

Package ‘MKmisc’

April 27, 2009

Type Package

Title Miscellaneous Functions from M. Kohl

Version 0.4

Date 2009-04-27

Author Matthias Kohl

Maintainer Matthias Kohl <Matthias.Kohl@stamats.de>

Description Miscellaneous Functions from M. Kohl

Depends R(>= 2.7.0), stats, graphics, robustbase, RColorBrewer

Suggests gplots

License LGPL-3

URL <http://www.stamats.de/>

Repository CRAN

Date/Publication 2009-04-27 11:58:50

R topics documented:

MKmisc-package	2
binomCI	2
corDist	4
corPlot	6
fiveNS	8
heatmapCol	9
IQrange	10
madMatrix	11
madPlot	12
oneWayAnova	13
pairwise.fc	14

qboxplot	16
qbxp.stats	19
repMeans	21
twoWayAnova	22

Index	24
--------------	-----------

MKmisc-package	<i>Miscellaneous Functions from M. Kohl.</i>
----------------	--

Description

Miscellaneous Functions from M. Kohl.

Details

Package: MKmisc
 Type: Package
 Version: 0.4
 Date: 2009-01-10
 Depends: R(>= 2.7.0), stats, graphics, robustbase, RColorBrewer
 Suggests: gplots
 License: LGPL-3
 URL: <http://www.stamats.de/>

require(MKmisc)

Author(s)

Matthias Kohl <http://www.stamats.de>
 Maintainer: Matthias Kohl <matthias.kohl@stamats.de>

binomCI	<i>Confidence Intervals for Binomial Proportions</i>
---------	--

Description

This functions can be used to compute confidence intervals for binomial proportions.

Usage

```
binomCI(x, n, conf.level = 0.95, method = "wilson", rand = 123)
```

Arguments

<code>x</code>	number of successes
<code>n</code>	number of trials
<code>conf.level</code>	confidence level
<code>method</code>	character string specifying which method to use; see details.
<code>rand</code>	seed for random number generator; see details.

Details

The Wald interval is obtained by inverting the acceptance region of the Wald large-sample normal test.

The Wilson interval, which is the default, was introduced by Wilson (1927) and is the inversion of the CLT approximation to the family of equal tail tests of $p = p_0$. The Wilson interval is recommended by Agresti and Coull (1998) as well as by Brown et al (2001).

The Agresti-Coull interval was proposed by Agresti and Coull (1998) and is a slight modification of the Wilson interval. The Agresti-Coull intervals are never shorter than the Wilson intervals; cf. Brown et al (2001).

The Jeffreys interval is an implementation of the equal-tailed Jeffreys prior interval as given in Brown et al (2001).

The modified Wilson interval is a modification of the Wilson interval for x close to 0 or n as proposed by Brown et al (2001).

The modified Jeffreys interval is a modification of the Jeffreys interval for $x == 0 \mid x == 1$ and $x == n-1 \mid x == n$ as proposed by Brown et al (2001).

The Clopper-Pearson interval is based on quantiles of corresponding beta distributions. This is sometimes also called exact interval.

The arcsine interval is based on the variance stabilizing distribution for the binomial distribution.

The logit interval is obtained by inverting the Wald type interval for the log odds.

The Witting interval (cf. Beispiel 2.106 in Witting (1985)) uses randomization to obtain uniformly optimal lower and upper confidence bounds (cf. Satz 2.105 in Witting (1985)) for binomial proportions.

For more details we refer to Brown et al (2001) as well as Witting (1985).

Value

A list with components

<code>estimate</code>	the estimated probability of success.
<code>CI</code>	a confidence interval for the probability of success.

Note

A first version of this function appeared in R package SLMisc.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

A. Agresti and B.A. Coull (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *American Statistician*, **52**, 119-126.

L.D. Brown, T.T. Cai and A. Dasgupta (2001). Interval estimation for a binomial proportion. *Statistical Science*, **16**(2), 101-133.

H. Witting (1985). *Mathematische Statistik I*. Stuttgart: Teubner.

See Also

[binom.test](#), [binconf](#)

Examples

```
binomCI(x = 42, n = 43, method = "wald")
binomCI(x = 42, n = 43, method = "wilson")
binomCI(x = 42, n = 43, method = "agresti-coull")
binomCI(x = 42, n = 43, method = "jeffreys")
binomCI(x = 42, n = 43, method = "modified wilson")
binomCI(x = 42, n = 43, method = "modified jeffreys")
binomCI(x = 42, n = 43, method = "clopper-pearson")
binomCI(x = 42, n = 43, method = "arcsine")
binomCI(x = 42, n = 43, method = "logit")
binomCI(x = 42, n = 43, method = "witting")

## the confidence interval computed by binom.test
## corresponds to the Clopper-Pearson interval
binomCI(x = 42, n = 43, method = "clopper-pearson")$CI
binom.test(x = 42, n = 43)$conf.int
```

corDist

Correlation Distance Matrix Computation

Description

The function computes and returns the correlation and absolute correlation distance matrix computed by using the specified distance measure to compute the distances between the rows of a data matrix.

