

Package ‘pubh’

November 2, 2018

Title A Toolbox for Public Health and Epidemiology

Version 0.4.3

Maintainer Josie Athens <josie.athens@otago.ac.nz>

Description A toolbox for making R functions and capabilities more accessible to students and professionals from Epidemiology and Public Health related disciplines. Includes a function to report coefficients and confidence intervals from models using robust standard errors (when available), functions that expand lattice plots and functions relevant for introductory papers in Epidemiology or Public Health. Please note that use of the provided data sets is for educational purposes only.

Depends R (>= 3.5.0), lattice, latticeExtra, tactile, stats

Imports car, bookdown, descTable, effects, Epi, epiR, epitools, knitr, latex2exp, lmtest, lme4, MASS, multcomp, nlme, nnet, ordinal, pander, papeR, sandwich, survival

Suggests epibasix, epiDisplay, haven, ISwR, psych, rmarkdown

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 6.1.0

VignetteBuilder knitr

URL <https://github.com/josie-athens/pubh>

BugReports <https://github.com/josie-athens/pubh/issues>

NeedsCompilation no

Author Josie Athens [aut, cre],
Deepayan Sarkar [ctb],
Mark Stevenson [ctb],
Simon Firestone [ctb],
Cord Heuer [ctb],
Jim Robison-Cox [ctb],
Frank Harell [ctb],

John Fox [ctb],
 Kazuki Yoshida [ctb],
 Tomas Aragon [ctb],
 Johan Larsson [ctb],
 R-Core [ctb]

Repository CRAN

Date/Publication 2018-11-02 18:10:06 UTC

R topics documented:

| | |
|----------------|----|
| bar_error | 3 |
| Bernard | 4 |
| bland_altman | 5 |
| box_plot | 6 |
| Brenner | 6 |
| bst | 7 |
| chisq.fisher | 8 |
| coef_det | 8 |
| coef_plot | 9 |
| contingency | 10 |
| contingency2 | 11 |
| diag_test | 12 |
| diag_test2 | 13 |
| estat | 14 |
| expand_df | 15 |
| Fentress | 16 |
| freq_cont | 16 |
| gen_bst_df | 17 |
| geo_mean | 18 |
| glm_coef | 19 |
| harm_mean | 20 |
| hist_norm | 21 |
| Hodgkin | 22 |
| inv_logit | 22 |
| jack_knife | 23 |
| Kirkwood | 24 |
| knife_mean | 24 |
| leverage | 25 |
| logistic_gof | 26 |
| Macmahon | 26 |
| mhor | 27 |
| odds_trend | 28 |
| Oncho | 29 |
| panel.errbars | 30 |
| panel.errbars1 | 30 |
| predict_inv | 31 |
| prop_or | 32 |

| | |
|---------------------------|----|
| pseudo_r2 | 32 |
| qq_plot | 33 |
| rank_influence | 34 |
| rank_leverage | 35 |
| reference_range | 36 |
| rel_dis | 36 |
| Roberts | 37 |
| Rothman | 38 |
| round_pval | 39 |
| rvf_plot | 39 |
| Sandler | 40 |
| Sharples | 41 |
| ss_jk | 41 |
| stats_quotes | 42 |
| strip_error | 43 |
| Thall | 44 |
| Tuzson | 45 |
| Vanderpump | 45 |
| xymultiple | 46 |

| | |
|--------------|-----------|
| Index | 48 |
|--------------|-----------|

| | |
|-----------|------------------------------------|
| bar_error | <i>Bar charts with error bars.</i> |
|-----------|------------------------------------|

Description

bar_error constructs bar charts in lattice with error bars showing 95 confidence intervals around mean values. High of bars represent mean values.

Usage

```
bar_error(formula, data, col = "gray70", aspect = 3/4, ...)
```

Arguments

| | |
|---------|--|
| formula | A formula of the form $y \sim x$ or $y \sim x z$ where y is a numerical variable and both x and z are factors. |
| data | A data frame where the variables in the formula can be found. |
| col | Colour passed to barchart . |
| aspect | Physical aspect ratio passed to barchart . |
| ... | Further arguments passed to barchart . |

Details

Limits for the y-axis have to be estimated; lower limit should be zero and upper limit higher than the maximum upper confidence interval.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
 Deepayan Sarkar, R-Core.

Examples

```
data(birthwt, package = "MASS")
birthwt$smoke <- factor(birthwt$smoke, labels = c("Non-smoker", "Smoker"))
gen_bst_df(bwt ~ smoke, data = birthwt) # To estimate limits of y-axis.
bar_error(bwt ~ smoke, data = birthwt, ylab = "Birth weight (g)", ylim = c(0, 3500))

birthwt$race <- factor(birthwt$race, labels = c("White", "African American", "Other"))
gen_bst_df(bwt ~ race|smoke, data = birthwt) # To estimate limits of y-axis.
bar_error(bwt ~ race|smoke, data = birthwt, ylab = "Birth weight (g)", ylim = c(0, 3800))

bar_error(bwt ~ race|smoke, data = birthwt, ylab = "Birth weight (g)", ylim = c(0, 3800),
          col = c("gray95", "gray20", "gray50"))
```

 Bernard

Survival of patients with sepsis.

Description

A randomised, double-blind, placebo-controlled trial of intravenous ibuprofen in 455 patients who had sepsis, defined as fever, tachycardia, tachypnea, and acute failure of at least one organ system.

Usage

Bernard

Format

A data frame with 455 rows and 9 variables:

id Patient ID

treat Treatment, factor with levels "Placebo" and "Ibuprofen".

race Race/ethnicity, factor with levels "White", "African American" and "Other".

fate Mortality status at 30 days, factor with levels "Alive" and "Dead".

apache Baseline APACHE score.

o2del Oxygen delivery at baseline.

followup Follow-up time in hours.

temp0 Baseline temperature in centigrades.

temp10 Temperature after 36 hr in centigrades.

Source

Bernard, GR, et al. (1997) The effects of ibuprofen on the physiology and survival of patients with sepsis, *N Engl J Med* 336: 912–918.

Examples

```
contingency(fate ~ treat, data = Bernard)
```

| | |
|--------------|--------------------------------------|
| bland_altman | <i>Bland-Altman agreement plots.</i> |
|--------------|--------------------------------------|

Description

Bland-Altman agreement plots.

Usage

```
bland_altman(formula, data, transform = FALSE, aspect = 3/4, ...)
```

Arguments

| | |
|-----------|---|
| formula | A formula of the form $y \sim x$ (see details). |
| data | A data frame where the variables in the formula can be found. |
| transform | Logical, should ratios instead of difference be used to construct the plot? |
| aspect | Physical aspect ratio passed to xyplot . |
| ... | Further arguments passed to xyplot . |

Details

`bland_altman` constructs Bland-Altman agreement plots.

Variables in `formula` are continuous paired observations. When the distribution of the outcome is not normal, but becomes normal with a log-transformation, `bland_altman` can plot the ratio between outcomes (difference in the log scale) by using option `transform = TRUE`.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Deepayan Sarkar, R-Core.

