

## A new graphic for one-way ANOVA

Bob Pruzek, SUNY Albany

This document describes and illustrates a new 'elemental' graphic for one-way analysis of variance, *i.e.*, ANOVA. The primary motivation for developing the central function was to facilitate a deeper understanding of the key features of analysis of variance by focusing on the central question of the method in the context of using modern graphics that can facilitate sound data analyses. It is also hoped that use of this function will facilitate development of modern data-analytic thinking and skills in ANOVA applications.

The key (omnibus)  $F$  statistic at the heart of any *inferential* application of the simplest ANOVA model implies a particular way to compare means, one based on ***data-based contrasts***. Indeed, there appears to be one distinctive approach to comparing several groups of quantitative data that is wholly consistent with the central question that drives (one-way) ANOVA, and that leads directly to an ***elemental graphic*** for this method. The method used permits visualization of all data points for any number of groups where all group means necessarily lie on a straight line for any data system. (It is also straightforward to generalize the display to accommodate rows and columns in any two-way ANOVA as well, although that will not be done here; see the Note on page eight below.) Additionally, this one-way ANOVA graphic can ordinarily be used to visualize residuals, basic 'effects,' as well as all data points, remaining faithful to the central question that drives the method. A key feature of the graphic is that it facilitates ***visualization of the central variance estimates***, *i.e.*, the Mean Squares Between and Within, by displaying certain ***squares***, the sides of which are based on standard deviation units. It follows that the conventional  $F$  statistic can be ***seen as a ratio of the areas*** of these squares.

It is anticipated that students using this function will better understand the basic principles of analysis of variance, and applied researchers will be able to better understand their data. In addition, as discussed briefly below, the function is easily used in to study the effects of various changes in one's data, or to visualize the results of simulations, or the effects of repeated sampling (as in bootstrapping). Also, since any data system admits to alternative transformations, or re-expressions, the graphic can facilitate a better understanding of how choice of transformation effects not just means and variances and other summary statistics, but can help qualify or support inferences, or see the role of individual data points in their respective groups.

The initial illustration of the function uses data from a One Way ANOVA website: <http://espse.ed.psu.edu/statistics/Chapters/Chapter11/Chap11.html#A%20Statlets%20Problem> whence the new graphic function (`gr.owaov`) is used.

Data A: 26 25 29 21 20 18 ← each **row** is a treatment group,  $n = 6$  each  
B: 24 17 16 13 21 19 [NB: input of these data readily done  
C: 32 34 29 19 27 28 if input '**yy**' is *transpose* of this matrix]  
D: 23 29 26 20 24 25

The data source (the website) states: “The data [were] collected by a research group investigating the whitening power of four new toothpaste formulas. The dependent variable ... (Dep\_Var) is a whiteness measure where the lower the number, the whiter the teeth. The independent variable Group codes the four different toothpaste formulas using letters” A - D. (It is not known whether the data are real or not.)

The function, which works with R or Splus, first prints a number of standard numerical results and then generates the graphic. In its simplest forms, with all but one logical argument set to FALSE, the function provides only the essentials of ANOVA, each group of score values being centered on its mean. (Details about the arguments for the function are provided below.) Given that  $M(j)$  represents the arithmetic mean of the  $j$ th group, and  $M$ , no subscript, represents the grand mean the **contrast coefficients** are constructed, each of form  $c(j) = M(j) - M$ , ordered from the smallest to the largest mean, say, from  $j = 1 \dots J$ . Such contrasts are generally called the ‘effects’ in one-way ANOVA, and the statistical (F) test for this method simply examines whether these effects cum contrasts are large enough in magnitude, relative to variation within groups, to conclude that their population counterparts are not all zero. The graphic uses the fact that when all  $n$ ’s are the same, the Sum of Squares Between in the numerator of the MS(Between) of the  $F$ -statistic, has the form  $n \sum (M(j) - M)^2 = n \sum (M(j) - M) M(j)$ ; so the latter can be written as  $n \sum c(j) M(j)$  using contrast coefficients. The same idea works when  $n$ ’s differ from group to group, except that SS(Between) must then incorporate different  $n$ ’s; viz.,  $SS(\text{Between}) = \sum n(j) c(j) M(j)$  so that contrast coefficients generally reflect differences in the sample sizes as well as differences among means. In any case, SS(Between) can always be constructed as a **data-based contrast** of group means. Of course the MS(Within) is just the average of the respective subgroup variances if  $n$ ’s are equal, a weighted average of group variances if the  $n$ ’s differ across groups.

