Package 'granova'

March 22, 2023

Version 2.2

Date 2023-03-21

Title Graphical Analysis of Variance

Author Frederic Bertrand [cre] (<https://orcid.org/0000-0002-0837-8281>),
Robert M. Pruzek [aut],
James E. Helmreich [aut]

Maintainer Frederic Bertrand <frederic.bertrand@utt.fr>

Depends R (>= 3.1.1), car (>= 2.0-21)

Suggests mgcv, rgl, tcltk, MASS

Description This small collection of functions provides what we call elemental graphics for display of analysis of variance
results, David C. Hoaglin, Frederick Mosteller and John W. Tukey (1991, ISBN:978-0-471-52735-0), Paul R. Rosenbaum (1989)
<doi:10.2307/2684513>, Robert M. Pruzek and James E. Helmreich <https://jse.amstat.org/v17n1/helmreich.html>.
The term elemental derives from the fact that each function is aimed at construction of graphical displays that afford direct visualizations of data with respect to the fundamental questions that drive the particular analysis of variance methods. These functions can be particularly helpful for students and non-statistician analysts. But these methods should be quite generally helpful for work-a-day applications of all kinds, as they can help to identify outliers, clusters or patterns, as well as highlight the role of non-linear transformations of data.

License GPL (>= 2)

LazyLoad yes

LazyData yes

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2023-03-22 11:30:02 UTC

R topics documented:

---

granova-package   *Graphical Analysis of Variance*

Description

This small collection of functions provides what we call elemental graphics for display of anova results. The term elemental derives from the fact that each function is aimed at construction of graphical displays that afford direct visualizations of data with respect to the fundamental questions that drive the particular anova methods. The two main functions are granova.1w (a graphic for one way anova) and granova.2w (a corresponding graphic for two way anova). These functions were written to display data for any number of groups, regardless of their sizes (however, very large data sets or numbers of groups can be problematic). For these two functions a specialized approach is used to construct data-based contrast vectors for which anova data are displayed. The result is that the graphics use straight lines, and when appropriate flat surfaces, to facilitate clear interpretations while being faithful to the standard effect tests in anova. The graphic results are complementary to standard summary tables for these two basic kinds of analysis of variance; numerical summary results of analyses are also provided as side effects. Two additional functions are granova.ds (for comparing two dependent samples), and granova.contr (which provides graphic displays for a priori contrasts). All functions provide relevant numerical results to supplement the graphic displays of anova data. The graphics based on these functions should be especially helpful for learning how the methods have been applied to answer the question(s) posed. This means they can be particularly helpful for students and non-statistician analysts. But these methods should be quite generally helpful for work-a-day applications of all kinds, as they can help to identify outliers, clusters or patterns, as well as highlight the role of non-linear transformations of data. In the case of granova.1w and granova.ds especially, several arguments are provided to facilitate flexibility in the construction of graphics that accommodate diverse features of data, according to their corresponding display requirements. See the help files for individual functions.

Details

Package:	granova
Version:	2.2
License:	GPL (>= 2)

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

James E. Helmreich <James.Helmreich@Marist.edu>

Maintainer: Frederic Bertrand <frederic.bertrand@utt.fr>

See Also

[granova.1w](#) [granova.2w](#) [granova.ds](#) [granova.contr](#)

---

anorexia.sub	*Family Treatment Weight change data for young female anorexia patients.*

Description

The MASS package includes the dataset anorexia, containing pre and post treatment weights for young female anorexia patients. This is a subset of those data, containing only those patients who received Family Treatment.

Usage

```
data(anorexia.sub)
```

Format

A dataframe with 17 observations on the following 2 variables, no NAs.

Prewt Pretreatment weight of subject, in pounds.

Postwt Postreatment weight of subject, in pounds.

Source

Hand, D. J., Daly, F., McConway, K., Lunn, D. and Ostrowski, E. eds (1993) A Handbook of Small Data Sets. Chapman & Hall, Data set 285 (p. 229)

References

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

---

arousal	*Arousal in Rats*

Description

40 rats were given divided randomly into four groups and assigned to one of four treatments: placebo, drug A, drug B, or both drug A and drug B. Response is a standard measure of physiological arousal.