Examples

```
data(wright, package = "ISwR")
bland_altman(mini.wright ~ std.wright, data = wright, pch = 16)
bland_altman(mini.wright ~ std.wright, data = wright, pch = 16,
             ylab = "Large-mini expiratory flow rate (l/min)",
             xlab = "Mean expiratory flow rate (l/min)")
```

| | |
|----------|---|
| box_plot | <i>Construct "pretty" box plots in lattice.</i> |
|----------|---|

Description

box_plot is a wrap function that calls [bwplot](#) to construct more aesthetic box plots.

Usage

```
box_plot(formula, data, box.fill = "gray70", box.ratio = 0.7,  
         aspect = 3/4, ...)
```

Arguments

| | |
|-----------|---|
| formula | A formula of the form $y \sim x$ where y is a numerical variable and x is a factor. |
| data | A data frame where the variables in the formula can be found. |
| box.fill | Colour used for the box passed to bwplot . |
| box.ratio | Ratio of box passed to bwplot . |
| aspect | Physical aspect ratio passed to bwplot . |
| ... | Further arguments passed to bwplot . |

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Deepayan Sarkar, R-Core.

Examples

```
data(kfm, package = "ISwR")  
box_plot(dl.milk ~ sex, data = kfm, ylab = "Breast-milk intake (dl/day)")
```

| | |
|---------|---|
| Brenner | <i>Prevalence of Helicobacter pylori infection in preschool children.</i> |
|---------|---|

Description

A data set containing the prevalence of *Helicobacter pylori* infection in preschool children according to parental history of duodenal or gastric ulcer.

Usage

Brenner

Format

A data frame with 863 rows and 2 variables:

Ulcer History of duodenal or gastric ulcer, factor with levels "No" and "Yes".

Infected Infected with *Helicobacter pylori*, factor with levels "No" and "Yes".

Source

Brenner H, Rothenbacher D, Bode G, Adler G (1998) Parental history of gastric or duodenal ulcer and prevalence of *Helicobacter pylori* infection in preschool children: population based study. *BMJ* 316:665.

Examples

```
data(Brenner)
contingency(Infected ~ Ulcer, data = Brenner, method = "cross.sectional")
```

bst *Bootstrap Confidence Intervals.*

Description

bst estimates confidence intervals around the [mean](#), [median](#) or [geo_mean](#).

Usage

```
bst(x, stat = "mean", n = 1000, CI = 95)
```

Arguments

| | |
|------|---|
| x | A numerical variable. Missing observations are removed by default. |
| stat | Statistic, either "mean" (default), "median" or "gmean" (geometric mean). |
| n | Number of replicates for the bootstrap (n=1000 by default). |
| CI | Confidence intervals (CI=95 by default). |

Value

A data frame with the estimate and confidence intervals.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
data(IgM, package="ISwR")
bst(IgM, "median")

bst(IgM, "gmean")
```

| | |
|--------------|--|
| chisq.fisher | <i>Internal test for chi-squared assumption. Fisher (2 by 2). If results = T, it fails</i> |
|--------------|--|

Description

chisq.fisher is an internal function called by contingency and contingency2 that uses the Fisher exact test if results from the assumptions for the chi-squared test fail.

Usage

```
chisq.fisher(tab)
```

Arguments

| | |
|-----|-----------------------------|
| tab | A numeric two by two table. |
|-----|-----------------------------|

| | |
|----------|--------------------------------------|
| coef_det | <i>Coefficient of determination.</i> |
|----------|--------------------------------------|

Description

coef_det estimates the coefficient of determination (r-squared) from fitted (predicted) and observed values. Outcome from the model is assumed to be numerical.

Usage

```
coef_det(obs, fit)
```

Arguments

| | |
|-----|--|
| obs | Vector with observed values (numerical outcome). |
| fit | Vector with fitted (predicted) values. |

Value

A scalar, the coefficient of determination (r-squared).

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
## Linear regression:
Riboflavin <- seq(0, 80, 10)
OD <- 0.0125*Riboflavin + rnorm(9, 0.6, 0.03)
titration <- data.frame(Riboflavin, OD)
model1 <- lm(OD ~ Riboflavin, data=titration)
summary(model1)
coef_det(titration$OD, fitted(model1))

## Non-linear regression:
library(nlme)
data(Puromycin)
mm.tx <- gnls(rate ~ SSmicmen(conc, Vm, K), data=Puromycin, subset=state=="treated")
summary(mm.tx)
coef_det(Puromycin$rate[1:12], mm.tx$fitted)
```

coef_plot

*Plot of model coefficients.***Description**

coef_plot Constructs plot displaying estimates of parameters with bars representing confidence intervals.

Usage

```
coef_plot(model, labels = NULL, Exp = FALSE, CI = 0.95, pch = 20,
  col = 1, rotx = 45, ...)
```

Arguments

| | |
|--------|--|
| model | A regression object (like glm). |
| labels | Vector for the coefficient's names (intercept/constant should be omitted). |
| Exp | Logical, should estimates and confidence intervals be exponentiated? |
| CI | Proportion representing the confidence intervals. |
| pch | Point character passed to xyplot . |
| col | Colour passed to xyplot . |
| rotx | Degrees to rotate x-labels. |
| ... | Further arguments passed to xyplot . |

Details

coef_plot does not show estimate for the constant (intercept). Estimates and confidence intervals can be optionally exponentiated, in which case estimates would represent ratios instead of differences.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Deepayan Sarkar, R-Core.

Examples

```
data(birthwt, package = "MASS")
birthwt$smoke <- factor(birthwt$smoke, labels=c("Non-smoker", "Smoker"))
birthwt$race <- factor(birthwt$race > 1, labels=c("White", "Non-white"))
model1 <- glm(bwt ~ smoke + race, data = birthwt)
glm_coef(model1, labels=c("Constant", "Smoker vs Non-smoker", "Non-white vs White"))
coef_plot(model1)
```

contingency

Measures of association from two by two contingency tables (formula).

Description

contingency is a wrap that calls [epi.2by2](#) from package `epiR`.

Usage

```
contingency(formula, data, method = "cohort.count", ...)
```

Arguments

| | |
|---------|---|
| formula | A formula of the form <code>outcome ~ exposure</code> . |
| data | A data frame where the variables in the formula can be found. |
| method | A character string with options: <code>"cohort.count"</code> , <code>"cohort.time"</code> , <code>"case.control"</code> , or <code>"cross.sectional"</code> . |
| ... | Further arguments passed to epi.2by2 . |

Details

contingency uses a formula as a way to input variables.

contingency displays the contingency table as a way for the user to check that the reference levels in the categorical variables (outcome and exposure) are correct. Then displays measures of association (table from [epi.2by2](#)). It also reports either chi-squared test or exact Fisher's test; contingency checks which one of the tests two is appropriate.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Mark Stevenson, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Australia.

Cord Heuer, EpiCentre, IVABS, Massey University, Palmerston North, New Zealand.

Jim Robison-Cox, Department of Math Sciences, Montana State University, Montana, USA.

Kazuki Yoshida, Brigham and Women's Hospital, Boston Massachusetts, USA.

Simon Firestone, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Australia.

Examples

```
## A case-control study on the effect of alcohol on oesophageal cancer.
Freq <- c(386, 29, 389, 171)
status <- gl(2, 1, 4, labels=c("Control", "Case"))
alcohol <- gl(2, 2, labels=c("0-39", "40+"))
cancer <- data.frame(Freq, status, alcohol)
cancer <- expand_df(cancer)
contingency(status ~ alcohol, data = cancer, method = "case.control")
```

| | |
|--------------|---|
| contingency2 | <i>Measures of association from two by two contingency tables (direct input).</i> |
|--------------|---|

Description

contingency2 is a wrap that calls [epi.2by2](#) from package epiR.