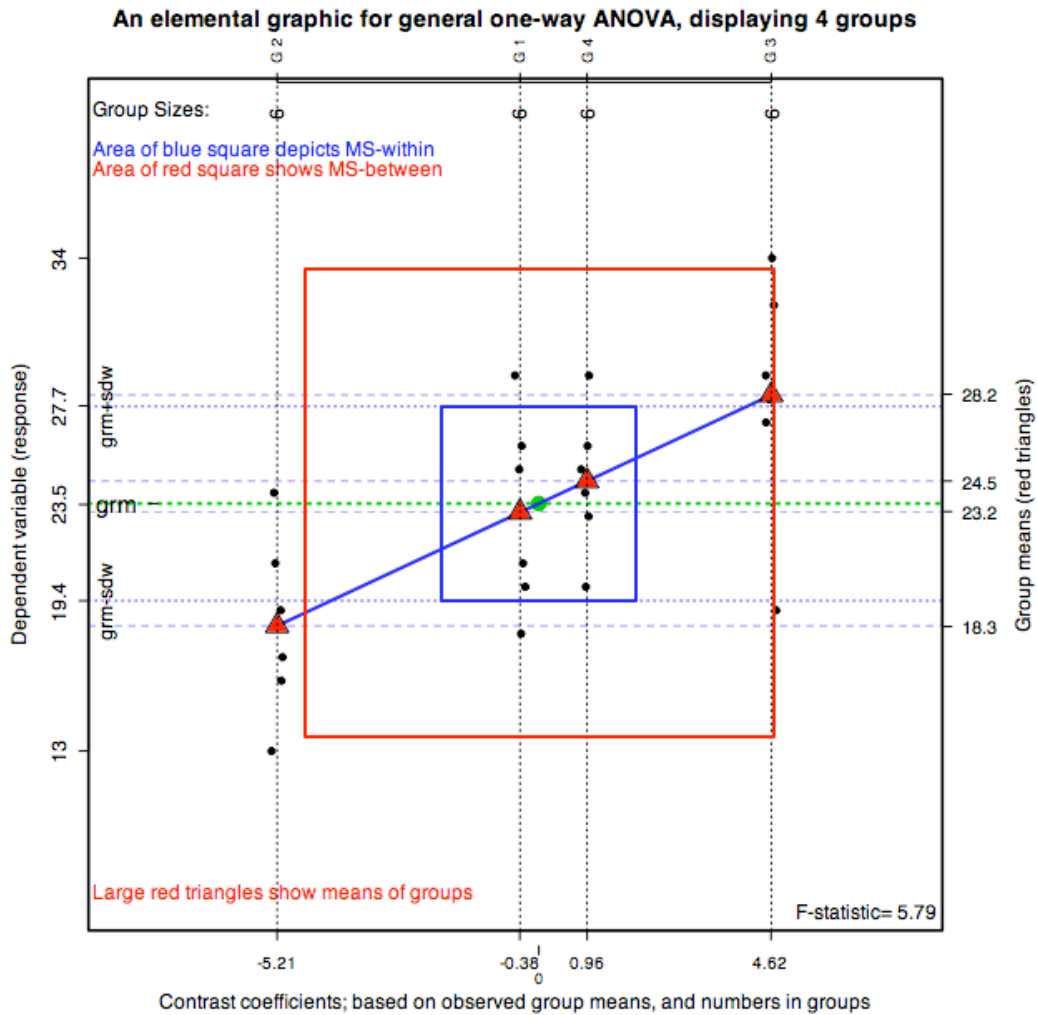
By organizing the groups according to the sizes of the contrasts, here ordered from smallest to largest, and plotting all scores, *i.e.*, the  $y(i,j)$ , vertically, and using a special symbol (a red triangle) to denote group means, these symbols necessarily fall on a straight line because the intervals between the means vary in a way that reflects their differences. The two sets of numerical outputs provided are shown in Table 1.

**Table 1 Summary statistics for an illustrative application of one-way ANOVA**

\$mvys					\$gsummary				
<u>Grp.means</u>	<u>Contr.coefs</u>	<u>Grp.variance</u>	<u>Grp.stdevs</u>		<u>Grand.Mn</u>	<u>df.betw</u>	<u>df.with</u>		
[2,]	18.3	-5.21	15.1	3.88	23.54	3	2		
[1,]	23.2	-0.38	17.4	4.17	<u>MS.Betw</u>	<u>MS.within</u>	<u>F.stat</u>	<u>prob.F</u>	
[4,]	24.5	0.96	9.1	3.02	99.15	17.125	5.79	0.005	
[3,]	28.2	4.62	27.0	5.19	<b>(NB: Focus first on numerical information, what it alone provides, then on how the graphic presentation below complements &amp; extends understanding.)</b>				
rows ordered by means,									
Grp2, Grp1, Grp4, Grp3									
<b>(see labels across top of graphic below)</b>									

The graphic that follows illustrates how this works in the case of a particularly simple set of data, in this case one with four groups, each of size six, for the toothpaste data. The function is called as: `gr.owaov(teeth.whtns.dt)`.

**Figure 1** A graphic that illustrates use of function `gr.owaov`



This figure depicts a scatterplot of *contrast coefficients* (col. two of **mvys** above) for the ordered groups, smallest means to largest, versus the sets of scores for the respective groups, for the horizontal and vertical axes respectively. Each contrast coefficient attaches to a single score, according to that score's group so there are  $n(j)$  contrast coefficients for each group. The contrast coefficients have been jittered to help distinguish points (see below) and the **means** of groups are signified by (red) triangles within each of the score-sets; also the values of these means are printed on the right margin. The plot of the means necessarily depicts a **straight line** because the means are essentially being plotted against themselves, except that the grand mean is initially subtracted from the first set of means. The grand mean [here, 23.54] is coupled

with the contrast coefficient of zero and can be visualized as the **green** point at the center of the plot.