Usage

```
data(arousal)
```

Format

A data frame with 40 observations, 10 in each of 4 columns the corresponding to placebo, drug A, drug B and both drug A and drug B; no NAs.

Placebo Rats receiving a placebo treatment.

Drug.A Rats receiving only drug A.

Drug.B Rats receiving only drug B.

Drug.A.B Rats receiving both drug A and drug B.

Source

Richard Lowry. Concepts & Applications of Inferential Statistics. Vassar College, Poughkeepsie, N.Y., 2010, http://faculty.vassar.edu/lowry/webtext.html

---

blood_lead	*Blood lead levels of lead workers' children matched with similar control children.*

Description

Children of parents who had worked in a factory where lead was used in making batteries were matched by age, exposure to traffic, and neighborhood with children whose parents did not work in lead-related industries. Whole blood was assessed for lead content yielding measurements in mg/dl

Usage

```
data(anorexia.sub)
```

Format

A dataframe with 33 observations on the following 2 variables, no NAs.

Exposed   Blood lead level of exposed child, mg/dl.

Control   Blood lead level of exposed child, mg/dl.

Source

Morton, D., Saah, A., Silberg, S., Owens, W., Roberts, M., Saah, M. (1982). Lead absorption in children of employees in a lead related industry. American Journal of Epidemiology, 115:549-555.

References

See discussion in Section 2.5 of Enhancing Dependent Sample Analyses with Graphics, Journal of Statistics Education Volume 17, Number 1 (March 2009).

---

granova.1w	*Graphic display for one-way ANOVA*

Description

Graphic to display data for a one-way analysis of variance, and also to help understand how ANOVA works, how the F statistic is generated for the data in hand, etc. The graphic may be called 'elemental' or 'natural' because it is built upon the key question that drives one-way ANOVA.

Usage

```
granova.1w(data, group = NULL, dg = 2, h.rng = 1.25, v.rng = 0.2,
    box = FALSE, jj = 1, kx = 1, px = 1, size.line = -2.5,
    top.dot = 0.15, trmean = FALSE, resid = FALSE, dosqrs = TRUE,
    ident = FALSE, pt.lab = NULL, xlab = NULL, ylab = NULL,
    main = NULL, ...)
```

Arguments

data	Dataframe or vector. If a dataframe, the two or more columns are taken to be groups of equal size (whence group is NULL). If data is a vector, group must be a vector, perhaps a factor, that indicates groups (unequal group sizes allowed with this option).
group	Group indicator, generally a factor in case data is a vector.
dg	Numeric; sets number of decimal points in output display, default = 2.
h.rng	Numeric; controls the horizontal spread of groups, default = 1.25
v.rng	Numeric; controls the vertical spread of points, default = 0.25.
box	Logical; provides a bounding box (actually a square) to the graph; default FALSE.

jj	Numeric; sets horizontal jittering level of points; when pairs of ordered means are close to one another, try jj < 1; default = 1.
kx	Numeric; controls relative sizes of cex, default = 1.0
px	Numeric; controls relative sizes of cex.axis, default = 1.0
size.line	Numeric; controls vertical location of group size and name labels, default = -2.5.
top.dot	Numeric; controls hight of end of vertical dotted lines through groups; default = .15.
trmean	Logical; marks 20% trimmed means for each group (as green cross) and prints out those values in output window, default = FALSE.
resid	Logical; displays marginal distribution of residuals (as a 'rug') on right side (wrt grand mean), default = FALSE.
dosqrs	Logical; ensures plot of squares (for variances); when FALSE or the number of groups is 2, squares will be suppressed, default = TRUE.
ident	Logical; allows user to identify specific points on the plot, default = FALSE.
pt.lab	Character vector; allows user to provide labels for points, else the rownames of xdata are used (if defined), or if not labels are 1:N (for N the total number of all data points), default = NULL.
xlab	Character; horizontal axis label, default = NULL.
ylab	Character; vertical axis label, default = NULL.
main	Character; main label, top of graphic; can be supplied by user, default = NULL, which leads to printing of generic title for graphic.
...	Optional arguments to be passed to identify, for example offset