Usage

```
corDist(x, method = "pearson", diag = FALSE, upper = FALSE, abs = FALSE,
        use = "pairwise.complete.obs", ...)
```

Arguments

x	a numeric matrix or data frame
method	the correlation distance measure to be used. This must be one of "pearson", "spearman", "kandall", "cosine", "mcd" or "ogk", respectively. Any unambiguous substring can be given.
diag	logical value indicating whether the diagonal of the distance matrix should be printed by 'print.dist'.
upper	logical value indicating whether the upper triangle of the distance matrix should be printed by 'print.dist'.
abs	logical, compute absolute correlation distances
use	character, corresponds to argument use of function cor
...	further arguments to functions covMcd or covOGK, respectively.

Details

The function computes the Pearson, Spearman, Kendall or Cosine sample correlation and absolute correlation; confer Section 12.2.2 of Gentleman et al (2005). For more details about the arguments we refer to functions `dist` and `cor`. Moreover, the function computes the minimum covariance determinant or the orthogonalized Gnanadesikan-Kettenring estimator. For more details we refer to functions `covMcd` and `covOGK`, respectively.

Value

'corDist' returns an object of class "dist"; cf. `dist`.

Note

A first version of this function appeared in package SLMisc.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

- Gentleman R. Ding B., Dudoit S. and Ibrahim J. (2005). Distance Measures in DNA Microarray Data Analysis. In: Gentleman R., Carey V.J., Huber W., Irizarry R.A. and Dudoit S. (editors) Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer.
- P. J. Rousseeuw and A. M. Leroy (1987). Robust Regression and Outlier Detection. Wiley.
- P. J. Rousseeuw and K. van Driessen (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212-223.
- Pison, G., Van Aelst, S., and Willems, G. (2002), Small Sample Corrections for LTS and MCD, *Metrika*, 55, 111-123.
- Maronna, R.A. and Zamar, R.H. (2002). Robust estimates of location and dispersion of high-dimensional datasets; *Technometrics* 44(4), 307-317.
- Gnanadesikan, R. and John R. Kettenring (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* 28, 81-124.

See Also

[plot.cor](#)

Examples

```
## only a dummy example
M <- cor(matrix(rnorm(1000), ncol = 20))
D <- corDist(M)
```

corPlot

Plot of similarity matrix based on correlation

Description

Plot of similarity matrix. This function is a slight modification of function `plot.cor` of package "sma".

Usage

```
corPlot(x, new = FALSE, col, minCor, labels = FALSE,
        labcols = "black", title = "", cex.title = 1.2,
        protocol = FALSE, cex.axis = 0.8,
        cex.axis.bar = 1, signifBar = 2, ...)
```

Arguments

<code>x</code>	data or correlation matrix, respectively
<code>new</code>	If <code>new=FALSE</code> , <code>x</code> must already be a correlation matrix. If <code>new=TRUE</code> , the correlation matrix for the columns of <code>x</code> is computed and displayed in the image.
<code>col</code>	colors palette for image. If missing, the RdYlGn palette of RColorBrewer is used.
<code>minCor</code>	numeric value in [-1,1], used to adjust <code>col</code>
<code>labels</code>	vector of character strings to be placed at the tickpoints, labels for the columns of <code>x</code> .
<code>labcols</code>	colors to be used for the labels of the columns of <code>x</code> . <code>labcols</code> can have either length 1, in which case all the labels are displayed using the same color, or the same length as <code>labels</code> , in which case a color is specified for the label of each column of <code>x</code> .
<code>title</code>	character string, overall title for the plot.
<code>cex.title</code>	A numerical value giving the amount by which plotting text and symbols should be magnified relative to the default; cf. par , <code>cex.main</code> .
<code>protocol</code>	logical, display color bar without numbers
<code>cex.axis</code>	The magnification to be used for axis annotation relative to the current setting of 'cex'; cf. par .

`cex.axis.bar` The magnification to be used for axis annotation of the color bar relative to the current setting of 'cex'; cf. [par](#).

`signifBar` integer indicating the precision to be used for the bar.

... graphical parameters may also be supplied as arguments to the function (see [par](#)). For comparison purposes, it is good to set `zlim=c(-1,1)`.

