1 Introduction

In statistics, there are many methods to describe the data, such as inferring about population parameters, comparing population parameters and hypothesis testing of population parameters. But underlying theory is mostly difficult to understand by non statisticians. As a solution it can be produced interactive and meaningful graphs to understand and trust the relevant statistical methods by non statisticians. There are many softwares which are successful in producing graphical representations for many statistical methods.

This report discusses about one such graphical representation tool in R statistical software, for most commonly used statistical method by almost all type of researchers such as agriculture researchers, medical researchers, food science researchers, econometricians etc. ANOVA (Analysis of Variance) is a very useful statistical tool which can be used to compare group means. Many softwares have the ability to produce ANOVA results by one click and immediately can be concluded about group means, but it is quiet hard to understand why those data give such statistical values and pattern of the data. To give a graphical representation of the ANOVA, R statistical software has a package called granove.

2 Overview of the Package

The package granova itself has 4 functions such as granova.1w, granova.2w, granova.contr and granova.ds.

• granova.1w : Graphical display of data for one-way ANOVA
• granova.2w : Graphical display of data for two-way ANOVA
• granova.contr : Graphic Display of Contrast Effect of ANOVA
• granova.ds : Granova for Display of Dependent Sample Data

2.1 Graphical display of data for one-way ANOVA

This function produces one-way ANOVA results and two-dimensional graph with more information than a simple scatter plot.
The function itself has many parameters and those parameters can be adjusted according to the requirement and finally produce a useful representation of the data to understand the one way ANOVA result. Although it produces one way ANOVA results also, in other words by one line code it produces the one way ANOVA results and graph.

Example 1: Comparing 4 independent sample means; here it compares 3 types of drug effects with the placebo group. Among 100 patients randomly assigned each drug for 25 patients and for another 25 patient assigned placebos and fed those medicine and measured their temperature. Here the temperatures are randomly generated normal distribution.

```r
#Before running Sweave file, have to install "granova",
#"rgl" packages
#install.packages("granova")
#install.packages("rgl")
library("granova")
#Data generated using normal distribution for 3 different
drug groups and placebo group.
#To produce the same result seed set for the random
#number generator.
set.seed(301205)
a<-rnorm(25,20,4)
b<-rnorm(25,25,1)
c<-rnorm(25,22,5)
d<-rnorm(25,18,4)
Data1<-data.frame(Placebo=a,Drug.A=b,Drug.B=c,Drug.C=d)
granova.1w(Data1)

$grandsum
  Grandmean df.bet df.with MS.bet MS.with
     21.86    3.00  96.00  160.13  11.39
F.stat F.prob SS.bet/SS.tot
     14.06    0.00       0.30

$stats
  Size Contrast Coef Wtd Mean Mean Trimd Mean Var. St. Dev.
Drug.C 25  -2.91  18.95  18.95  18.90  10.48    3.24
Placebo 25  -1.04  20.82  20.82  20.78  10.34    3.22
Drug.B 25   1.00  22.86  22.86  23.30  24.03    4.90
Drug.A 25   2.95  24.80  24.80  24.77   0.71    0.84

This function gives
• the p-value of the ANOVA table,
and some descriptive statistics for every groups such as
• sample size
• contrast coefficient (individual group mean - grand mean)
• weighted mean
• trimmed (20%) mean
• variance
• standard deviation

Figure 1: Graphical display of data for one-way ANOVA

In Figure 1 (p. 3), MS-within indicates the mean sum of squares and MS-between indicates
the error sum of squares, and according to the graph the ratio between blue and red boxes
gives the F-ratio. Contrast coefficient is just the different between grand mean and the
individual group means, this measure is used plot the graph and graphically it shows that
every group means are how far away from the grand mean, rather than plotting them in
equally spaced on horizontal axis.

2.2 Graphical display of data for two-way ANOVA

This function produces one-way ANOVA results and three-dimensional graph with more
information than a three-dimensional scatter plot.

\[
ganova.2w(data, formula = NULL, fit = "linear", ident = FALSE,
offset = NULL, ...)
\]

The function itself has many parameters, like previous function and those parameters can
be adjusted according to the requirement and finally produces a useful representation of the
data to understand the two way ANOVA result. Although it produces two way ANOVA
results also, in other words by one line code it produces the two way ANOVA results and
three dimensional graph.

Example 2: An experiment which has two categorical factors and one quantitative re-
response measurement. In this example first explanatory variable has 3 factor levels and other
one has 4 factor levels. To identify whether all combination has same response, doing two
way ANOVA is the suitable way.

```r
> # Factor 1 has 3 levels
> Fact1<-as.factor(rep(rep(1:3,each=4),len=48))
> # Factor 2 has 4 levels
> Fact2<-as.factor(c(rep(1,len=12),rep(2,len=12),rep(3,len=12),
+ rep(4,len=12)))
> Response<-matrix(c(0.31,0.45,0.46,0.43,0.36,0.29,0.40,0.23,
+ 0.22,0.21,0.18,0.23,0.82,1.10,0.88,0.72,
+ 0.92,0.61,0.49,1.24,0.30,0.37,0.38,0.29,
+ 0.43,0.45,0.63,0.76,0.44,0.35,0.31,0.40,
+ 0.23,0.25,0.24,0.22,0.45,0.71,0.66,0.62,
+ 0.56,1.02,0.71,0.38,0.30,0.36,0.31,0.33),
+ nrow=48,ncol=1,byrow=TRUE)
> Data2<-data.frame(Response,Fact1,Fact2)
> #1Data2
> granova.2w(Data2)