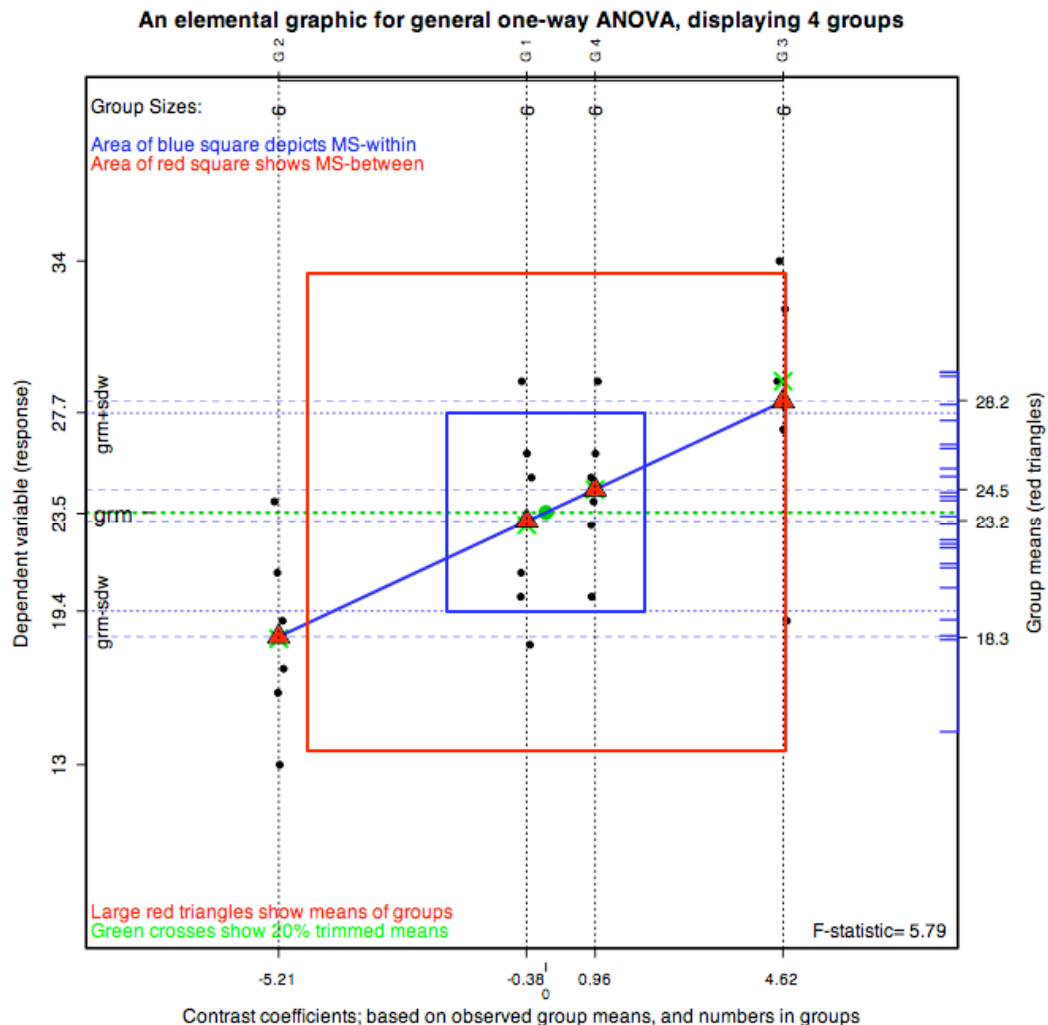
The '**concentric**' squares in the center of the plot correspond to the **within group estimate of variance (blue)** [MS(Within)] & the **between group (red) variance estimate** [MS(Between)]; each side (of each square) corresponds to **two** standard deviation units so that **their areas correspond to variances**. The left side of the figure shows numerical values of edges for the MS(Within). The ratio of the areas of the two squares, **the red area divided by the blue**, is just the  $F$ -statistic. In this case, the area of the **red square** is 5.79 times larger than area of **blue square**, so **it is seen** that  $F = 5.79$ . Study of the two squares, preferably for several examples, is likely to help you see, perhaps for the first time, how variances (*i.e.*, mean squares) can be represented visually; and take note that when means are homogeneous enough, the red square will be smaller than the blue in which case the  $F$  statistic will be less than unity (one). The range of scores is given on left/vertical axis; group means, which correspond to the red triangles in the graphic, are printed at the right margin. Here, the four means spread past two s.d.(within) units, *implying 'notable statistical effects,' which follows (loosely) from seeing that the red square is substantially larger than the blue.*

Figure 1a, below, shows the same basic graphic, but in this case a *rug plot* is shown along the inside right margin to show the distribution of residuals *centered on grand mean*. In addition, the green crosses show trimmed means, these generally being sound robust replacements for the arithmetic mean (that is clearly not a robust estimate of location). Note that the trimmed means tend to be near their non-trimmed counterparts, except for the case of the third treatment, treatment C. In the latter case a low-outlier appears, which is the reason the 20% trimmed mean (see the green cross for the 'G-3', label at top) is discernably larger than the corresponding non-trimmed mean.

As for an overall summary of what these data have to say, as seen in the graphic, the second treatment, *i.e.*, group B, appears to have had the 'best' effect (since smaller scores mean better results according to the author); furthermore, treatments A & D are not particularly competitive with B, and treatment C is notably less effective than the others. (These statements are based on the tacit but fundamental assumption that the 'material' had been randomly assigned to treatment groups so that the entities were more or less comparable from group to group before the introduction of treatments.) As noted, there seems to be (only) one anomalous data point, this for treatment C, but these data were chosen for this initial illustration to keep things relatively simple. Note that the variances or standard deviations are not too dissimilar, except that treatment has C yielded more variable scores. (Remember that formal inferential application of ANOVA entails an assumption of equal variances, as well as normality, in the so-called treatment populations.) Finally, given that the  $F$ -statistic is associated with a small  $p$ -value (see box on p. 2 above), the evidence is strong (assuming random assignments) that these treatments really do 'cause' distinctively different effects (in some putative system of populations), *i.e.*, the formal

inferential test implies that the population means corresponding to these four samples are different from one another.