Details

This functions generates the so called similarity matrix (based on correlation) for a microarray experiment.

If `min(x)`, respectively `min(cor(x))` is smaller than `minCor`, the colors in `col` are adjusted such that the minimum correlation value which is color coded is equal to `minCor`.

Value

`invisible()`

Note

A first version of this function appeared in package `SLmisc`.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

Sandrine Dudoit, Yee Hwa (Jean) Yang, Benjamin Milo Bolstad and with contributions from Natalie Thorne, Ingrid Loennstedt and Jessica Mar. `sma`: Statistical Microarray Analysis. <http://www.stat.berkeley.edu/users/terry/zarray/Html/smacode.html>

See Also

[plot.cor](#)

Examples

```
## only a dummy example
M <- cor(matrix(rnorm(1000), ncol = 20))
corPlot(M, minCor = min(M))
```

`fiveNS`*Five-Number Summaries*

Description

Function to compute five-number summaries (minimum, 1st quartile, median, 3rd quartile, maximum)

Usage

```
fiveNS(x, na.rm = TRUE, type = 7)
```

Arguments

<code>x</code>	numeric vector
<code>na.rm</code>	logical; remove NA before the computations.
<code>type</code>	an integer between 1 and 9 selecting one of nine quantile algorithms; for more details see quantile .

Details

In contrast to [fivenum](#) the functions computes the first and third quartile using function [quantile](#).

Value

A numeric vector of length 5 containing the summary information.

Author(s)

Matthias Kohl <Matthias.Kohl@stamats.de>

See Also

[fivenum](#), [quantile](#)

Examples

```
x <- rnorm(100)
fiveNS(x)
fiveNS(x, type = 2)
fivenum(x)
```

heatmapCol *Generate colors for heatmaps*

Description

This function modifies a given color vector as used for heatmaps.

Usage

```
heatmapCol(data, col, lim, na.rm = TRUE)
```

Arguments

data	matrix or data.frame; data which shall be displayed in a heatmap; ranging from negative to positive numbers.
col	vector of colors used for heatmap.
lim	constant colors are used for data below $-lim$ resp. above lim .
na.rm	logical; remove NA values.

Details

Colors below and above a specified value are kept constant. In addition, the colors are symmetrized.

Value

vector of colors

Note

A first version of this function appeared in package SLmisc.

Author(s)

Matthias Kohl <Matthias.Kohl@stamats.de>

Examples

```
data.plot <- matrix(rnorm(100*50, sd = 1), ncol = 50)
colnames(data.plot) <- paste("patient", 1:50)
rownames(data.plot) <- paste("gene", 1:100)
data.plot[1:70, 1:30] <- data.plot[1:70, 1:30] + 3
data.plot[71:100, 31:50] <- data.plot[71:100, 31:50] - 1.4
data.plot[1:70, 31:50] <- rnorm(1400, sd = 1.2)
data.plot[71:100, 1:30] <- rnorm(900, sd = 1.2)
nrcol <- 128
require(gplots)
require(RColorBrewer)
```

```
heatmap.2(data.plot, col = rev(colorRampPalette(brewer.pal(10, "RdBu"))(nrcol)), trace = "n",
  farbe <- heatmapCol(data = data.plot, col = rev(colorRampPalette(brewer.pal(10, "RdBu"))(nrcol)), trace = "n",
  heatmap.2(data.plot, col = farbe, trace = "none", tracecol = "black")
```

IQRange

*The Interquartile Range***Description**

computes interquartile range of the x values.

Usage

```
IQRange(x, na.rm = FALSE, type = 7)
```

Arguments

<code>x</code>	a numeric vector.
<code>na.rm</code>	logical. Should missing values be removed?
<code>type</code>	an integer between 1 and 9 selecting one of nine quantile algorithms; for more details see quantile .

Details

This function computes quartiles as $IQR(x) = \text{quantile}(x, 3/4) - \text{quantile}(x, 1/4)$. In contrast to [IQR](#) the argument `type` is added and can be used to select between different algorithms for the computation of quantiles. The default `type = 7` corresponds to the setting used in case of [IQR](#).

For normally $N(m, 1)$ distributed X , the expected value of $IQR(X)$ is $2 * \text{qnorm}(3/4) = 1.3490$, i.e., for a normal-consistent estimate of the standard deviation, use $IQR(x) / 1.349$.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.

See Also

[quantile](#), [IQR](#).

Examples

```
IQrange(rivers)

## identical to
IQR(rivers)

## but, e.g.
IQrange(rivers, type = 4)
IQrange(rivers, type = 5)
```

`madMatrix`*Compute MAD between columns of a matrix or data.frame*

Description

Compute MAD between columns of a matrix or data.frame. Can be used to create a similarity matrix for a microarray experiment.