Usage

```
contingency2(aa, bb, cc, dd, ...)
```

Arguments

| | |
|-----|--|
| aa | Number of cases where both exposure and outcome are present. |
| bb | Number of cases where exposure is present but outcome is absent. |
| cc | Number of cases where exposure is absent but outcome is present. |
| dd | Number of cases where both exposure and outcome are absent. |
| ... | Further arguments passed to epi.2by2 . |

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Mark Stevenson, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Australia.

Cord Heuer, EpiCentre, IVABS, Massey University, Palmerston North, New Zealand.

Jim Robison-Cox, Department of Math Sciences, Montana State University, Montana, USA.

Kazuki Yoshida, Brigham and Women's Hospital, Boston Massachusetts, USA.

Simon Firestone, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Australia.

Examples

```
## A case-control study on the effect of alcohol on oesophageal cancer.
Freq <- c(386, 29, 389, 171)
status <- gl(2, 1, 4, labels=c("Control", "Case"))
alcohol <- gl(2, 2, labels=c("0-39", "40+"))
cancer <- data.frame(Freq, status, alcohol)
cancer <- expand_df(cancer)

contingency2(171, 389, 29, 386, method = "case.control")
```

diag_test

Diagnostic tests from variables.

Description

diag_test is a wrap function that calls [epi.tests](#) from package epiR. It computes sensitivity, specificity and other statistics related with screening tests.

Usage

```
diag_test(formula, data, ...)
```

Arguments

| | |
|---------|---|
| formula | A formula of the form outcome ~ predictor (see details). |
| data | A data frame where the variables in the formula can be found. |
| ... | Further arguments passed to epi.tests . |

Details

For the formula, the outcome is the gold standard and the explanatory variable is the new (screening) test. See examples.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Mark Stevenson, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Australia.

Examples

```
## We compare the use of lung's X-rays on the screening of TB against the gold standard test.
Freq <- c(1739, 8, 51, 22)
BCG <- gl(2, 1, 4, labels=c("Negative", "Positive"))
Xray <- gl(2, 2, labels=c("Negative", "Positive"))
tb <- data.frame(Freq, BCG, Xray)
tb <- expand_df(tb)
diag_test(BCG ~ Xray, data=tb)
```

diag_test2

Diagnostic tests from direct input.

Description

diag_test2 is a wrap that calls [epi.tests](#) from package epiR. It computes sensitivity, specificity and other statistics related with screening tests.

Usage

```
diag_test2(aa, bb, cc, dd)
```

Arguments

| | |
|----|---|
| aa | Number of cases where both screening test and the gold standard are positive. |
| bb | Number of cases where screening test is positive but gold standard is negative. |
| cc | Number of cases where screening test is negative but gold standard is positive. |
| dd | Number of cases where both screening test and the gold standard are negative. |

Details

diag.test uses direct input variables.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Mark Stevenson, Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Australia.

Examples

```
## We compare the use of lung's X-rays on the screening of TB against the gold standard test.
diag_test2(22, 51, 8, 1739)
```

| | |
|-------|---|
| estat | <i>Descriptive statistics for continuous variables.</i> |
|-------|---|

Description

estat calculates descriptives of numerical variables.

Usage

```
estat(formula, data, digits = 2, label = NULL)
```

Arguments

| | |
|---------|--|
| formula | A formula of the form: $\sim x$ or $\sim x z$ (for groups). |
| data | A data frame where the variables in the formula can be found. |
| digits | Number of digits for rounding (default = 2). |
| label | Label used to display the name of the variable (see examples). |

Value

A data frame with descriptive statistics.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

See Also

[summary](#), `summarize`.

Examples

```
data(kfm, package = "ISwR")
estat(~ dl.milk, data = kfm, label = "Breast-milk intake (dl/day)")
estat(~ dl.milk|sex, data = kfm, label = "Breast-milk intake (dl/day)")
estat(~ weight|sex, data = kfm, label = "Weight of child (kg)")
```

| | |
|-----------|-----------------------------|
| expand_df | <i>Expand a data frame.</i> |
|-----------|-----------------------------|

Description

expand_df expands a data frame by a vector of frequencies.

Usage

```
expand_df(aggregate.data, index.var = "Freq", retain.freq = FALSE)
```

Arguments

aggregate.data A data frame.

index.var A numerical variable with the frequencies (counts).

retain.freq Logical expression indicating if frequencies should be kept.

Details

This is a generic function that resembles weighted frequencies in other statistical packages (for example, Stata). `expand_df` was adapted from a function developed by deprecated package `epicalc` (now package `epiDisplay`).

Value

An expanded data frame with replicates given by the frequencies.

Examples

```
Freq <- c(5032, 5095, 41, 204)
Mortality <- gl(2, 2, labels=c("No", "Yes"))
Calcium <- gl(2, 1, 4, labels=c("No", "Yes"))
anyca <- data.frame(Freq, Mortality, Calcium)
anyca
anyca.exp <- expand_df(anyca)
with(anyca.exp, table(Calcium, Mortality))
```

 Fentress

Migraine pain reduction.

Description

Randomised control trial on children suffering from frequent and severe migraine. Control group represents untreated children. The active treatments were either relaxation alone or relaxation with biofeedback.

Usage

Fentress

Format

A data frame with 18 rows and 2 variables:

pain Reduction in weekly headache activity expressed as percentage of baseline data.

group Group, a factor with levels "Untreated", "Relaxation" (alone) and "Biofeedback" (relaxation and biofeedback).

Source

Fentress, DW, et al. (1986) Biofeedback and relaxation-response in the treatment of pediatric migraine. *Dev Med Child Neurol* 28:1 39-46.

Altman, DA (1991) *Practical statistics for medical research*. Chapman & Hall/CRC.

Examples

```
strip_error(pain ~ group, data = Fentress, ylab = "Pain reduction (%)")
```

 freq_cont

Relative and Cumulative Frequency.

Description

freq_cont tabulates a continuous variable by given classes.

Usage

```
freq_cont(x, bks, dg = 2)
```


Arguments

| | |
|-----|--|
| x | A numerical (continuous) variable. Ideally, relatively long (greater than 100 observations). |
| bks | Breaks defining the classes (see example). |
| dg | Number of digits for rounding (default = 2). |

Value

A data frame with the classes, the mid-point, the frequencies, the relative and cumulative frequencies.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

See Also

[hist](#)

Examples

```
data(IgM, package="ISwR")
Ab <- data.frame(IgM)
estat(~ IgM, data = Ab)
freq_cont(IgM, seq(0, 4.5, 0.5))
```

gen_bst_df

Generate a data frame with estimate and bootstrap CIs.

Description

gen_bst_df is an internal function called by [bar_error](#) that generates a data frame with confidence intervals of a continuous variable by levels of one or two categorical ones (factors).