**Figure 1a** A counterpart of Figure 1 that adds residuals and trimmed means to the plot



The new graphic is somewhat more informative in that residual scores are now shown in the so-called 'rug' on the right side (but where instead of centering on zero, I have centered on the grand mean). It seems clearer now that the lowest point in the right-most group (G3, see top) stands out more clearly as an outlier; furthermore, this graphic shows that the trimmed mean is larger than its untrimmed counterpart, as seen in the green cross for this group residing above the red triangle for that group. That the other trimmed means are more or less the same as their untrimmed counterparts is also clear from this graphic.

Two final points: the argument `dosqrs`, defaulted as `T`, can be set at `FALSE` in which case the entire figure becomes somewhat less busy; this also affords the option of teaching about what ANOVA shows without reference to inference, this being the sole purpose of the F statistic, or the 'concentric' squares. Furthermore, it is

recognized that comparing only two (independent) groups of scores, is most often done in the framework of a two group t-test; thus, a t-statistic replaces F statistic in the lower right corner and the standardized effect size is also printed as part of the numerical output of running the function.

### **Suggestions for repeated use of this graphic function to gain experience learning about the ANOVA method**

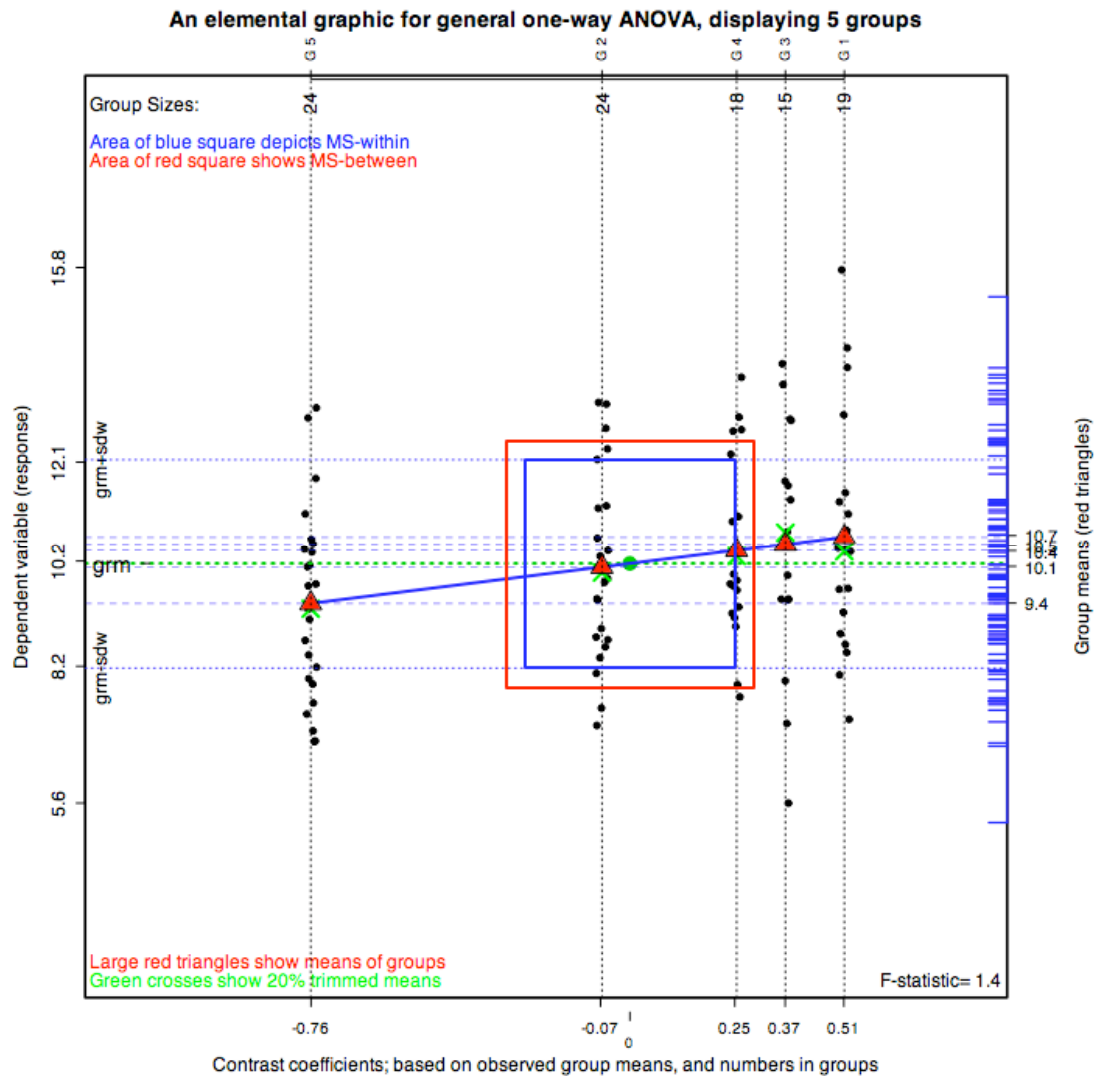
Since R software makes it so easy to simulate data, it is straightforward to use this function to visualize simulated data (and possibly to compare various re-expressed or sampled versions of the data using this graphic or others). For one-way ANOVA, a simulation might proceed initially by sampling so that approximate sample normality is a realistic expectation. For example, suppose we wish to generate data from a single normal population, say with a mean of 10 and a standard deviation of 2, and further, suppose the simulated data points are to be randomly assigned to five groups of varying sizes. This is easily done using the following command, as a type of applet where a total of 100 data points have been distributed randomly across five groups (using functions **rnorm** & **sample**):

```
>gr.owaov(yy = rnorm(n=100,mean=10, sd=2),  
          gp = sample(x=1:5,size=100,repl=T), p=T, m=T)
```

This yielded the graphic shown in Figure 2 in its first trial on my computer. [→Try this on your machine & you will get something similar, but with different numbers, groups, etc.!]