Usage

```
madMatrix(x)
```

Arguments

`x` matrix or data.frame

Details

This functions computes the so called similarity matrix (based on MAD) for a microarray experiment; cf. Bunes et. al. (2004).

Value

matrix of MAD values between columns of `x`

Note

A first version of this function appeared in package SLmisc.

Author(s)

Matthias Kohl <Matthias.Kohl@stamats.de>

References

Andreas Bunes, Wolfgang Huber, Klaus Steiner, Holger Sueltmann, and Annemarie Poustka. arrayMagic: two-colour cDNA microarray quality control and preprocessing. Bioinformatics Advance Access published on September 28, 2004. doi:10.1093/bioinformatics/bti052

See Also

plotMAD

Examples

```
## only a dummy example
M <- madMatrix(matrix(rnorm(1000), ncol = 10))
madPlot(M)
```

madPlot

Plot of similarity matrix based on MAD

Description

Plot of similarity matrix based on MAD between microarrays.

Usage

```
madPlot(x, new = FALSE, col, maxMAD = 3, labels = FALSE,
        labcols = "black", title = "", protocol = FALSE, ...)
```

Arguments

<code>x</code>	data or correlation matrix, respectively
<code>new</code>	If <code>new=FALSE</code> , <code>x</code> must already be a matrix with MAD values. If <code>new=TRUE</code> , the MAD matrix for the columns of <code>x</code> is computed and displayed in the image.
<code>col</code>	colors palette for image. If missing, the RdYlGn palette of RColorBrewer is used.
<code>maxMAD</code>	maximum MAD value displayed
<code>labels</code>	vector of character strings to be placed at the tickpoints, labels for the columns of <code>x</code> .
<code>labcols</code>	colors to be used for the labels of the columns of <code>x</code> . <code>labcols</code> can have either length 1, in which case all the labels are displayed using the same color, or the same length as <code>labels</code> , in which case a color is specified for the label of each column of <code>x</code> .
<code>title</code>	character string, overall title for the plot.
<code>protocol</code>	logical, display color bar without numbers
<code>...</code>	graphical parameters may also be supplied as arguments to the function (see par). For comparison purposes, it is good to set <code>zlim=c(-1,1)</code> .

Details

This functions generates the so called similarity matrix (based on MAD) for a microarray experiment; cf. Bunes et. al. (2004). The function is similar to `corPlot` and `plot.cor`, respectively.

Note

A first version of this function appeared in package SLmisc.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

Sandrine Dudoit, Yee Hwa (Jean) Yang, Benjamin Milo Bolstad and with contributions from Natalie Thorne, Ingrid Loennstedt and Jessica Mar. sma: Statistical Microarray Analysis.

<http://www.stat.berkeley.edu/users/terry/zarray/Html/smacode.html>

Andreas Bunes, Wolfgang Huber, Klaus Steiner, Holger Suetmann, and Annemarie Poustka. arrayMagic: two-colour cDNA microarray quality control and preprocessing. Bioinformatics Advance Access published on September 28, 2004. doi:10.1093/bioinformatics/bti052

See Also

[plot.cor](#), [corPlot](#)

Examples

```
## only a dummy example
M <- madMatrix(matrix(rnorm(1000), ncol = 10))
madPlot(M)
```

oneWayAnova

A function for Analysis of Variance

Description

This function is a slight modification of function [Anova](#) of package "genefilter".

Usage

```
oneWayAnova(cov, na.rm = TRUE)
```

Arguments

cov	The covariate. It must have length equal to the number of columns of the array that the result of oneWayAnova will be applied to.
na.rm	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

The function returned by `oneWayAnova` uses `lm` to fit a linear model of the form `lm(x ~ cov)`, where `x` is the set of gene expressions. The F statistic for an overall effect is computed and the corresponding p-value is returned.

The function `Anova` instead compares the computed p-value to a prespecified p-value and returns `TRUE`, if the computed p-value is smaller than the prespecified one.

Value

`oneWayAnova` returns a function with bindings for `cov` that will perform a one-way ANOVA.

The covariate can be continuous, in which case the test is for a linear effect for the covariate.

Note

A first version of this function appeared in package `SLmisc`.

Author(s)

Matthias Kohl <Matthias.Kohl@stamats.de>

References

R. Gentleman, V. Carey, W. Huber and F. Hahne (2006). `genefilter`: methods for filtering genes from microarray experiments. R package version 1.13.7.