Usage

```
gen_bst_df(formula, data, stat = "mean", ...)
```

Arguments

| | |
|---------|--|
| formula | A formula of the form $y \sim x$ or $y \sim x z$ where y is a numerical variable and both x and z are factors. |
| data | A data frame where the variables in the formula can be found. |
| stat | Statistic used for bst . |
| ... | Passes optional arguments to bst . |

Value

A data frame with the confidence intervals by level.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
data(kfm, package = "ISwR")
gen_bst_df(dl.milk ~ sex, data = kfm)
bar_error(dl.milk ~ sex, data = kfm, ylim = c(0,9), ylab = "Breast-milk intake (dl/day)")

data(birthwt, package = "MASS")
birthwt$smoke <- factor(birthwt$smoke, labels=c("Non-smoker", "Smoker"))
birthwt$Race <- 0
birthwt$Race[birthwt$race>1] <- 1
birthwt$Race <- factor(birthwt$Race, labels=c("White", "Non-white"))
gen_bst_df(bwt ~ smoke|Race, data = birthwt)
bar_error(bwt ~ smoke|Race, data = birthwt, ylim = c(0, 3800), ylab = "Birth weight (g)")
```

geo_mean

Geometric mean.

Description

Geometric mean.

Usage

```
geo_mean(x)
```

Arguments

x A numeric variable with no negative values.

Value

A scalar, the calculated geometric mean.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
data(IgM, package = "ISwR")
Ab <- data.frame(IgM)
estat(~ IgM, data = Ab)
geo_mean(IgM)
```

glm_coef *Table of coefficients from generalised linear models.*

Description

glm_coef displays estimates with confidence intervals and p-values from generalised linear models (see Details).

Usage

```
glm_coef(model, digits = 2, alpha = 0.05, labels = NULL,
         se.rob = TRUE, type = "cond")
```

Arguments

| | |
|--------|--|
| model | A model from any of the classes listed in the details section. |
| digits | A scalar, number of digits for rounding the results (default = 2). |
| alpha | Significant level (default = 0.05) used to calculate confidence intervals. |
| labels | An optional character vector with the names of the coefficients (including intercept). |
| se.rob | Logical, should robust errors be used to calculate confidence intervals? (default = TRUE). |
| type | Character, either "cond" (condensed) or "ext" (extended). See details. |

Details

glm_coef recognises objects (models) from the following classes: c1m, clogit, coxph, gee, glm, glmrMod, lm, lme, multinom, negbin, polr and survreg

For models from logistic regression (including conditional logistic, ordinal and multinomial), Poisson or survival analysis, coefficient estimates and corresponding confidence intervals are automatically exponentiated (back-transformed).

By default, glm_coef uses robust standard errors for calculating confidence intervals.

glm_coef can display two different data frames depending on the option of type, for type type = "cond" (the default), glm_coef displays the standard table of coefficients with confidence intervals and p-values; for type = "ext", glm_coef displays each number in a different column and includes standard errors.

Please read the Vignette on Regression for more details.

Value

A data frame with estimates, confidence intervals and p-values from glm objects.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
## Continuous outcome.
data(birthwt, package = "MASS")
birthwt$smoke <- factor(birthwt$smoke, labels=c("Non-smoker", "Smoker"))
birthwt$race <- factor(birthwt$race > 1, labels=c("White", "Non-white"))
model_norm <- glm(bwt ~ smoke + race, data = birthwt)
glm_coef(model_norm)
glm_coef(model_norm, labels=c("Constant", "Smoker vs Non-smoker", "Non-white vs White"))

## Logistic regression.
data(diet, package = "Epi")
model_binom <- glm(chd ~ fibre, data = diet, family = binomial)
glm_coef(model_binom, labels = c("Constant", "Fibre intake (g/day)"))
glm_coef(model_binom, labels = c("Constant", "Fibre intake (g/day)", type = "ext")

## Poisson regression.
library(MASS)
data(quine)
levels(quine$Eth) <- list(White = "N", Aboriginal = "A")
levels(quine$Sex) <- list(Male = "M", Female = "F")
model_pois <- glm(Days ~ Eth + Sex + Age, family = poisson, data = quine)
glm_coef(model_pois)
deviance(model_pois) / df.residual(model_pois) # to check for overdispersion

model_negbin <- glm.nb(Days ~ Eth + Sex + Age, data = quine)
unadj <- glm_coef(model_negbin, labels=c("Constant",
                                         "Race: Aboriginal/White",
                                         "Sex: Female/Male",
                                         "F1/Primary",
                                         "F2/Primary",
                                         "F3/Primary"))
unadj # Not-adjusted for multiple comparisons

## For more examples, please read the Vignette on Regression.
```

harm_mean

Harmonic mean.

Description

Harmonic mean.

Usage

```
harm_mean(x)
```

Arguments

x A numeric variable with no zero values.

Value

A scalar, the calculated harmonic mean.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
data(IgM, package = "ISwR")
Ab <- data.frame(IgM)
estat(~ IgM, data = Ab)
harm_mean(IgM)
```

hist_norm

Histogram with Normal density curve.

Description

hist_norm constructs histograms in lattice and adds corresponding Normal density curve.

Usage

```
hist_norm(formula, data = NULL, col = "gray70", aspect = 3/4, ...)
```

Arguments

| | |
|---------|---|
| formula | A formula of the form $\sim x$ where x is a numerical variable. |
| data | A data frame where the variables in the formula can be found. |
| col | Colour passed to histogram . |
| aspect | Physical aspect ratio passed to histogram . |
| ... | Further arguments passed to histogram . |

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Deepayan Sarkar, R-Core.

Examples

```
data(birthwt, package = "MASS")
hist_norm(~ bwt, data = birthwt, nint = 15, xlab = "Birth weight (g)")
```

Hodgkin

T-cell counts from Hodgkin's disease patients.

Description

Number of CD4+ T-cells and CD8+ T-cells in blood samples from patients in remission from Hodgkin's disease or in remission from disseminated malignancies.

Usage

Hodgkin

Format

A data frame with 40 rows and 3 variables:

CD4 Concentration of CD4+ T-cells (cells / mm³).

CD8 Concentration of CD8+ T-cells (cells / mm³).

Group Group, factor with levels "Non-Hodgkin" and "Hodgkin".

Source

Shapiro, CM, et al (1986) Immunologic status of patients in remission from Hodgkin's disease and disseminated malignancies. *Am J Med Sci* 293:366-370.

Altman, DA (1991) *Practical statistics for medical research*. Chapman & Hall/CRC.

Examples

```
data(Hodgkin)
Hodgkin$Ratio <- Hodgkin$CD4/Hodgkin$CD8
estat(~ Ratio|Group, data = Hodgkin, label = "CD4+/CD8+ T-cells")
qq_plot(~ Ratio|Group, data = Hodgkin, ylab = "CD4+/CD8+ T-cells", aspect = 1)
```

inv_logit

Inverse of the logit

Description

inv_logit Calculates the inverse of the logit (probability in logistic regression)

Usage

inv_logit(x)

Arguments

x Numerical value used to compute the inverse of the logit.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

jack_knife *Ranks leverage observations from Jackknife method.*

Description

jack_knife Ranks the squared differences between mean values from Jackknife analysis (arithmetic mean estimated by removing one observation at a time) and the original mean value.

Usage

```
jack_knife(x)
```

Arguments

x A numeric variable. Missing values are removed by default.