Another approach to construction of applets might replace the function **rnorm** above with **rt**, to sample randomly from a *t*-distribution with possibly small degrees of freedom. (The standard assumption used in the mathematical derivation of the ANOVA test function entails the assumption of population normality, whereas applications regularly lead to samples with longer-than-normal tails, so there may be good reasons in some contexts to study effects of various longer-than-normal tailed distributions on results when ANOVA is used.) This can be done using the same format as seen above, except that **rnorm( )** could be replaced with, say, **rt(n=100, df=3) + 10**, where the choices of 100, 3 and 10 are optional. If samples were to be held to the same size, the second argument for **gp** could be constructed as **rep(1:#gps,ea=n)**, in obvious notation. A further variation on this theme might entail use of ranks in place of the initial scores in the **yy** vector; for example, we define **yy = rank(rt(100,3))** for the first of the two arguments in the basic function; this would provide what is known in the basic literature of statistics as the Friedman version of non-parametric one-way ANOVA. Visual comparisons of rank-transformed data (especially for groups) with its unranked counterpart can be especially revealing.

Figure 2 Randomly simulated data for ANOVA with five groups of varying sizes



Another method of interest entails bootstrapping. The function `samp.mat.boot`, given below (with only two lines of code) is easily employed when the input data `yy` take the form of a matrix (and `gp` is left unspecified). Data (e.g. `teeth.whitns.dt` above) would be in the form of a matrix as in: `func..(samp.mat.boot(data-matrix))` so that 'new' data to be analyzed and plotted generally consists of bootstrap samples from respective columns (treatment groups) of the input matrix. How many groups to compare, of what sizes, possibly equal or varying, and what distributional specifications to make are among the choices that one must make, but applications based on S-language software are straightforward to make, as well as easily tailored to particular needs in defined applications in the context of **R**. It is a special virtue that the **R** software is freely downloadable from the internet [go to [www.r-project.org](http://www.r-project.org), click on CRAN,...].

In applied statistical practice, as distinguished from simulation studies, there is the possibility that covariate information will be available for individual entities. In such cases, it may be useful or revealing to employ a variety of colors, shapes, sizes, etc. to characterize individual scores, here seen as small black dots. Or individual i.d.s can be printed within the graphic. Once one decides to visualize data points, any information thought likely to be informative in the context of the application is of potential interest in such graphics. In observational studies, as contrasted with experiments, covariate information may be usefully summarized using propensity scores that in turn can be used in constructing graphics of this form.

There is great merit in the teachings of the perhaps the most brilliant of all 20<sup>th</sup> century statisticians John Tukey, who argued compellingly that basic **data** are the proper focus of applied science, not statistics, not statistical theory, and not necessarily formal inference. This generally implies that investigators should take account of all that is known about the data's source and context. Graphics in particular have potential to suggest new hypotheses, to help ensure sound qualifications, or even to show there may be good reasons to modify or revise initial research questions, depending what one's data have to say, and on the background information that is available. The foregoing elemental graphic has the virtue that it connects directly with the central question that drives one-way ANOVA, yet it can reveal far more than can any (standard ANOVA or other) data summaries; moreover, when generalized to the case of two-way ANOVA, several other advantages will be seen in sharper relief. Students, and applied researchers generally, do well to heed the counsel of John Tukey and so many others who have contributed to his data analytic philosophy, and who have elaborated on the many ways that sound work-a-day data analysis can improve applied science.

**Note:** This is the first of three related documents. The second is nearing completion; it concerns two-way analysis of variance, and graphics that implement the analyses. The third concerns planned comparison contrasts, as well as a number of thoughts and opinions about data analytic practice surrounding applied scientific uses of analysis of variance. Several design matters will be discussed, especially the matter of blocking, various ways showing how effectively to do it, and what can be gained from doing so.

*[Highlight, copy, and paste to your open console in R; and review the documentation.]*