See Also

[Anova](#)

Examples

```
set.seed(123)
af <- oneWayAnova(c(rep(1,5), rep(2,5)))
af(rnorm(10))
```

`pairwise.fc`*Compute pairwise fold changes*

Description

This function computes pairwise fold changes. It also works for logarithmic data.

Usage

```
pairwise.fc(x, g, ave = mean, log = TRUE, base = 2, mod.fc = TRUE, ...)
```

Arguments

<code>x</code>	numeric vector.
<code>g</code>	grouping vector or factor
<code>ave</code>	function to compute the group averages.
<code>log</code>	logical. Is the data logarithmic?
<code>base</code>	If <code>log = TRUE</code> , the base which was used to compute the logarithms.
<code>mod.fc</code>	logical. Return modified fold changes? (see details)
<code>...</code>	optional arguments to <code>ave</code> .

Details

The function computes pairwise fold-changes between groups, where the group values are aggregated using the function which is given by the argument `ave`.

The fold-changes are returned in a slightly modified form if `mod.fc = TRUE`. Fold changes FC which are smaller than 1 are reported as to $-1/FC$.

The implementation is comparable to [pairwise.t.test](#).

Value

Vector with pairwise fold-changes.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

See Also

[pairwise.t.test](#)

Examples

```
set.seed(13)
x <- rnorm(100) ## assumed as log2-data
g <- factor(sample(1:4, 100, replace = TRUE))
levels(g) <- c("a", "b", "c", "d")
pairwise.fc(x, g)

## some small checks
res <- by(x, list(g), mean)
2^(res[[1]] - res[[2]]) # a vs. b
-1/2^(res[[1]] - res[[3]]) # a vs. c
2^(res[[1]] - res[[4]]) # a vs. d
-1/2^(res[[2]] - res[[3]]) # a vs. d
-1/2^(res[[2]] - res[[4]]) # b vs. d
2^(res[[3]] - res[[4]]) # c vs. d
```

qboxplot

*Box Plots***Description**

Produce box-and-whisker plot(s) of the given (grouped) values. In contrast to `boxplot` quartiles are used instead of hinges (which are not necessarily quartiles) the rest of the implementation is identical to `boxplot`.

Usage

```
qboxplot(x, ...)

## S3 method for class 'formula':
qboxplot(formula, data = NULL, ..., subset, na.action = NULL, type = 7)

## Default S3 method:
qboxplot(x, ..., range = 1.5, width = NULL, varwidth = FALSE,
         notch = FALSE, outline = TRUE, names, plot = TRUE,
         border = par("fg"), col = NULL, log = "",
         pars = list(boxwex = 0.8, staplewex = 0.5, outwex = 0.5),
         horizontal = FALSE, add = FALSE, at = NULL, type = 7)
```

Arguments

<code>formula</code>	a formula, such as <code>y ~ grp</code> , where <code>y</code> is a numeric vector of data values to be split into groups according to the grouping variable <code>grp</code> (usually a factor).
<code>data</code>	a <code>data.frame</code> (or list) from which the variables in <code>formula</code> should be taken.
<code>subset</code>	an optional vector specifying a subset of observations to be used for plotting.
<code>na.action</code>	a function which indicates what should happen when the data contain NAs. The default is to ignore missing values in either the response or the group.
<code>x</code>	for specifying data from which the boxplots are to be produced. Either a numeric vector, or a single list containing such vectors. Additional unnamed arguments specify further data as separate vectors (each corresponding to a component boxplot). NAs are allowed in the data.
<code>...</code>	For the <code>formula</code> method, named arguments to be passed to the default method. For the default method, unnamed arguments are additional data vectors (unless <code>x</code> is a list when they are ignored), and named arguments are arguments and graphical parameters to be passed to <code>bxp</code> in addition to the ones given by argument <code>pars</code> (and override those in <code>pars</code>).
<code>range</code>	this determines how far the plot whiskers extend out from the box. If <code>range</code> is positive, the whiskers extend to the most extreme data point which is no more than <code>range</code> times the interquartile range from the box. A value of zero causes the whiskers to extend to the data extremes.
<code>width</code>	a vector giving the relative widths of the boxes making up the plot.

