

Package ‘granova’

July 29, 2010

Version 1.4

Date 2010-07-27

Title Graphical Analysis of Variance

Author Robert M. Pruzeck <RMPruzeck@yahoo.com> and James E. Helmreich
<James.Helmreich@Marist.edu>

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

Depends R (>= 2.5.0), car

Suggests mgcv, rgl, tcltk, DAAG, MASS

Description This small collection of functions provides distinctive graphics for display of anova results. The two principal functions are `granova.lw` (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 are likely to be problematic). For these two functions a specialized approach is used to construct data-based contrast vectors with respect to which anova data are displayed. The result is that the graphics use straight lines, and flat surfaces, to help ensure 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. Two additional functions are `granova.ds` (for comparing two dependent samples of data), 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 are aimed at students and non-statistician analysts; but they can be generally helpful for identifying outliers, clusters, trends or the role of non-linear transformations of data. In the case of `granova.lw` and `granova.ds` especially, several arguments have been provided to facilitate construction of graphics that accommodate diverse features of data, and their corresponding display requirements.

License GPL (>= 2)

Repository CRAN

Date/Publication 2010-07-29 09:13:48

R topics documented:

granova-package	2
granova.lw	3
granova.2w	5
granova.contr	8
granova.ds	10
rat	12
Index	13

granova-package	<i>Graphical Analysis of Variance</i>
-----------------	---------------------------------------

Description

This small collection of functions provides distinctive graphics for display of anova results. The two principal functions are granova.lw (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 with respect to which anova data are displayed. The result is that the graphics use straight lines, and (when possible) flat surfaces, to help ensure 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. Two additional functions are granova.ds (for comparing two dependent samples of data), 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 are should be especially helpful for learning how the various methods have used the data to answer the question(s) posed by the various methods, so they can be particularly helpful for students and non-statistician analysts. The graphics can be generally helpful for identifying outliers, clusters, trends or the role of non-linear transformations of data. In the case of granova.lw and granova.ds especially, several arguments are provided to facilitate construction of graphics that accommodate diverse features of data, and their corresponding display requirements. See the help files for individual functions.

Details

Package: granova
Version: 1.3
License: GPL (>= 2)

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

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

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

See Also

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

granova.lw

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 computed, etc. The graphic may be called 'elemental' or 'natural' because it is built around the key question that drives one-way ANOVA.

Usage

```
granova.lw(yy, 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

yy	Dataframe or vector. If a dataframe, the two or more columns are taken to be groups of equal size (whence 'group' is NULL). If yy 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 yy 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 height 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.

<code>resid</code>	Logical; displays marginal distribution of residuals (as a 'rug') on right side (wrt grand mean), default = FALSE.
<code>dosqrs</code>	Logical; ensures inclusion of the squares (for variances); when FALSE and the number of groups = 2, squares will be suppressed, default = TRUE.
<code>ident</code>	Logical; allows user to identify specific points on the plot, default = FALSE.
<code>pt.lab</code>	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.
<code>xlab</code>	Character; horizontal axis label, default = NULL.
<code>ylab</code>	Character; vertical axis label, default = NULL.
<code>main</code>	Character; main label, top of graphic; can be supplied by user, default = NULL, which leads to printing of a generic title for graphic.
<code>...</code>	Optional arguments to be passed to <code>identify</code> , for example <code>offset</code>

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. Additionally, data points are displayed (re: vertical axis) with respect to these group means for each group. Once this elemental plot has been generated the one-way ANOVA residuals can be displayed as a rug plot (on the right margin), whence it is recognized that the pooled standard deviation of the residuals, when squared, is just the mean square within. The conventional F statistic is just a ratio of the between to the within mean squares, or variances, and it is straightforward to represent each variance as the area of a square in the graphic. Specifically, the blue square, centered on the grand mean vertically and zero for the X-axis, corresponds to the mean square within (with side based on [twice] the pooled standard deviation); the second square corresponds to the mean square between, and is also centered on the grand mean. The ratio of the areas of these squares is the F-statistic. By using effects to locate the groups in the order of the observed means, from left to right for smallest to largest, an 'elemental' graphic is defined for this commonly used statistical method.

Groups need not be of the same sizes, nor do data need to reflect any particular distributional characteristics. But skewness, outliers, clustering of scores, and various other features of the data can often be readily seen in this framework. Trimmed means (20%) can also be displayed if desired. Finally, by redisplaying the response data in two or more versions of the graphic it will often be possible to visualize the effects of transformations. Finally, individual points can be labeled in the graphic (`ident=TRUE`).

Value

Returns a list with two components:

<code>grandsum</code>	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

```
library(MASS)
wt.gain <- anorexia[,3] - anorexia[,2]
granova.lw(wt.gain, group = anorexia[,1], size.line = -3)
#
attach(ChickWeight)
CW0 <- subset(ChickWeight, Time==0)[,c(1,4)]
granova.lw(yy = CW0[,1], group = CW0[,2], jj = 2)
#
CW21 <- subset(ChickWeight, Time==21)[,c(1,4)]
granova.lw(yy = CW21[,1], group = CW21[,2])
```

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(formula = NULL, data.A.B, fit = "linear", ident = FALSE,
           offset = NULL, ...)
```