Value

Data frame with the ranked squared differences.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

See Also

[rank_leverage](#)

Examples

```
x <- rnorm(10, 170, 8)
x
mean(x)
jack_knife(x)

x <- rnorm(100, 170, 8)
mean(x)
head(jack_knife(x))
```

Kirkwood *Body weight and plasma volume.*

Description

Body weight and plasma volume in eight healthy men.

Usage

Kirkwood

Format

A data frame with 8 rows and 3 variables:

subject Subject ID.

weight Body weight in kg.

volume Plasma volume in litres.

Source

Kirkwood, BR and Sterne, JAC (2003) Essential Medical Statistics. Second Edition. Blackwell.

Examples

```
rbind(estat(~ weight, data = Kirkwood, label = c("Weight (kg)")),
      estat(~ volume, data = Kirkwood, label = c("Volume (l)")))
```

knife_mean *Jackknife for means.*

Description

knife_mean is an internal function. Calculates arithmetic means by removing one observation at a time.

Usage

knife_mean(x)

Arguments

x A numerical variable. Missing values are removed for the mean calculation.

Value

A vector with the mean calculations.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
x <- rnorm(10, 170, 8)
x
mean(x)
knife_mean(x)
```

leverage

Leverage.

Description

leverage is an internal function called by [rank_leverage](#).

Usage

```
leverage(x)
```

Arguments

x A numeric variable. Missing values are removed by default.

Details

Estimates the leverage of each observation around the arithmetic mean.

Value

Variable with corresponding leverage estimations

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
x <- rnorm(10, 170, 8)
x
mean(x)
leverage(x)
rank_leverage(x)
```

`logistic_gof`*Goodness of fit for Logistic Regression.*

Description

`logistic_gof` performs the Hosmer and Lemeshow test to test the goodness of fit of a logistic regression model. This function is part of `residuals.lrm` from package `rms`.

Usage

```
logistic_gof(model)
```

Arguments

`model` A logistic regression model object.

Author(s)

Frank Harrell, Vanderbilt University <f.harrell@vanderbilt.edu>

References

Hosmer DW, Hosmer T, Lemeshow S, le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat in Med* 16:965–980, 1997.

Examples

```
data(diet, package = "Epi")
model <- glm(chd ~ fibre, data = diet, family = binomial)
glm_coef(model, labels = c("Constant", "Fibre intake (g/day)"))
logistic_gof(model)
```

`Macmahon`*Breast cancer and age of childbirth.*

Description

An international case-control study to test the hypothesis that breast cancer is related to the age that a woman gives childbirth.

Usage

```
Macmahon
```

Format

A data frame with 185 rows and 2 variables:

Cancer Diagnosed with breast cancer, a factor with levels "No" and "Yes".

Age Age mother gives childbirth, factor with levels "<20", "20-24", "25-29", "30-34" and ">34".

Source

Macmahon, B. et al. (1970). Age at first birth and breast cancer risk. Bull WHO 43, 209-221.

Examples

```
odds_trend(Cancer ~ Age, data = Macmahon)
```

| | |
|------|------------------------------------|
| mhor | <i>Mantel-Haenszel odds ratio.</i> |
|------|------------------------------------|

Description

mhor computes odds ratios by levels of the stratum variable as well as the Mantel-Haenszel pooled odds ratio. The test for effect modification (test for interaction) is also displayed.

Usage

```
mhor(formula, data)
```

Arguments

formula A formula expressed as outcome ~ stratum/exposure.

data A data frame containing the variables used in formula.

Value

Odds ratios with 95 outcome by levels of stratum. The Mantel-Haenszel pooled OR and the test for effect modification is also reported.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

See Also

[mh](#)

Examples

```
data(oswego, package = "epitools")
oswego$ill <- factor(oswego$ill)
oswego$sex <- factor(oswego$sex)
oswego$chocolate.ice.cream <- factor(oswego$chocolate.ice.cream)
mhor(ill ~ sex/chocolate.ice.cream, data = oswego)
```

odds_trend

Function to calculate OR using Wald CI, and plot trend.

Description

odds_trend calculates the odds ratio with confidence intervals (Wald) for different levels (three or more) of the exposure variable, constructs the corresponding plot and calculates if the trend is significant or not.

Usage

```
odds_trend(formula, data, ...)
```

Arguments

| | |
|---------|---|
| formula | A formula of the form outcome ~ exposure. |
| data | A data frame where the variables in the formula can be found. |
| ... | Passes optional arguments to oddsratio . |

Details

odds_trend is a wrap function that calls [oddsratio](#) from package epitools.

Value

Displays odds ratio, analysis of trend and plot.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Tomas Aragon, University of Berkeley, USA.

See Also

[oddsratio](#)

Examples

```
## A cross-sectional study looked at the association between obesity and a biopsy resulting
## from mammography screening.

Freq <- c(3441, 34, 39137, 519, 20509, 280, 12149, 196, 11882, 199)
Biopsy <- gl(2, 1, 10, labels = c("No", "Yes"))
Weight <- gl(5, 2, 10, labels = c("Underweight", "Normal", "Over (11-24%)",
                                "Over (25-39%)", "Over (> 39%)"))
breast <- data.frame(Freq, Biopsy, Weight)
breast

breast <- expand_df(breast)
odds_trend(Biopsy ~ Weight, data = breast)
```

Oncho

Onchocerciasis in Sierra Leone.

Description

Study of onchocerciasis ("river blindness") in Sierra Leone, in which subjects were classified according to whether they lived in villages in savannah or rainforest area.

Usage

Oncho

Format

A data frame with 1302 rows and 7 variables:

id Subject ID.

mf Infected with *Onchocerciasis volvulus*, factor with levels "Not-infected" and "Infected".

area Area of residence, factor with levels "Savannah" and "Rainforest".

agegrp Age group in years, factor with levels "5-9", "10-19", "20-39" and "40+".

sex Subject sex, factor with levels "Male" and "Female".

mflod Microfilariae load.

lesions Severe eye lesions, factor with levels "No" and "Yes".

Source

McMahon, JE, Sowa, SIC, Maude, GH and Kirkwood BR (1988) *Onchocerciasis in Sierra Leone 2: a comparison of forest and savannah villages*. *Trans Roy Soc Trop Med Hyg* 82: 595-600.

Kirkwood, BR and Sterne, JAC (2003) *Essential Medical Statistics*. Second Edition. Blackwell.

Examples

```
odds_trend(mf ~ agegrp, data = Oncho)
```

panel.errbars *Internal function for displaying error bars in lattice plots.*

Description

panel.errbars is an internal function called by bar_error and strip_error.

Usage

```
panel.errbars(x, y, ..., panel.xy = panel.xyplot,
             make.grid = c("horizontal", "vertical", "both", "none"), ewidth = 0)
```

Arguments

| | |
|-----------|--|
| x | A numeric vector with the x positions for the bars. |
| y | A numeric matrix with the mid values, lower values and upper values of the bars. |
| ... | Further arguments passed to lattice. |
| panel.xy | A lattice panel. |
| make.grid | Type of grid, a character with options: "horizontal", "vertical", "both" and "none". |
| ewidth | An integer. |

panel.errbars1 *Internal function for displaying error bars in lattice plots.*

Description

panel.errbars1 is an internal function called by panel.errbars1.