Details

The central idea of the graphic is to use the fact that a one way analysis of variance F statistic is the ratio of two variances each of which can usefully be presented graphically. In particular, the sum of squares between (among) can be represented as the sum of products of so-called effects (each being a group mean minus the grand mean) and the group means; when these effects are themselves plotted against the group means a straight line necessarily ensues. The group means are plotted as (red triangles along this line. Data points (jittered) for groups are displayed (vertical axis) with respect to respective group means. One-way ANOVA residuals can be displayed (set resid=TRUE) as a rug plot (on right margin); the standard deviation of the residuals, when squared, is just the mean square within, which corresponds to area of blue square. The conventional F statistic is just a ratio of the between to the within mean squares, or variances, each of which corresponds to areas of squares in the graphic. The blue square, centered on the grand mean vertically and zero for the X-axis, corresponds to mean square within (with side based on [twice] the pooled standard deviation); the red square corresponds to the mean square between, also centered on the grand mean. Use of effects to locate the groups in the order of the observed means, from left to right (by increasing size) yields this 'elemental' graphic for this commonly used statistical method.

Groups need not be of the same sizes, nor do data need to reflect any particular distributional charac-teristics. Skewness, outliers, clustering of data points, and various other features of the data may be seen in this graphic, possibly identified using point labels. Trimmed means (20%) can also be dis-played if desired. Finally, by redisplaying the response data in two or more versions of the graphic it can be useful to visualize various effects of non-linear data transformations. (ident=TRUE).

Value

Returns a list with two components:

grandsum   Contains the basic ANOVA statistics: the grandmean, the degrees of freedom and mean sums of squares between and within groups, the F statistic, F probability and the ratio between the sum of squares between groups and the total sum of squares.

stats   Contains a table of statistics by group: the size of each group, the contrast coefficients used in plotting the groups, the weighted means, means, and 20% trimmed means, and the group variances and standard deviations.

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>,

James E. Helmreich <James.Helmreich@Marist.edu>

References

Fundamentals of Exploratory Analysis of Variance, Hoaglin D., Mosteller F. and Tukey J. eds., Wiley, 1991.

See Also

granova.2w, granova.contr, granova.ds

Examples

```
data(arousal)
#Drug A
granova.1w(arousal[,1:2], h.rng = 1.6, v.rng = 0.5, top.dot = .35)

#########################

data(anorexia, package="MASS")
wt.gain <- anorexia[, 3] - anorexia[, 2]
granova.1w(wt.gain, group = anorexia[, 1], size.line = -3)

##########################

data(poison)
##Note violation of constant variance across groups in following graphic.
granova.1w(poison$SurvTime, group = poison$Group, ylab = "Survival Time")
##RateSurvTime = SurvTime^-1
granova.1w(poison$RateSurvTime, group = poison$Group,
ylab = "Survival Rate = Inverse of Survival Time")

##Nonparametric version: RateSurvTime ranked and rescaled
##to be comparable to RateSurvTime;
##note labels as well as residual (rug) plot below.
granova.1w(poison$RankRateSurvTime, group = poison$Group,
```

```
ylab = "Ranked and Centered Survival Rates",
main = "One-way ANOVA display, poison data (ignoring 2-way set-up)",
res = TRUE)
```

---

granova.2w   *Graphical display of data for two-way analysis of variance*

Description

Produces a rotatable graphic (controlled by the mouse) to display all data points for any two way analysis of variance.

Usage

```
granova.2w(data, formula = NULL, fit = "linear", ident = FALSE,
        offset = NULL, ...)
```

Arguments

data	An N x 3 dataframe. (If it is a matrix, it will be converted to a dataframe.) Column 1 must contain response values or scores for all groups, N in all; columns 2 and 3 should be factors (or will be coerced to factors) showing levels of the two treatments. If rows are named, then for ident= TRUE, points can be identified with those labels, otherwise the row number of data is used. Note that factor levels will (generally) be reordered.
formula	Optional formula used by [aov](#) to produce the summary 2-way ANOVA table provided as output. Not used in the scatterplot.
fit	Defines whether the fitted surface will be linear (default) or some more complicated surface, e.g., quadratic, or smooth; see below.
ident	Logical, if TRUE allows interactive identification of individual points using rownames of data on graphic. If rownames are not provided then 1:N is used. Click and hold right mouse button while dragging over point. Right click white space to end.
offset	Number; if NULL then default for identify3d is used.
...	Optional arguments to be passed to scatter3d.