[NOTE: The following version is for the Mac OS X; a pc version follows]

```
gr.owaov <-
function(yy, gp = NULL, dg = 2, jj=1,tit = T, mrktrm=F, pltrds=F,dosqrs=T){
# NB: graphical analysis here corresponds to conventional one-way ANOVA, either vector or matrix
input
# if yy is matrix, n's are taken as equal to one another, and gp input is NULL, but is computed;
# if yy is a vector of scores for all groups, group membership must be given by input vector gp.
# Function plots grouped data from yy, according to group memberships. Note that group data
# sets are first reordered according to sizes of their means, low (left) to high (right).
# arg dg sets no. of decimal points in output display (& can have fundamental importance)
# arg jj sets level off jittering of points/scores around vertical (dashed) lines
# setting tit = F allows user to specify context-specific title instead of generic one (default)
# par is set to make plot square; reset to default after graphic is done. The central boxes
correspond to
# within group variance estimate (blue) and between group estimate (red), TWO s.e. units on side
for each.
```



```

# The F statistic is reflected by the RATIO of the areas of red box to the blue box.
# If pltrds = T (default is F) the marginal distribution of residuals given on right side (around
gr-mean)
# Please address questions, and make suggestions for improvements to: rmpruzek@yahoo.com
rnd2<-function(x) round(x,2)
rnd3<-function(x) round(x,3)
par(pty='s')
mtx<-is.matrix(yy)
if(mtx) ggpp <- rep(1:ncol(yy),ea=nrow(yy))
if(!mtx) yr<-yy
      else yr<-as.vector(yy)
if(mtx) gp<-ggpp
gpf<-factor(gp)
ngps<-length(unique(gpf))
#####
if(ngps==2 & !dosqrs)efftsz<-TRUE
      else efftsz<-FALSE
stats<-matrix(unlist(tapply(yr,list(gpf),function(x)yy<-c(mean(x),var(x),sd(x))))),byrow=T,ncol=3)
gpn<-as.numeric(gpf)
yrm<-stats[,1][gpn]
tabc<-table(gpf)
grm<-mean(yr)
tabc.dm<-tabc/mean(tabc)
mn.n<-mean(tabc)
stats<-cbind(tabc,stats[,1]-grm,tabc.dm*stats[,1],stats)
statsro<-stats[order(stats[,4]),]
yro<-yr
stats.vc<-yrm-grm
rng.v<-range(yro)
rng.h<-range(stats.vc)
rng.rt<-diff(rng.v)/diff(rng.h)
rng.vv<-c(rng.v[1]-.3*diff(rng.v),rng.v[2]+.3*diff(rng.v))
rng.rt2<-diff(rng.vv)/diff(rng.h)
rng.h2<-1.6*rng.h
rng.rt2<-diff(rng.vv)/diff(rng.h2)
rng.sts<-range(stats[,2])
ammt<-(jj)*(.012)*diff(rng.sts)
#i.e.jitter range defaulted at 1.2% range(contrast values (x's)); can be modified w/ jj
stats.vcj<-jitter(stats.vc,am=ammt)
plot(stats.vcj,yro,pch=16,cex=.9,xlab="",ylab="",xlim=rng.h2,ylim=rng.vv,axes=F)
box(lwd=1.7)
title(xlab="Contrast coefficients; based on observed group means, and numbers in groups")
title(ylab="Dependent variable (response)")
mtext(side=4,'Group means (red triangles)',line=3.5,cex.axis=0.7)
mtext(side=3,text=as.character(statsro[,1]),at=statsro[,2],las=2,line=-2)
points(0,grm,pch=16,cex=1.9,col=3) #puts (heavy) green dot at center, for grand mean
axis(side=1,at=rnd2(statsro[,2]),cex.axis=.8)
axis(side=1,at=0,line=.75,cex.axis=.75)
axis(side=2,at=round(grm,dg-1))
abline(h=grm,lty=3,lwd=2,col=3)
axis(side=4,at=round(stats[,4],dg-1),las=2,cex.axis=.9)
#might reset digits to dg instead of dg-1
axis(side=2,at=round(range(yro),dg-1))
trmd.mn<-tapply(yro,list(gpf),mean,tr=.2)
if(mrktrm){points(statsro[,2],trmd.mn[order(stats[,4])],pch=4,cex=2.3,lwd=2,col='green')}
mtext('Green crosses show 20% trimmed means',adj=0,side=1,line=-1.5,col='green')}
mtext('Large red triangles show means of groups',adj=0,side=1,line=-2.5,col='red')}
mtext('Group Sizes:',adj=0,line=-2)
if(dosqrs)mtext('Area of blue square depicts MS-within',adj=0,line=-4,col='blue')}
if(dosqrs)mtext('Area of red square shows MS-between',adj=0,line=-5,lwd=1.5,col='red')}
abline(h=stats[,4],lty=2,lwd=.5,col=4)
lines(stats[,2],stats[,4],lwd=2,col=4)
points(stats[,2],stats[,4],pch=2,cex=2.3)
points(stats[,2],stats[,4],pch=17,cex=2,col='red')#original means, triangles
abline(v=stats[,2],lty=3,lwd=1,col=1)
mmy<-grm
if(tit) title(paste('An elemental graphic for general one-way ANOVA,
displaying',ngps,'groups'),line=2.6,cex=1.2)
axis(side=3,at=statsro[,2],paste("G",order(stats[,4])),line=-.2,las=2,cex.axis=0.75)
r.xy.sqd<-cor(yro,stats.vc)^2
#old way to compute SS.bet, etc., but works well, so have left it in

