summary.pod *(pod)*, 36
swan96, 50

titanic, 51
transact, 51
tukey.nonadd.test *(alr3-deprecated)*, 5
turk0, 52
turkey, 53
twins, 53

ufc, 54
ufcdf *(ufc)*, 54
ufcgf *(ufc)*, 54
ufcwc *(ufc)*, 54
UN1, 55
UN2, 56
UN3, *55*, 56

vcov.pod *(pod)*, 36

walleye, 57
water, 58
wblake, 59
wblake2 *(wblake)*, 59
wm1, 60
wm2, 60
wm3, 61
wm4, 62
wool, 63