Usage

```
panel.errbars1(x, y0, y1, ewidth = 0)
```

Arguments

| | |
|--------|---|
| x | A numeric vector with the x positions for the bars. |
| y0 | A numeric vector with the lower limits of the bars. |
| y1 | A numeric vector with the upper limits of the bars. |
| ewidth | An integer. |

| | |
|-------------|--|
| predict_inv | <i>Given y solve for x in a simple linear model.</i> |
|-------------|--|

Description

predict_inv Calculates the value the predictor x that generates value y with a simple linear model.

Usage

```
predict_inv(model, y)
```

Arguments

| | |
|-------|--|
| model | A simple linear model object (class lm). |
| y | A numerical scalar, the value of the outcome for which we want to calculate the predictor x. |

Value

The estimated value of the predictor.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
## Spectrophotometry example. Titration curve for riboflavin (nmol/ml). The sample has an absorbance
## of 1.15. Aim is to estimate the concentration of riboflavin in the sample.

Riboflavin <- seq(0, 80, 10)
OD <- 0.0125 * Riboflavin + rnorm(9, 0.6, 0.03)
titration <- data.frame(Riboflavin, OD)

xyplot(OD ~ Riboflavin, data = titration, pch = 16, col = 1, aspect = 3/4) +
  layer(panel.smoother(lwd = 1.5, col = 2, method = "lm", ...))

## Model with intercept different from zero:
model <- lm(OD ~ Riboflavin, data = titration)
glm_coef(model)
predict_inv(model, 1.15)
```

| | |
|---------|--|
| prop_or | <i>Proportion, p1 from proportion p2 and OR.</i> |
|---------|--|

Description

prop_or is a simple function to calculate a proportion, from another proportion and the odds ratio between them.

Usage

```
prop_or(p2, or)
```

Arguments

| | |
|----|--|
| p2 | The value of a proportion in the unexposed group (p2). |
| or | The odds ratio of p1/p2. |

Value

p1, the proportion in the exposed group (p1).

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago.

Examples

```
flu <- matrix(c(20, 80, 220, 140), nrow = 2)
colnames(flu) <- c("Yes", "No")
rownames(flu) <- c("Vaccine", "Placebo")
flu

or <- (20 * 140) / (80 * 220)
p2 <- 80 / 220
prop_or(p2 = p2, or = or)
20 / 240
```

| | |
|-----------|--|
| pseudo_r2 | <i>Pseudo R2 (logistic regression) pseudo_r2 Calculates R2 analogues (pseudo R2) of logistic regression.</i> |
|-----------|--|

Description

Pseudo R2 (logistic regression) pseudo_r2 Calculates R2 analogues (pseudo R2) of logistic regression.

Usage

```
pseudo_r2(model)
```

Arguments

model A logistic regression model.

Details

pseudo_r2 calculates three pseudo R2 of logistic regression models: 1) Nagelkerke, 2) Cox and Snell, 3) Hosmer and Lemeshow.

Value

A data frame with the calculated pseudo R2 values.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
model_oncho <- glm(mf ~ area, data = Oncho, binomial)
glm_coef(model_oncho, labels = c("Constant", "Area (rainforest/savannah)"))
pseudo_r2(model_oncho)
```

qq_plot

Quantile-quantile plots against the standard Normal distribution.

Description

qq_plot constructs quantile-quantile plots against the standard normal distribution (also known as quantile-normal plots).

Usage

```
qq_plot(formula, data = NULL, pch = 20, col = 1, aspect = 3/4, ...)
```

Arguments

formula A formula of the form $y \sim x$ or $y \sim x|z$ where x is a numerical variable and z is a factor.

data A data frame where the variables in the formula can be found.

pch Point character passed to [qqmath](#).

col Colour passed to [qqmath](#).

aspect Physical aspect ratio passed to [qqmath](#).

... Further arguments passed to [qqmath](#).

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Deepayan Sarkar, R-Core.

Examples

```
data(kfm, package = "ISwR")
qq_plot(~ dl.milk, data = kfm, ylab = "Breast-milk intake (dl/day)")
qq_plot(~ dl.milk|sex, data = kfm, ylab = "Breast-milk intake (dl/day)", aspect = 1)
```

| | |
|----------------|--|
| rank_influence | <i>Ranks observations based upon influence measures on models.</i> |
|----------------|--|

Description

rank_influence calculates influence measures of each data observation on models and then ranks them.

Usage

```
rank_influence(model)
```

Arguments

model A generalised linear model object.

Details

rank_influence is a wrap function that calls [influence.measures](#), ranks observations on their significance influence on the model and displays the 10 most influential observations (if they are significant).

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Several R core team members and John Fox, originally in his car package.

Examples

```
data(diet, package = "Epi")
model <- glm(chd ~ fibre, data = diet, family = binomial)
rank_influence(model)
```

| | |
|---------------|--|
| rank_leverage | <i>Ranks observations by leverage.</i> |
|---------------|--|

Description

rank_leverage ranks observations by their leverage (influence) on the arithmetic mean.

Usage

```
rank_leverage(x)
```

Arguments

x A numeric variable. Missing values are removed by default.

Value

A data frame ranking observations by their leverage around the mean.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

See Also

[jack_knife](#)

Examples

```
x <- rnorm(10, 170, 8)
x
mean(x)
rank_leverage(x)

x <- rnorm(100, 170, 8)
mean(x)
head(rank_leverage(x))
```

| | |
|-----------------|--|
| reference_range | <i>Reference range (reference interval).</i> |
|-----------------|--|

Description

reference_range estimates the reference range (reference interval) of a numerical variable.

Usage

```
reference_range(avg, std)
```

Arguments

| | |
|-----|--|
| avg | The arithmetic mean (a scalar numerical value). |
| std | The standard deviation (a scalar numerical value). |

Details

The reference range assumes normality and represents the limits that would include 95 observations.

Value

A data frame with the reference range limits.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
x <- rnorm(100, 170, 8)
round(mean(x), 2)
round(sd(x), 2)

round(reference_range(mean(x), sd(x)), 2)
```

| | |
|---------|-----------------------------|
| rel_dis | <i>Relative Dispersion.</i> |
|---------|-----------------------------|

Description

Calculates the coefficient of variation (relative dispersion) of a variable. The relative dispersion is defined as the standard deviation over the arithmetic mean.

Usage

```
rel_dis(x)
```

Arguments

x A numerical variable. NA's observations are removed by default.

Value

The coefficient of variation (relative dispersion).

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
height <- rnorm(100, 170, 8)
rel_dis(height)
```

Roberts

Extracorporeal membrane oxygenation in neonates.

Description

A clinical trial on the value of extracorporeal membrane oxygenation for term neonates with severe respiratory failure. RCT compares active treatment against conventional management.

Usage

Roberts

Format

A data frame with 185 rows and 2 variables:

EMO Extracorporeal membrane oxygenation treatment, factor with levels "No" and "Yes".

Survived One year survival, factor with levels "No" and "Yes".

Source

Roberts, TE (1998) Extracorporeal Membrane Oxygenation Economics Working Group. Economic evaluation and randomised controlled trial of extracorporeal membrane oxygenation: UK collaborative trial. *Brit Med J* 317:911-16.