Arguments

<code>data.A.B</code>	An N X 3 dataframe. (If it is a matrix, it will be converted to a df.) Column 1 must contain response values or scores for all groups, N in all; columns 2 and three must be vectors with integers showing levels of factors A and B, respectively. If rows are named uniquely, then for <code>ident= TRUE</code> , points can be identified with those labels, otherwise the row number of <code>data.A.B</code> is used. Note that factor levels will (generally) be reordered.
<code>formula</code>	Optional formula used by <code>ao</code> to produce the summary 2-way ANOVA table provided as output. Not used in the scatterplot.
<code>fit</code>	Defines whether the fitted surface will be 'linear' (default) or some more complicated surface, e.g., quadratic, or smooth; see below.
<code>ident</code>	Logical, if <code>TRUE</code> allows interactive identification of individual points using row-names of <code>data.A.B</code> on graphic. If rownames are not provided then 1:N is used.
<code>offset</code>	Number; if <code>NULL</code> then default for <code>identify3d</code> is used.
<code>...</code>	Optional arguments to be passed to <code>scatter3d</code> .

Details

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.lw](#)) 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 where 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. Function `scatter3d` is used from `Rcmdr` (thanks, John Fox). The `fit` is defaulted to `linear` whence interactions are depicted 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. The table of counts for the cell means is printed (with respect to 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 `ao` 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:

<code>A.effects</code>	Reordered factor A effects (deviations of A-level means from grand mean)
------------------------	--

`B.effects` Reordered factor B 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.lw](#), [granova.contr](#), [granova.ds](#)

Examples

```
#Random 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(data.A.B = rdat1)
#
rdat2 <- cbind(rnorm(64,10,2), sample(1:4, 64, repl = TRUE),
              sample(1:3, 64, repl = TRUE))
granova.2w(data.A.B = rdat2)
#
granova.2w(formula = breaks ~ wool * tension, data.A.B = warpbreaks)
```

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(resp, con, ngrp = nrow(con),
              npg = length(resp)/ngrp, jj = 1)
```

Arguments

<code>resp</code>	Vector of scores for all equally sized groups.
<code>con</code>	Matrix of column contrasts with dimensions (number of groups [G]) x (number of contrasts) [generally G X G-1].
<code>ngrp</code>	Numeric; default = <code>nrow(resp)</code> . The number of rows = number ‘cells’ or groups [G].
<code>npg</code>	Numeric; default is <code>length(resp)/ngrp</code> . If the number of contrasts is NOT equal to the number of groups less one (i.e., to df ‘among’), then <code>npg</code> must be set equal to the number in each group (presumed equal for all groups). The default <code>npg</code> setting assumes <code>con</code> has one fewer column than the number of rows.
<code>jj</code>	Numeric; controls <code>jitter</code> 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 so-called a priori 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. resp`). The matrix of contrast vectors `con` should have G rows (the number of groups), and a number of columns equal to the number of a priori contrasts, at most G-1. If the number of columns of `con` is G-1, then the number per group, or cell size, is taken to be `length(resp)/G`, where `G = nrow(con)`.

If the number of columns of `con` 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`.

<code>summary.lm</code>	Summary results for a linear model analysis based on the R function <code>lm</code> (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 <code>lm</code> summary results.)
<code>means.pos.neg.coef</code>	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
<code>means.pos.neg.coef</code>	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.
<code>contrasts</code>	Contrast matrix used.
<code>group.means.sds</code>	Group means and standard deviations.
<code>data</code>	Input data in matrix form.

Author(s)

Robert M. Pruzek <RMPruzek@yahoo.com>

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

See Also

[granova.lw](#), [granova.2w](#), [granova.ds](#)

Examples

```
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], con = dat6)
```

```
#Polynomial Contrasts
granova.contr(rat[,1],con = contr.poly(6))

#based on random data (not run)
resp <- rt(64, 5)
granova.contr(resp, con = contr.helmert(8))
```

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(xdata, revc = FALSE, sw = 0.4, ne = 0.5, ptpch=c(19,3),
  ptcex=c(1,1.4), 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

xdata	is an n X 2 dataframe. First column defines X (initially for horizontal 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(xdata) 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 xdata.
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 subtitle (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. 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. 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.

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: $\text{mean}(D)/\text{SD}(D)$
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.

Examples

```
#Data analysis and graphics, by John McDonald
library(DAAG)
#length of stretched rubber bands; scores show length
#after being heated (X), as well as at ambient temperature (Y).
data(pair65)
granova.ds(pair65,
  main = "Dependent sample assessment plot for pair65 data, n = 9")
```

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

*Topic **datasets**

rat, [12](#)

*Topic **hplot**

granova-package, [2](#)

granova.lw, [3](#)

granova.2w, [5](#)

granova.contr, [7](#)

granova.ds, [9](#)

*Topic **htest**

granova.lw, [3](#)

granova.ds, [9](#)

aov, [6](#)

granova (*granova-package*), [2](#)

granova-package, [2](#)

granova.lw, [2](#), [3](#), [6](#), [7](#), [9](#)

granova.2w, [2](#), [5](#), [5](#), [9](#)

granova.contr, [2](#), [5](#), [7](#), [7](#)

granova.ds, [2](#), [5](#), [7](#), [9](#), [9](#)

jitter, [8](#)

rat, [12](#)