Details

The function depicts data points graphically in a window using the row by column set-up for a two-way ANOVA; the graphic is rotatable, controlled by the mouse. Data-based contrasts (cf. description for one-way ANOVA: [granova.1w](#)) are used to ensure a flat surface – corresponding to an additive fit (if fit = linear; see below) – for all cells. Points are displayed vertically (initially) with respect to the fitting surface. In particular, (dark blue) spheres are used to show data points for all groups. The mean for each cell is shown as a white sphere. The graphic is based on rgl and scatter3d; the graphic display can be zoomed in and out by scrolling, where the mouse is used to rotate the entire figure in a 3d representation. The row and column (factor A and B) effects have

been used for spacing of the cells on the margins of the fitting surface. As noted, the first column of the input data frame must be response values (scores); the second and third columns should be integers that identify levels of the A and B factors respectively. Based on the row and column means, factor levels are first ordered (from small to large) separately for the row and column means; levels are assumed not to be ordered at the outset.

The function scatter3d is used from car (thanks, John Fox). The value of fit is passed to scatter3d and determines the surface fit to the data. The default value of fit is linear, so that interactions may be seen as departures of the cell means from a flat surface. It is possible to replace linear with any of quadratic, smooth, or additive; see help for scatter3d for details. Note in particular that a formula specified by the user (or the default) has no direct effect on the graphic, but is reflected in the console output.

For data sets above about 300 or 400 points, the default sphere size (set by sphere.size) can be quite small. The optional argument sphere.size = 2 or a similar value will increase the size of the spheres. However, the sphere sizes possible are discrete.

The table of counts for the cell means is printed (with respect the the reordered rows and columns); similarly, the table of cell means is printed (also, based on reordered rows and columns). Finally, numerical summary results derived from function aov are also printed. Although the function accommodates the case where cell counts are not all the same, or when the data are unbalanced with respect to the A & B factors, the surface can be misleading, especially in highly unbalanced data. Machine memory for this function has caused problems with some larger data sets. The authors would appreciate reports of problems or successes with larger data sets.

Value

Returns a list with four components:

A.effects	Reordered factor A (second column of data) effects (deviations of A-level means from grand mean)
B.effects	Reordered factor B (third column of data) effects (deviations of B-level means from grand mean)
CellCounts.Reordered	
	Cell sizes for all A-level, B-level combinations, with rows/columns reordered according to A.effects and B.effects.
CellMeans.Reordered	
	Means for all cells, i.e., A-level, B-level combinations, with rows/columns reordered according to A.effects and B.effects
anova.summary	Summary aov results, based on input data

Note

Right click on the graphic to terminate identify and return the output from the function.

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

James E. Helmreich <James.Helmreich@Marist.edu>

References

Fundamentals of Exploratory Analysis of Variance, Hoaglin D., Mosteller F. and Tukey J. eds.,
Wiley, 1991.

See Also

granova.1w, granova.contr, granova.ds

Examples

```
# using the R dataset warpbreaks; see documentation
#(first surface flat since fit = 'linear' (default);
#second surface shows curvature)
granova.2w(warpbreaks)
granova.2w(warpbreaks, formula = breaks ~ wool + tension)
granova.2w(warpbreaks, formula = breaks ~ wool + tension,
fit = 'quadratic')

# Randomly generated data
resp <- rnorm(80, 0, .25) + rep(c(0, .2, .4, .6), ea = 20)
f1 <- rep(1:4, ea = 20)
f2 <- rep(rep(1:5, ea = 4), 4)
rdat1 <- cbind(resp, f1, f2)
granova.2w(rdat1)
#
rdat2 <- cbind(rnorm(64, 10, 2), sample(1:4, 64, repl = TRUE),
   sample(1:3, 64, repl = TRUE))
granova.2w(rdat2)
#
#

data(poison)
#Raw Survival Time as outcome measure:
granova.2w(poison[, c(4, 1, 2)])
# Now with quadratic surface (helpful for this poor metric):
granova.2w(poison[, c(4, 1, 2)], fit = 'quadratic')
#
#Inverse of Survival Time as outcome measure
#(actually rate of survival, a better version of response, clearly):
granova.2w(poison[, c(5, 1, 2)])
#Now curvature is minimal (confirming adequacy of
#linear model fit for this metric):
granova.2w(poison[, c(5, 1, 2)], fit = 'quadratic')
#
#Ranked Version of Inverse:
granova.2w(poison[, c(6, 1, 2)])
```