Examples

```
barchart(with(Roberts, table(EMO, Survived)), horizontal = FALSE, stack = FALSE,
         xlab = "EMO treatment", ylab = "Number of subjects",
         auto.key = list(title = "Survived", space = "right", cex = 0.8))
```

Rothman

*Oral contraceptives and stroke.***Description**

A case-control study of oral contraceptives and stroke in young women with presence or absence of hypertension. Cases represent thrombotic stroke and controls are hospital controls. The group of no hypertension includes normal blood pressure (<140/90 mm Hg) and borderline hypertension (140-159/90-94 mm Hg). Hypertension group includes moderate hypertension (160-179/95-109 mm Hg) and severe hypertension (180+/110+ mm Hg). This data has been used as an example of join exposure by Rothman for measuring interactions (see examples).

Usage

Rothman

Format

A data frame with 477 rows and 3 variables:

stroke Thrombotic stroke, factor with levels "No" and "Yes".

oc Current user of oral contraceptives, factor with levels "Non-user" and "User".

ht Hypertension, factor with levels "No" (<160/95 mm Hg) and "Yes".

Source

Collaborative Group for the Study of Stroke in Young Women (1975) Oral contraceptives and stroke in young women. JAMA 231:718-722.

Rothman, KJ (2002) Epidemiology. An Introduction. Oxford University Press.

Examples

```
data(Rothman)
mhor(stroke ~ ht/oc, data = Rothman)

## Model with standard interaction term:
model1 <- glm(stroke ~ ht*oc, data = Rothman, family = binomial)
glm_coef(model1)

## Model considering join exposure:
Rothman$join <- 0
Rothman$join[Rothman$oc == "Non-user" & Rothman$ht == "Yes"] <- 1
Rothman$join[Rothman$oc == "User" & Rothman$ht == "No"] <- 2
Rothman$join[Rothman$oc == "User" & Rothman$ht == "Yes"] <- 3
Rothman$join <- factor(Rothman$join, labels=c("Unexposed", "Hypertension", "OC user",
"OC and hypertension"))
with(Rothman, table(join))
```

```
model2 <- glm(stroke ~ join, data = Rothman, family = binomial)
glm_coef(model2)
odds_trend(stroke ~ join, data = Rothman)
```

round_pval *Rounding p-values.*

Description

round_pval is an internal function called by glm_coef to round p-values from model coefficients.

Usage

```
round_pval(pval)
```

Arguments

pval vector of p-values, numeric.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

rvf_plot *Residual vs Fitted plot.*

Description

rvf_plot plots studentized residuals against fitted values from [glm](#) objects using package `lattice`.

Usage

```
rvf_plot(model, pch = 20, col = 1, aspect = 3/4, ...)
```

Arguments

model A [glm](#) or [lm](#) object with a numerical outcome.
pch Point character passed to [xyplot](#).
col Colour passed to [xyplot](#).
aspect Physical aspect ratio passed to [xyplot](#).
... Further arguments passed to [xyplot](#).

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Deepayan Sarkar, R-Core.

Examples

```
data(thuesen, package = "ISwR")
model <- lm(short.velocity ~ blood.glucose, data = thuesen)
plot(model, which = 1)
rvf_plot(model)
```

Sandler

Passive smoking in adulthood and cancer risk.

Description

A case-control study to investigate the effects of passive smoking on cancer. Passive smoking was defined as exposure to the cigarette smoke of a spouse who smoked at least one cigarette per day for at least 6 months.

Usage

Sandler

Format

A data frame with 998 rows and 3 variables:

passive Passive smoker, factor with levels "No" and "Yes".

cancer Diagnosed with cancer, factor with levels "No" and "Yes".

smoke Active smoker, factor with levels "No" and "Yes".

Source

Sandler, DP, Everson, RB, Wilcox, AJ (1985). Passive smoking in adulthood and cancer risk. *Amer J Epidemiol*, 121: 37-48.

Examples

```
mhor(cancer ~ smoke/passive, data = Sandler)
```

Sharples

Measured and self-reported weight in New Zealand.

Description

Data on measured and self-reported weight from 40–50 year old participants in the 1989/1990 Life In New Zealand Survey.

Usage

Sharples

Format

A data frame with 343 rows and 4 variables:

srweight Self-reported weight in kg.

weight Measured weight in kg.

srbmi Body mass index calculated from self-reported weight and self-reported height in kg/m².

mbmi Body mass index calculated from measured weight and measured height in kg/m².

Source

Sharples, H, et al. (2012) Agreement between measured and self-reported height, weight and BMI in predominantly European middle-aged New Zealanders: findings from a nationwide 1989 survey. *New Zealand Med J* 125: 60-69.

Examples

```
bland_altman(srweight ~ weight, data = Sharples, transform = TRUE, pch = 20, col = 1,
             xlab = "Mean of weights (kg)", ylab = "Measured weight / Self-reported weight")
```

ss_jk

Sum of squares for Jackknife.

Description

ss_jk is an internal function called by `jack_knife`. It calculates the squared difference of a numerical variable around a given value (for example, the mean).

Usage

```
ss_jk(obs, stat)
```

Arguments

obs A numerical vector with no missing values (NA's).
stat The value of the statistic that is used as a reference.

Value

The squared difference between a variable and a given value.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Examples

```
x <- rnorm(10, 170, 8)
x
mean(x)
ss_jk(x, mean(x))
jack_knife(x)
```

stats_quotes

Internal function to calculate descriptive statistics.

Description

stats_quotes is an internal function called by estat.

Usage

```
stats_quotes(x, data2, digits = 2)
```

Arguments

x a numeric variable
data2 A data frame where x can be found.
digits Number of digits for rounding.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

strip_error

*Strip plots with error bars.***Description**

strip_error constructs strip plots in lattice with error bars showing 95 confidence intervals around mean values.

Usage

```
strip_error(formula, data, pch = 20, aspect = 3/4, ...)
```

Arguments

| | |
|---------|--|
| formula | A formula of the form $y \sim x$ or $y \sim x z$ where y is a numerical variable and both x and z are factors. |
| data | A data frame where the variables in the formula can be found. |
| pch | Point character passed to stripplot . |
| aspect | Physical aspect ratio passed to stripplot . |
| ... | Further arguments passed to stripplot . |

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.

Deepayan Sarkar, R-Core.

Examples

```
data(energy, package="ISwR")
strip_error(expend ~ stature, data = energy, xlab = "Stature", ylab = "Energy expenditure (MJ)")

## Adding an horizontal line to show significant difference:
fig <- strip_error(expend~stature, data=energy, xlab="Stature",
                  ylab="Energy expenditure (MJ)", ylim=c(5.5,14))
fig + layer(panel.segments(1, 13.3, 2, 13.3, lwd=1.5)) + layer(panel.text(1.5, 13.5, "*"))

data(birthwt, package = "MASS")
birthwt$smoke <- factor(birthwt$smoke, labels = c("Non-smoker", "Smoker"))
birthwt$Race <- factor(birthwt$race > 1, labels = c("White", "Non-white"))
strip_error(bwt ~ Race|smoke, data = birthwt, ylab = "Birth weight (g)")
```

Thall *RCT on the treatment of epilepsy.*

Description

Randomised control trial of an antiepileptic drug (prograbide), in which the number of seizures of 59 patients at baseline and other four follow-up visits were recorded.

Usage

Thall

Format

A data frame with 59 rows and 8 variables:

id Subject ID.

treat Treatment, factor with levels "Control" and "Prograbide".

base Number of seizures at baseline.

age Age in years at baseline.

y1 Number of seizures at year one follow-up.

y2 Number of seizures at year two follow-up.

y3 Number of seizures at year three follow-up.

y4 Number of seizures at year four follow-up.