varwidth	if <code>varwidth</code> is <code>TRUE</code> , the boxes are drawn with widths proportional to the square-roots of the number of observations in the groups.
notch	if <code>notch</code> is <code>TRUE</code> , a notch is drawn in each side of the boxes. If the notches of two plots do not overlap this is ‘strong evidence’ that the two medians differ (Chambers <i>et al.</i> , 1983, p. 62). See <code>boxplot.stats</code> for the calculations used.
outline	if <code>outline</code> is not true, the outliers are not drawn (as points whereas S+ uses lines).
names	group labels which will be printed under each boxplot. Can be a character vector or an expression (see <code>plotmath</code>).
boxwex	a scale factor to be applied to all boxes. When there are only a few groups, the appearance of the plot can be improved by making the boxes narrower.
staplewex	staple line width expansion, proportional to box width.
outwex	outlier line width expansion, proportional to box width.
plot	if <code>TRUE</code> (the default) then a boxplot is produced. If not, the summaries which the boxplots are based on are returned.
border	an optional vector of colors for the outlines of the boxplots. The values in <code>border</code> are recycled if the length of <code>border</code> is less than the number of plots.
col	if <code>col</code> is non-null it is assumed to contain colors to be used to colour the bodies of the box plots. By default they are in the background colour.
log	character indicating if x or y or both coordinates should be plotted in log scale.
pars	a list of (potentially many) more graphical parameters, e.g., <code>boxwex</code> or <code>outpch</code> ; these are passed to <code>bxp</code> (if <code>plot</code> is true); for details, see there .
horizontal	logical indicating if the boxplots should be horizontal; default <code>FALSE</code> means vertical boxes.
add	logical, if true <i>add</i> boxplot to current plot.
at	numeric vector giving the locations where the boxplots should be drawn, particularly when <code>add = TRUE</code> ; defaults to <code>1:n</code> where <code>n</code> is the number of boxes.
type	an integer between 1 and 9 selecting one of nine quantile algorithms; for more details see quantile .

Details

The generic function `qboxplot` currently has a default method (`qboxplot.default`) and a formula interface (`qboxplot.formula`).

If multiple groups are supplied either as multiple arguments or via a formula, parallel boxplots will be plotted, in the order of the arguments or the order of the levels of the factor (see [factor](#)).

Missing values are ignored when forming boxplots.

Value

List with the following components:

stats	a matrix, each column contains the extreme of the lower whisker, the lower hinge, the median, the upper hinge and the extreme of the upper whisker for one group/plot. If all the inputs have the same class attribute, so will this component.
n	a vector with the number of observations in each group.
conf	a matrix where each column contains the lower and upper extremes of the notch.
out	the values of any data points which lie beyond the extremes of the whiskers.
group	a vector of the same length as out whose elements indicate to which group the outlier belongs.
names	a vector of names for the groups.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) *The New S Language*. Wadsworth & Brooks/Cole.

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis*. Wadsworth & Brooks/Cole.

Murrell, P. (2005) *R Graphics*. Chapman & Hall/CRC Press.

See also [boxplot.stats](#).

See Also

[qbxp.stats](#) which does the computation, [bxp](#) for the plotting and more examples; and [stripchart](#) for an alternative (with small data sets).

Examples

```
## adapted examples from boxplot

## qboxplot on a formula:
qboxplot(count ~ spray, data = InsectSprays, col = "lightgray")
# *add* notches (somewhat funny here):
qboxplot(count ~ spray, data = InsectSprays,
          notch = TRUE, add = TRUE, col = "blue")

qboxplot(decrease ~ treatment, data = OrchardSprays,
          log = "y", col = "bisque")

rb <- qboxplot(decrease ~ treatment, data = OrchardSprays, col="bisque")
title("Comparing boxplot()s and non-robust mean +/- SD")

mn.t <- tapply(OrchardSprays$decrease, OrchardSprays$treatment, mean)
sd.t <- tapply(OrchardSprays$decrease, OrchardSprays$treatment, sd)
xi <- 0.3 + seq(rb$n)
points(xi, mn.t, col = "orange", pch = 18)
```

```

arrows(xi, mn.t - sd.t, xi, mn.t + sd.t,
       code = 3, col = "pink", angle = 75, length = .1)

## boxplot on a matrix:
mat <- cbind(Uni05 = (1:100)/21, Norm = rnorm(100),
            `5T` = rt(100, df = 5), Gam2 = rgamma(100, shape = 2))
qboxplot(as.data.frame(mat),
         main = "qboxplot(as.data.frame(mat), main = ...)")
par(las=1)# all axis labels horizontal
qboxplot(as.data.frame(mat), main = "boxplot(*, horizontal = TRUE)",
         horizontal = TRUE)

## Using 'at = ' and adding boxplots -- example idea by Roger Bivand :

qboxplot(len ~ dose, data = ToothGrowth,
         boxwex = 0.25, at = 1:3 - 0.2,
         subset = supp == "VC", col = "yellow",
         main = "Guinea Pigs' Tooth Growth",
         xlab = "Vitamin C dose mg",
         ylab = "tooth length",
         xlim = c(0.5, 3.5), ylim = c(0, 35), yaxs = "i")
qboxplot(len ~ dose, data = ToothGrowth, add = TRUE,
         boxwex = 0.25, at = 1:3 + 0.2,
         subset = supp == "OJ", col = "orange")
legend(2, 9, c("Ascorbic acid", "Orange juice"),
      fill = c("yellow", "orange"))

```

qbxp.stats

Box Plot Statistics

Description

This functions works identical to `boxplot.stats`. It is typically called by another function to gather the statistics necessary for producing box plots, but may be invoked separately.