```

```

SS.tot<- (length(yro)-1)*var(yro)
SS.bet<-r.xy.sqd*SS.tot
df.b<-ngps-1
df.w<-length(yro)-1-df.b
SS.w<-SS.tot-SS.bet
MS.w<-SS.w/df.w
MS.b<-SS.bet/df.b
resds<-rnd3(yr-stats.vc)
if(pltrsd) rug(resds,side=4,ti=.02,lwd=1.3,col=4)
sd.resd<-sd(resds)*sqrt((length(yr)-1)/df.w)
print(paste('sd.resids = s.d. within =',rnd3(sd.resd)))
grm.pm.sdw<-c(grm-sd.resd,grm+sd.resd) #this interval based on pooled standard error within
grm.pm.sewR<-round(grm.pm.sdw,dg-1)
axis(side=2,at=grm.pm.sewR,cex.axis=1)
axis(side=2,at=grm,labels='grm',las=2,line=-3.5,cex.axis=1.2)
mtext(c('grm-sdw','grm+sdw'),at=grm.pm.sdw,side=2,line=-1.3,cex=1) #change cex as desired
abline(h=grm.pm.sdw,lty=3,lwd=.9,col='blue')
F.stat <- MS.b/MS.w
efsz<-2*sqrt(F.stat)/sqrt(df.w)
p.F <- 1 - pf(F.stat, df.b,df.w)
sqrF<-sqrt(F.stat)
sqrS<-2*sqrt(MS.w)/rng.rt2
if(dosqrs) symbols(0,grm,squares=sqrs,lwd=2,inches=F,fg='blue',lty=1,,add=T)
if(dosqrs) symbols(0,grm,squares=sqrs*sqrt(F.stat),lwd=2,inches=F,fg='red',lty=1,,add=T)
if(mrktrm){print('The 20% trimmed means, ordered as in stats below, are:')
  print(rnd2(as.matrix(trmd.mn[order(stats[,4])])))}
if(!efftsz) legend('bottomright',paste('F-statistic=',rnd2(F.stat)),cex=1,bty='n')
  else legend('bottomright',paste('t-statistic=',rnd2(sqrt(F.stat))),cex=1,bty='n')
gsummary<-array(c(grm,df.b,df.w,MS.b,MS.w,F.stat,p.F))
if(efftsz)print(paste('Magnitude of standardized effect size =',rnd3(efsz)))
print(paste('Ratio of SS.bet to SS.tot is:',rnd3(r.xy.sqd)))
dimnames(gsummary)<-list(c('grand
mean','df.betw','df.with','MS.betw','MS.with','F.statistic','prob.F'))
dimnames(statsro)[2]<-list(c('no. in group','contrast coef','wt-d grp means','grp mean','grp
variance','grp s.d.'))
list(gsummary=round(gsummary,dg),stats=round(statsro,dg))
}

-----#and now the PC version (minor changes only)

gr.owaovpc <- function(yy, gp = NULL, dg = 2, jj=1,tit = T, mrktrm=F, pltrsd=F,dosqrs=T){
# Graphical analysis here corresponds to conventional one-way ANOVA, either vector or matrix input
# if yy is matrix, n's are taken as equal to one another, and gp input is NULL, but is computed;
# if yy is a vector of scores for all groups, group membership must be given by input vector gp.
# Function plots grouped data from yy, according to group memberships. Note that group data
# sets are first reordered according to sizes of their means, low (left) to high (right).
# arg dg sets no. of decimal points in output display (& can have fundamental importance)
# arg jj sets level off jittering of points/scores around vertical (dashed) lines
# setting tit = F allows user to specify context-specific title instead of generic one (default)
# par is set to make plot square; reset to default after graphic is done. The central boxes
#correspond to
# within group variance estimate (blue)& between group estimate (red),TWO s.e. units on each side
#The F statistic is reflected by the RATIO of the areas of red box to the blue box.
#If pltrsd = T (default is F) the marginal distribution of residuals given on right side (around
#gr-mean)
# Please address questions, and make suggestions for improvements to: rmpruzek@yahoo.com
rnd2<-function(x) round(x,2)
rnd3<-function(x) round(x,3)
par(pty='s')
mtx<-is.matrix(yy)
if(mtx) ggpp <- rep(1:ncol(yy),ea=nrow(yy))
if(!mtx) yr<-yy
else yr<-as.vector(yy)
if(mtx) gp<-ggpp
gpf<-factor(gp)
ngps<-length(unique(gpf))
if(ngps==2 & !dosqrs)efftsz<-TRUE
else efftsz<-FALSE
stats<-matrix(unlist(tapply(yr,list(gpf),function(x)yy<-c(mean(x),var(x),sd(x))),byrow=T,ncol=3))
gpn<-as.numeric(gpf)
yrm<-stats[,1][gpn]