Source

Thall, PF and Vail, SC (1990) Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 46: 657-671.

Stukel, TA (1993) Comparison of methods for the analysis of longitudinal data. *Statistics Med* 12: 1339-1351.

Shoukri, MM and Chaudhary, MA (2007) *Analysis of correlated data with SAS and R*. Third Edition. Chapman & Hall/CRC.

Examples

```
data(Thall)
```

```
c1 <- cbind(Thall[, c(1:5)], count = Thall$y1[, c(1:4, 6)])
c2 <- cbind(Thall[, c(1:4, 6)], count = Thall$y2[, c(1:4, 6)])
c3 <- cbind(Thall[, c(1:4, 7)], count = Thall$y3[, c(1:4, 6)])
c4 <- cbind(Thall[, c(1:4, 8)], count = Thall$y3[, c(1:4, 6)])
epilepsy <- rbind(c1, c2, c3, c4)
```

```
library(lme4, warn.conflicts = FALSE)
model_glmmer <- glmmer(count ~ treat + base + I(age - mean(age, na.rm = TRUE)) +
```

```

      (1|id), data=epilepsy, family=poisson)
glm_coef(model_glmmer, labels = c("Treatment (Prograbide/Control)",
      "Baseline count", "Age (years)"))

```

Tuzson *Peak knee velocity in walking at flexion and extension.*

Description

Data of peak knee velocity in walking at flexion and extension in studies about functional performance in cerebral palsy.

Usage

Tuzson

Format

A data frame with 18 rows and 2 variables:

flexion Peak knee velocity in gait: flexion (degree/s).

extension Peak knee velocity in gait: extension (degree/s).

Source

Tuzson, AE, Granata, KP, and Abel, MF (2003) Spastic velocity threshold constrains functional performance in cerebral palsy. Arch Phys Med Rehabil 84: 1363-1368.

Examples

```
cor.test(~ flexion + extension, data = Tuzson)
```

Vanderpump *Smoking and mortality in Wickham, England.*

Description

Data represents women participating in a health survey in Wickham, England in 1972-1974.

Usage

Vanderpump

Format

A data frame with 1314 rows and 3 variables:

vstatus Vitality status, factor with levels "Alive" and "Death".

smoker Smoking status, factor with levels "Non-smoker" and "Smoker".

agegrp Age group, factor with levels "18-44", "45-64" and "64+".

Source

Vanderpump, MP, et al (1996) *Thyroid*, 6:155-160.

Vittinghoff, E, Glidden, DV, Shiboski, SC and McCulloh, CE (2005) *Regression methods in Biostatistics*. Springer.

Examples

```
mhor(vstatus ~ agegrp/smoker, data = Vanderpump)
```

xymultiple

Multiple comparisons with plot.

Description

xymultiple displays results from post-doc analysis and constructs corresponding plot.

Usage

```
xymultiple(model.glht, method = "bonferroni", Exp = FALSE, dg = 2,
  plot = TRUE, rotx = 45, ...)
```

Arguments

| | |
|------------|--|
| model.glht | An object of class "glht" (from multiple comparisons). |
| method | Method passed to summary.glht (see details). |
| Exp | Logical, should results be exponentiated? (default = FALSE). |
| dg | Number of digits for rounding (default = 2). |
| plot | Logical, should a plot be constructed? (default = TRUE). |
| rotx | Degrees to rotate x-labels (default = 45). |
| ... | Passes additional information to xyplot. |

Details

The default adjusting method is "bonferroni". Other options are: "single-step", "Shaffer", "free", "holm", "hochberg", "hommel", "Westfall".

Value

A data frame with CIs and p-values adjusted for multiple comparisons.

Author(s)

Josie Athens, Department of Preventive and Social Medicine, University of Otago, New Zealand.
Deepayan Sarkar, R-Core.

See Also

`glht`, `glht-methods`.

Examples

```
library(multcomp)
data(birthwt)
birthwt$race <- factor(birthwt$race, labels = c("White", "Black", "Other"))
model1 <- aov(bwt ~ race, data = birthwt)
model1_glht <- glht(model1, linfct = mcp(race = "Tukey"))
xymultiple(model1_glht)

model2 <- glm(low ~ race, data = birthwt, family = binomial)
model2_glht <- glht(model2, linfct = mcp(race="Tukey"))
xymultiple(model2_glht, Exp = TRUE)
```

Index

*Topic **datasets**

- Bernard, [4](#)
 - Brenner, [6](#)
 - Fentress, [16](#)
 - Hodgkin, [22](#)
 - Kirkwood, [24](#)
 - Macmahon, [26](#)
 - Oncho, [29](#)
 - Roberts, [37](#)
 - Rothman, [38](#)
 - Sandler, [40](#)
 - Sharples, [41](#)
 - Thall, [44](#)
 - Tuzson, [45](#)
 - Vanderpump, [45](#)
-
- bar_error, [3](#), [17](#)
 - barchart, [3](#)
 - Bernard, [4](#)
 - bland_altman, [5](#)
 - box_plot, [6](#)
 - Brenner, [6](#)
 - bst, [7](#), [17](#)
 - bwplot, [6](#)
-
- chisq.fisher, [8](#)
 - coef_det, [8](#)
 - coef_plot, [9](#)
 - contingency, [10](#)
 - contingency2, [11](#)
-
- diag_test, [12](#)
 - diag_test2, [13](#)
-
- epi.2by2, [10](#), [11](#)
 - epi.tests, [12](#), [13](#)
 - estat, [14](#)
 - expand_df, [15](#)
-
- Fentress, [16](#)
 - freq_cont, [16](#)
-
- gen_bst_df, [17](#)
 - geo_mean, [7](#), [18](#)
 - glm, [9](#), [39](#)
 - glm_coef, [19](#)
-
- harm_mean, [20](#)
 - hist, [17](#)
 - hist_norm, [21](#)
 - histogram, [21](#)
 - Hodgkin, [22](#)
-
- influence.measures, [34](#)
 - inv_logit, [22](#)
-
- jack_knife, [23](#), [35](#), [41](#)
-
- Kirkwood, [24](#)
 - knife_mean, [24](#)
-
- leverage, [25](#)
 - lm, [39](#)
 - logistic_gof, [26](#)
-
- Macmahon, [26](#)
 - mean, [7](#)
 - median, [7](#)
 - mh, [27](#)
 - mhor, [27](#)
-
- odds_trend, [28](#)
 - oddsratio, [28](#)
 - Oncho, [29](#)
-
- panel.errbars, [30](#)
 - panel.errbars1, [30](#)
 - predict_inv, [31](#)
 - prop_or, [32](#)
 - pseudo_r2, [32](#)
-
- qq_plot, [33](#)
 - qqmath, [33](#)

rank_influence, 34
rank_leverage, 23, 25, 35
reference_range, 36
rel_dis, 36
Roberts, 37
Rothman, 38
round_pval, 39
rvf_plot, 39

Sandler, 40
Sharples, 41
ss_jk, 41
stats_quotes, 42
strip_error, 43
stripplot, 43
summary, 14
summary.glt, 46

Thall, 44
Tuzson, 45

Vanderpump, 45

xymultiple, 46
xyplot, 5, 9, 39