Usage

```
qbxp.stats(x, coef = 1.5, do.conf = TRUE, do.out = TRUE, type = 7)
```

Arguments

<code>x</code>	a numeric vector for which the boxplot will be constructed (NAs and NaNs are allowed and omitted).
<code>coef</code>	it determines how far the plot ‘whiskers’ extend out from the box. If <code>coef</code> is positive, the whiskers extend to the most extreme data point which is no more than <code>coef</code> times the length of the box away from the box. A value of zero causes the whiskers to extend to the data extremes (and no outliers be returned).
<code>do.conf</code>	logical; if FALSE, the <code>conf</code> component will be empty in the result.

`do.out` logical; if `FALSE`, `out` component will be empty in the result.

`type` an integer between 1 and 9 selecting one of nine quantile algorithms; for more details see [quantile](#).

Details

The notches (if requested) extend to $\pm 1.58 \text{ IQR} / \sqrt{n}$. This seems to be based on the same calculations as the formula with 1.57 in Chambers *et al.* (1983, p. 62), given in McGill *et al.* (1978, p. 16). They are based on asymptotic normality of the median and roughly equal sample sizes for the two medians being compared, and are said to be rather insensitive to the underlying distributions of the samples. The idea appears to be to give roughly a 95% confidence interval for the difference in two medians.

Value

List with named components as follows:

`stats` a vector of length 5, containing the extreme of the lower whisker, the first quartile, the median, the third quartile and the extreme of the upper whisker.

`n` the number of non-NA observations in the sample.

`conf` the lower and upper extremes of the ‘notch’ (`if(do.conf)`). See the details.

`out` the values of any data points which lie beyond the extremes of the whiskers (`if(do.out)`).

Note that `$stats` and `$conf` are sorted in *increasing* order, unlike `S`, and that `$n` and `$out` include any $\pm \text{Inf}$ values.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

- Tukey, J. W. (1977) *Exploratory Data Analysis*. Section 2C.
- McGill, R., Tukey, J. W. and Larsen, W. A. (1978) Variations of box plots. *The American Statistician* **32**, 12–16.
- Velleman, P. F. and Hoaglin, D. C. (1981) *Applications, Basics and Computing of Exploratory Data Analysis*. Duxbury Press.
- Emerson, J. D and Strenio, J. (1983). Boxplots and batch comparison. Chapter 3 of *Understanding Robust and Exploratory Data Analysis*, eds. D. C. Hoaglin, F. Mosteller and J. W. Tukey. Wiley.
- Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983) *Graphical Methods for Data Analysis*. Wadsworth & Brooks/Cole.

See Also

[quantile](#), [boxplot.stats](#)

Examples

```
## adapted example from boxplot.stats
x <- c(1:100, 1000)
(b1 <- qbxp.stats(x))
(b2 <- qbxp.stats(x, do.conf=FALSE, do.out=FALSE))
stopifnot(b1$stats == b2$stats) # do.out=F is still robust
qbxp.stats(x, coef = 3, do.conf=FALSE)
## no outlier treatment:
qbxp.stats(x, coef = 0)

qbxp.stats(c(x, NA)) # slight change : n is 101
(r <- qbxp.stats(c(x, -1:1/0)))
stopifnot(r$out == c(1000, -Inf, Inf))
```

repMeans

*Compute mean of replicated spots***Description**

Compute mean of replicated spots where additionally spot flags may incorporated.

Usage

```
repMeans(x, flags, use.flags = NULL, ndups, spacing, method, ...)
```

Arguments

<code>x</code>	matrix or data.frame of expression values
<code>flags</code>	matrix or data.frame of spot flags; must have same dimension as <code>x</code>
<code>use.flags</code>	should flags be included and in which way; cf. section details
<code>ndups</code>	integer, number of replicates on chip. The number of rows of <code>x</code> must be divisible by <code>ndups</code>
<code>spacing</code>	the spacing between the rows of 'x' corresponding to replicated spots, <code>spacing = 1</code> for consecutive spots; cf. function unwrapdups in package "limma"
<code>method</code>	function to aggregate the replicated spots. If missing, the mean is used.
<code>...</code>	optional arguments to <code>method</code> .