---

granova.contr   *Graphic Display of Contrast Effect of ANOVA*

Description

Provides graphic displays that shows data and effects for a priori contrasts in ANOVA contexts; also corresponding numerical results.

Usage

```
granova.contr(data, contrasts, ylab = "Outcome (response)",
xlab = NULL, jj = 1)
```

Arguments

data	Vector of scores for all equally sized groups, or a data.fame or matrix where each column represents a group.
contrasts	Matrix of column contrasts with dimensions (number of groups [G]) x (number of contrasts) [generally (G x G-1)].
ylab	Character; y axis lable.
xlab	Character vector of length number of contrast columns. To name the specific contrast being made in all but last panel of graphic. Default = NULL
jj	Numeric; controls [jitter](#) and confers the possibility of controlling the amount of jitter in the panel plots for the contrasts Default is 1.

Details

Function provides graphic displays of contrast effects for prespecified contrasts in ANOVA. Data points are displayed as relevant for each contrast based on comparing groups according to the positive and negative contrast coefficients for each contrast on the horizontal axis, against response values on the vertical axis. Data points corresponding to groups not being compared in any contrast (coefficients of zero) are ignored. For each contrast (generally as part of a 2 x 2 panel) a line segment is given that compares the (weighted) mean of the response variable for the negative coefficients versus the positive coefficients. Standardized contrasts are used, wherein the sum of (magnitudes) of negative coefficients is unity; and the same for positive coefficients. If a line is 'notably' different from horizontal (i.e. slope of zero), a 'notable' effect has been identified; however, the question of statistical significance generally depends on a sound context-based estimate of standard error for the corresponding effect. This means that while summary aov numerical results and test statistics are presented (see below), the appropriateness of the default standard error generally requires the analyst's judgment. The response values are to be input in (a stacked) form, i.e. as a vector, for all cells (cf. arg. ylab). The matrix of contrast vectors contrasts must have G rows (the number of groups), and a number of columns equal to the number of prespecified contrasts, at most G-1. If the number of columns of contrasts is G-1, then the number per group, or cell size, is taken to be length(data)/G, where G = nrow(contrasts).

If the number of columns of contrasts is less than G-1 then the user must stipulate npg, the number in each group or cell. The function is designed for the case when all cell sizes are the same, and may

be most helpful when the a priori contrasts are mutually orthogonal (e.g., in power of 2 designs, or their fractional counterparts; also when specific row or column comparisons, or their interactions (see the example below based on rat weight gain data)). It is not essential that contrasts be mutually orthogonal; but mutual linear independence is required. (When factor levels correspond to some underlying continuum a standard application might use con = contr.poly(G), for G the number of groups; consider also contr.helmert(G).) The final plot in each application shows the data for all groups or cells in the design, where groups are simply numbered from 1:G, for G the number of groups, on the horizontal axis, versus the response values on the vertical axis.

Value

Two sets of numerical results are presented: Weighted cell means for positive and negative coefficients for each a priori contrast, and summary results from lm.

summary.lm   Summary results for a linear model analysis based on the R function lm (When effects are simple, as in an equal n's power of 2 design, mean differences will generally correspond to the linear regression coefficients as seen in the lm summary results.)

means.pos.neg.coeff

          table showing the (weighted) means for positive and negative coefficients for each (row) contrast, and for each row, the difference between these means in the final column

means.pos.neg.coeff

          Table showing the (weighted) means for positive and negative coefficients for each (row) contrast, and for each row, the difference between these means, and the standardized effect size in the final column.

contrasts   Contrast matrix used.

group.means.sds

          Group means and standard deviations.

data   Input data in matrix form.