```

```

tabc<-table(gpf)
grm<-mean(yr)
tabc.dm<-tabc/mean(tabc)
mn.n<-mean(tabc)
stats<-cbind(tabc,stats[,1]-grm,tabc.dm*stats[,1],stats)
statsro<-stats[order(stats[,4]),]
yro<-yr
stats.vc<-yrm-grm
rng.v<-range(yro)
rng.h<-range(stats.vc)
rng.rt<-diff(rng.v)/diff(rng.h)
rng.vv<-c(rng.v[1]-.3*diff(rng.v),rng.v[2]+.3*diff(rng.v))
rng.rt2<-diff(rng.vv)/diff(rng.h)
rng.h2<-1.6*rng.h
rng.rt2<-diff(rng.vv)/diff(rng.h2)
rng.sts<-range(stats[,2])
#rng.sts<-abs(diff(range(stats[,2])))
ammt<- (jj/100)*diff(rng.sts)
stats.vcj<-jitter(stats.vc,am=ammt)
plot(stats.vcj,yro,pch=16,cex=.7,xlab="",ylab="",xlim=rng.h2,ylim=rng.vv,axes=F)
box(lwd=1.5)
title(xlab="Contrast coefficients; based on observed group means (and numbers in rroups)",cex=.5)
title(ylab="Dependent variable (response)",cex=.5,line=3.7)
mtext(side=4,'Group means (red triangles)',line=2.3,las=0,cex.axis=0.85,cex=.75)
mtext(side=3,text=as.character(statsro[,1]),at=statsro[,2],las=2,adj=1,line=-,cex.axis=.85,cex=.7)
points(0,grm,pch=16,cex=1.7,col=3) #puts (heavy) green dot at center, for grand mean
axis(side=1,at=rnd2(statsro[,2]),cex.axis=.75)
axis(side=1,at=0,line=.75,cex.axis=.75)
axis(side=2,at=round(grm,dg-1),cex=.55)
abline(h=grm,lty=3,lwd=2,col=3)
axis(side=4,at=round(stats[,4],dg-1),las=2,cex.axis=.7)
#might reset digits to dg instead of dg-1
axis(side=2,at=round(range(yro),dg-1),cex=.55)
trmd.mn<-tapply(yro,list(gpf),mean,tr=.2)
if(mrktrm){points(statsro[,2],trmd.mn[order(stats[,4])],pch=4,cex=2,lwd=2.5,col='green')}
mtext('Green crosses show 20% trimmed means',adj=0,side=1,line=-1.5,col='green',cex=.8)}
mtext('Large red triangles show means of groups',adj=0,side=1,line=-2.4,col='red',cex=.8)
mtext('Group Sizes:',adj=0,line=-2.3,cex=.65)
if(dosqrs)mtext('Area of blue square depicts MS-within',adj=0,line=-4,col='blue',cex=.8)
if(dosqrs)mtext('Area of red square shows MS-between',adj=0,line=-5,lwd=1.5,col='red',cex=.8)
abline(h=stats[,4],lty=2,lwd=.4,col=4)
lines(stats[,2],stats[,4],lwd=1.6,col=4)
points(stats[,2],stats[,4],pch=2,cex=1.45)
points(stats[,2],stats[,4],pch=17,cex=1.3,col='red')#original means, triangles
abline(v=stats[,2],lty=3,lwd=1,col=1)
mmy<-grm
if(tit) title(paste('An elemental graphic for general one-way ANOVA,
displaying',ngps,'groups'),line=2.2,cex=.55)
axis(side=3,at=statsro[,2],paste("G",order(stats[,4])),line=-.4,las=2,cex.axis=0.7)
r.xy.sqd<-cor(yro,stats.vc)^2
SS.tot<-(length(yro)-1)*var(yro)
SS.bet<-r.xy.sqd*SS.tot
df.b<-ngps-1
df.w<-length(yro)-1-df.b
SS.w<-SS.tot-SS.bet
MS.w<-SS.w/df.w
MS.b<-SS.bet/df.b
resds<-rnd3(yr-stats.vc)
if(pltrsd) rug(resds,side=4,ti=.02,lwd=1.3,col=4)
sd.resd<-sd(resds)*sqrt((length(yr)-1)/df.w)
print(paste('sd.resids = s.d. within =',rnd3(sd.resd)))
grm.pm.sdw<-c(grm-sd.resd,grm+sd.resd) #this interval based on pooled standard error within
grm.pm.sewR<-round(grm.pm.sdw,dg-1)
axis(side=2,at=grm.pm.sewR,cex.axis=.7)
axis(side=2,at=grm,labels='grm',las=2,line=-2.2,cex.axis=.7)
mtext(c('grm-sdw','grm+sdw'),at=grm.pm.sdw,side=2,line=-2.6,cex=.7) #change cex as desired
abline(h=grm.pm.sdw,lty=3,lwd=.4,col='blue')
F.stat<-MS.b/MS.w
efsz<-2*sqrt(F.stat)/sqrt(df.w)
p.F<-1-pf(F.stat,df.b,df.w)
sqrF<-sqrt(F.stat)

```

```

sqrts<-2*sqrt(MS.w)/rng.rt2
if(dosqrs)symbols(0,grm,squares=sqrts,lwd=2,inches=F,fg='blue',lty=1,,add=T,cex=.8)
if(dosqrs)symbols(0,grm,squares=sqrts*sqrt(F.stat),lwd=2,inches=F,fg='red',lty=1,,add=T,cex=.8)
if(mrktrm){print('The 20% trimmed means, ordered as in stats below, are:')
print(rnd2(as.matrix(trmd.mn[order(stats[,4])])))}
if(!efftsz)legend('bottomright',paste('F-statistic=',rnd2(F.stat)),cex=.75,bty='n')
else legend('bottomright',paste('t-statistic=',rnd2(sqrt(F.stat))),cex=.75,bty='n')
gsummary<-array(c(grm,df.b,df.w,MS.b,MS.w,F.stat,p.F))
if(efftsz)print(paste('Magnitude of standardized effect size =',rnd3(efs)))
print(paste('Ratio of SS.bet to SS.tot is:',rnd3(r.xy.sqd)))
dimnames(gsummary)<-list(c('grand
mean','df.betw','df.with','MS.betw','MS.with','F.statistic','prob.F'))
dimnames(statsro)[2]<-list(c('no. in group','contrast coef','wt-d grp means','grp mean','grp
variance','grp s.d.'))
list(gsummary=round(gsummary,dg),stats=round(statsro,dg))
}

```

#----AN AUXILLARY FUNCTION TO FACILITATE BOOTSTRAPPING; use only when input yy is a matrix.

```

samp.mat.boot <- function(x){
# samples each column of input data matrix x, with replacement, outputs matrix
# of the same dimension; this function intended for use w/ grph.g.owaov,
# when yy is an input matrix (and n's are therefore equal)
y<-x
for(i in 1:ncol(y))y[,i]<-sample(x[,i],nrow(x),repl=T)
list(y=y)
}

```