Details

The incorporation of spot flags is controlled via argument `use.flags`.

`NULL`: flags are not used; minimum flag value of replicated spots is returned

`"max"`: only spots with flag value equal to the maximum flag value of replicated spots are used

`"median"`: only spots with flag values larger or equal to median of replicated spots are used

`"mean"`: only spots with flag values larger or equal to mean of replicated spots are used

Value

LIST with components

<code>exprs</code>	mean of expression values
<code>flags</code>	flags for mean expression values

Note

A first version of this function appeared in package `SLmisc`.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

See Also

[unwrapdups](#)

Examples

```
## only a dummy example
M <- matrix(rnorm(1000), ncol = 10)
FL <- matrix(rpois(1000, lambda = 10), ncol = 10) # only for this example
res <- repMeans(x = M, flags = FL, use.flags = "max", ndups = 5, spacing = 20)
```

twoWayAnova

A function for Analysis of Variance

Description

This function is a slight modification of function `Anova` of package "genefilter".

Usage

```
twoWayAnova(cov1, cov2, interaction, na.rm = TRUE)
```

Arguments

<code>cov1</code>	The first covariate. It must have length equal to the number of columns of the array that the result of <code>twoWayAnova</code> will be applied to.
<code>cov2</code>	The second covariate. It must have length equal to the number of columns of the array that the result of <code>twoWayAnova</code> will be applied to.
<code>interaction</code>	logical, should interaction be considered
<code>na.rm</code>	a logical value indicating whether 'NA' values should be stripped before the computation proceeds.

Details

The function returned by `twoWayAnova` uses `lm` to fit a linear model of the form `lm(x ~ cov1*cov2)`, where `x` is the set of gene expressions. The F statistics for the main effects and the interaction are computed and the corresponding p-values are returned.

Value

`twoWayAnova` returns a function with bindings for `cov1` and `cov2` that will perform a two-way ANOVA.

Note

A first version of this function appeared in package `SLmisc`.

Author(s)

Matthias Kohl (Matthias.Kohl@stamats.de)

References

R. Gentleman, V. Carey, W. Huber and F. Hahne (2006). `genefilter`: methods for filtering genes from microarray experiments. R package version 1.13.7.

See Also

[Anova](#)

Examples

```
set.seed(123)
af1 <- twoWayAnova(c(rep(1,6), rep(2,6)), rep(c(rep(1,3), rep(2,3)), 2))
af2 <- twoWayAnova(c(rep(1,6), rep(2,6)), rep(c(rep(1,3), rep(2,3)), 2), interaction = FALSE)
x <- matrix(rnorm(12*10), nrow = 10)
apply(x, 1, af1)
apply(x, 1, af2)
```

Index

- *Topic **distribution**
 - fiveNS, 7
 - IQrange, 9
- *Topic **dplot**
 - qbxp.stats, 18
- *Topic **hplot**
 - heatmapCol, 8
 - madPlot, 11
 - qboxplot, 15
- *Topic **htest**
 - oneWayAnova, 13
 - twoWayAnova, 22
- *Topic **models**
 - oneWayAnova, 13
 - twoWayAnova, 22
- *Topic **multivariate**
 - corDist, 4
- *Topic **package**
 - MKmisc-package, 1
- *Topic **robust**
 - IQrange, 9
- *Topic **univar**
 - binomCI, 2
 - corPlot, 6
 - fiveNS, 7
 - IQrange, 9
 - madMatrix, 10
 - pairwise.fc, 14
 - repMeans, 20
- Anova, 13, 14, 22
- binconf, 3
- binom.test, 3
- binomCI, 2
- boxplot, 15
- boxplot.stats, 16, 17, 19, 20
- bxp, 16, 17
- cor, 4, 5
- corDist, 4
- corPlot, 6, 12
- covMcd, 4, 5
- covOGK, 4, 5
- dist, 5
- expression, 16
- factor, 17
- fiveNS, 7
- fivenum, 8
- heatmapCol, 8
- IQR, 10
- IQrange, 9
- lm, 13, 22
- madMatrix, 10
- madPlot, 11
- MKmisc (MKmisc-package), 1
- MKmisc-package, 1
- NA, 16, 19
- NaN, 19
- oneWayAnova, 13
- pairwise.fc, 14
- pairwise.t.test, 14, 15
- par, 6, 12
- plot.cor, 5, 7, 12
- plotmath, 16
- qboxplot, 15
- qbxp.stats, 17, 18
- quantile, 8–10, 17, 19, 20
- repMeans, 20

stripchart, *17*

twoWayAnova, [22](#)

unwrapdups, *21*