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

James E. Helmreich <James.Helmreich@Marist.edu>

See Also

granova.1w, granova.2w, granova.ds

Examples

```
data(arousal)
contrasts22 <- data.frame( c(-.5,-.5,.5,.5),
c(-.5,.5,-.5,.5), c(.5,-.5,-.5,.5) )
names(contrasts22) <- c("Drug.A", "Drug.B", "Drug.A.B")
granova.contr(arousal, contrasts = contrasts22)

data(rat)
```

```
dat6 <- matrix(c(1, 1, 1, -1, -1, -1, -1, 1, 0, -1, 1, 0, 1, 1, -2,
    1, 1, -2, -1, 1, 0, 1, -1, 0, 1, 1, -2, -1, -1, 2), ncol = 5)
granova.contr(rat[,1], contrasts = dat6, ylab = "Rat Weight Gain",
  xlab = c("Amount 1 vs. Amount 2", "Type 1 vs. Type 2",
  "Type 1 & 2 vs Type 3", "Interaction of Amount and Type 1 & 2",
  "Interaction of Amount and  Type (1, 2), 3"))
#Polynomial Contrasts
granova.contr(rat[,1],contrasts = contr.poly(6))

#based on random data
data.random <- rt(64, 5)
granova.contr(data.random, contrasts = contr.helmert(8),
ylab = "Random Data")
```

---

granova.ds	*Granova for Display of Dependent Sample Data*

Description

Plots dependent sample data beginning from a scatterplot for the X,Y pairs; proceeds to display difference scores as point projections; also X and Y means, as well as the mean of the difference scores. Also prints various summary statistics including: effect size, means for X and Y, a 95% confidence interval for the mean difference as well as the t-statistic and degrees of freedom.

Usage

```
granova.ds(data, revc = FALSE, sw = 0.4, ne = 0.5, ptpch=c(19,3),
        ptcex=c(.8,1.4), labcex = 1, ident = FALSE,
        colors = c(1,2,1,4,2,'green3'), pt.lab = NULL,
        xlab = NULL, ylab = NULL, main = NULL, sub = NULL,
        par.orig = TRUE)
```

Arguments

data	is an n X 2 dataframe or matrix. First column defines X (intially for horzontal axis), the second defines Y.
revc	reverses X,Y specifications.
sw	extends axes toward lower left, effectively moving data points to the southwest.
ne	extends axes toward upper right, effectively moving data points to northeast. Making both sw and ne smaller moves points farther apart, while making both larger moves data points closer together.
ptpch	controls the pch of the (X,Y) points and of differences score points.
ptcex	controls the cex of the (X,Y) points and of differences score points.
labcex	controls size of axes labels.
ident	logical, default FALSE. Allows user to identify individual points.

colors	vector defining colors of six components of the plot: (X,Y) points, horizontal and vertical dashed lines representing means of the two groups, light dashed diagonal lines connecting (X,Y) points and projections differences dotplot, differences arranged as a dotplot, heavy dashed diagonal line representing the mean of differences, confidence interval.
pt.lab	optional character vector defining labels for points. Only used if ident is TRUE. If NULL, rownames(data) are used if available; if not 1:n is used.
xlab	optional label (as character) for horizontal axis. If not defined, axis labels are taken from colnames of data.
ylab	optional label (as character) for vertical axis.
main	optional main title (as character); if not supplied by user generic title is provided.
sub	optional subtile (as character).
par.orig	returns par to original settings; if multipanel plots it is advisable to specify FALSE.

Details

Paired X & Y values are plotted as scatterplot. The identity reference line (for Y=X) is drawn. Since the better data view often entails having X's > Y's the revc argument facilitates reversal of the X, Y specifications. Parallel projections of data points to (a lower-left) line segment show how each point relates to its X-Y = D difference; blue 'crosses' are used to display the distribution of difference scores and the mean difference is displayed as a heavy dashed (red) line, parallel to the identity reference line. Means for X and Y are also plotted (as thin dashed vertical and horizontal lines), and rug plots are shown for the distributions of X (at the top of graphic) and Y (on the right side). Several summary statistics are plotted as well, to facilitate both description and inference; see below. The 95% confidence interval for the population mean difference is also shown graphically. Because all data points are plotted relative to the identity line, and summary results are shown graphically, clusters, data trends, outliers, and possible uses of transformations are readily seen, possibly to be accommodated.

Value

A list is returned with the following components:

mean(X)	Mean of X values
mean(Y)	Mean of Y values
mean(D=X-Y)	Mean of differences D = X - Y
SD(D)	Standard deviation of differences D
ES(D)	Effect Size for differences D: mean(D)/SD(D)
r(X,Y)	Correlation based on X,Y pairs
r(x+y,D)	Correlation based on X+Y,D pairs
LL 95%CI	Lower bound for 95% confidence interval for population mean(D)
UL 95%CI	Upper bound for 95% confidence interval for population mean(D)
t(D-bar)	t-statistic associated w/ test of hypothesis that population mean(D) = 0.0
df.t	Degrees of freedom for the t-statistic
pval.t	P-value for two sided t-test of null hypothesis that population mean(D) does not equal zero.

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

James E. Helmreich <James.Helmreich@Marist.edu>

References

Exploratory Plots for Paired Data, Rosenbaum P., The American Statistician, May 1989, vol. 43, no. 2, pp. 108-9.

Enhancing Dependent Sample Analyses with Graphics, Pruzek, R. and Helmreich, J., Journal of Statistics Education, March 2009, Vol. 17, no. 1.

http://www.amstat.org/publications/jse/v17n1/helmreich.pdf

Examples

```
### See discussion of anorexia graphic in EDSAG, J. Statistics Ed.
data(anorexia.sub)

granova.ds(anorexia.sub, revc = TRUE,
main = "Assessment Plot for weights to assess Family Therapy treatment
for Anorexia Patients")
# If labels for four unusual points at lower left are desired:
granova.ds(anorexia.sub, revc = TRUE,
main = "Assessment Plot for weights to assess Family Therapy treatment
for Anorexia Patients", ident = TRUE)


## See discussion of blood lead graphic in EDSAG, J. Statistics Ed.
data(blood_lead)

granova.ds(blood_lead, sw = .1,
   main = "Dependent Sample Assessment Plot
   Blood Lead Levels of Matched Pairs of Children")
```

---

poison	*Poison data from Biological Experiment*

Description

Survial times of animals in a 3 x 4 factorial experiment involving poisons (3 levels) and various treatments (four levels), as described in Chapter 8 of Box, Hunter and Hunter.

Usage

```
data(poison)
```

Format

This data frame was originally `poison.data` from the package BHH2, but as presented here has added columns; no NAs.

`Poison` Factor with three levels I, II, and III.

`Treatment` Factor with four levels, A, B, C, and D.

`Group` Factor with 12 levels, 1:12.

`SurvTime` Numeric; survival time.

`RateSurvTime` Numeric; inverse of SurvTime

`RankRateSurvTime` Numeric; `RateSurvTime` scores have been converted to ranks, and then rescaled to have the same median as and a spread comparable to `RateSurvTime`

Source

Box, G. E. P. and D. R. Cox, An Analysis of Transformations (with discussion), Journal of the Royal Statistical Society, Series B, Vol. 26, No. 2, pp. 211 - 254.

References

Box G. E. P, Hunter, J. S. and Hunter, W. C. (2005). Statistics for Experimenters II. New York: Wiley.

---

rat   *Weight gains of rats fed different diets*

Description

60 rats were fed varying diets to see which produced the greatest weight gain. Two diet factors were protein type: beef, pork, chicken and protein level: high and low.

Usage

```
data(rat)
```

Format

A data frame with 60 observations on the following 3 variables, no NAs.

`Weight.Gain` Weight gain (grams) of rats fed the diets.

`Diet.Amount` Amount of protein in diet: 1 = High, 2 = Low.

`Diet.Type` Type of protein in diet: 1 = Beef, 2 = Pork, 3 = Cereal.

Source

Fundamentals of Exploratory Analysis of Variance, Hoaglin D., Mosteller F. and Tukey J. eds., Wiley, 1991, p. 100; originally from Statistical Methods, 7th ed, Snedecor G. and Cochran W. (1980), Iowa State Press.

Index