[1] Response ~ Fact1 * Fact2
$Fact1.effects
     3     2     1
-0.203  0.065  0.138

$Fact2.effects
     1     3     4     2
```
-0.1650 -0.0869 0.0548 0.1970

$CellCounts.Reordered
Fact2
Fact1 1 3 4 2
     3 4 4 4 4
     2 4 4 4 4
     1 4 4 4 4

$CellMeans.Reordered
Fact2
Fact1 1 3 4 2
     3 0.210 0.235 0.325 0.335
     2 0.320 0.375 0.668 0.815
     1 0.412 0.568 0.610 0.880

$aov.summary
<table>
<thead>
<tr>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fact1</td>
<td>2</td>
<td>1.0330</td>
<td>0.5165</td>
<td>23.222</td>
</tr>
<tr>
<td>Fact2</td>
<td>3</td>
<td>0.9212</td>
<td>0.3071</td>
<td>13.806</td>
</tr>
<tr>
<td>Fact1:Fact2</td>
<td>6</td>
<td>0.2501</td>
<td>0.0417</td>
<td>1.874</td>
</tr>
<tr>
<td>Residuals</td>
<td>36</td>
<td>0.8007</td>
<td>0.0222</td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

This function gives

- the p-value of the ANOVA table

and some descriptive statistics for every groups such as

- sample sizes for each combination of explanatory factor levels
- means of responses for each combination of explanatory factor levels
- factor effects same as contrast coefficient in previous function

In Figure 2 (p. 6), white spheres indicate the group means and blue spheres indicate the responses. Similarly in the previous function the effect measure is used plot the graph and graphically it shows that every group means are how far away from the grand mean, rather than plotting them in equally spaced on horizontal axis. This graph can be rotatable and very useful to understand two way effect for the response. The blue color surface is drawn same as using the line which connected all the individual group means in Figure 1, for two factor it has two such lines and two lines used to draw this blue colored surface.
Figure 2: Graphical display of data for two-way ANOVA
2.3 Graphic Display of Contrast Effect of ANOVA

In ANOVA, the results conclude that whether all group’s means are equal or not only. But using contrast analysis can conclude about the relationship between linear combination of subset of group’s means and linear combination of other subset of group’s means.

\[
\text{ granova.contr(data, contrasts, ylab = "Outcome (response)", xlab = NULL, jj = 1) }
\]

The function itself has many parameters and those parameters can be adjusted according to the requirement and finally produce a useful representations for many contrast hypothesis of the data to understand the contrast analysis and relationship between the group effects which cannot be explained by the simple ANOVA.

Example 3: Contrast analysis for the data in Example 1, for following contrasts.

\[
\begin{align*}
\text{Contrast 1(C1)} & : \frac{\mu_{\text{Placebo}} + \mu_{\text{Drug.A}}}{2} = \frac{\mu_{\text{Drug.B}} + \mu_{\text{Drug.C}}}{2} \\
\text{Contrast 2(C2)} & : \frac{\mu_{\text{Placebo}} + \mu_{\text{Drug.B}}}{2} = \frac{\mu_{\text{Drug.A}} + \mu_{\text{Drug.C}}}{2} \\
\text{Contrast 3(C3)} & : \frac{\mu_{\text{Placebo}} + \mu_{\text{Drug.C}}}{2} = \frac{\mu_{\text{Drug.A}} + \mu_{\text{Drug.B}}}{2}
\end{align*}
\]

\[
> \text{ # Defining contrast matrix.}
> \text{ con<- data.frame( c(-.5,-.5,.5,.5),c(-.5,.5,-.5,.5),}
> \text{ + c(.5,-.5,-.5,.5))}
> \text{ names(con) <- c("C1", "C2", "C3")}
> \text{ granova.contr(Data1, contrasts = con,xlab =}
> \text{ + c("C1", "C2", "C3"),}
> \text{ + ylab="Temperature")}
\]

\[
\text{ summary.lm}
\]

Call:
\[
\text{ lm(formula = resp ~ contrst) }
\]

Residuals:
\[
\begin{array}{cccccc}
\text{ Min} & 1Q & \text{ Median} & 3Q & \text{ Max} \\
-9.9833 & -1.3691 & 0.0718 & 1.7775 & 8.9036
\end{array}
\]

Coefficients:
\[
\begin{array}{cccccc}
\text{ Estimate} & \text{ Std. Error} & \text{ t value} & \text{ Pr(>|t|)} \\
\text{(Intercept) 21.85724} & 0.33750 & 64.762 & < 2e-16 ***
\text{ contrast1} & -1.90435 & 0.67500 & -2.821 & 0.00581 **
\text{ contrast2} & 0.03855 & 0.67500 & 0.057 & 0.95457
\text{ contrast3} & -3.94808 & 0.67500 & -5.849 & 6.83e-08 ***
\end{array}
\]

---

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
Residual standard error: 3.375 on 96 degrees of freedom
Multiple R-squared: 0.3052, Adjusted R-squared: 0.2835
F-statistic: 14.06 on 3 and 96 DF, p-value: 1.141e-07

<table>
<thead>
<tr>
<th>$\text{means.pos.neg.coeff}$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>neg</td>
<td>22.81</td>
<td>20.91</td>
<td>-1.90</td>
<td>-0.56</td>
</tr>
<tr>
<td>pos</td>
<td>21.84</td>
<td>21.88</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>diff</td>
<td>23.83</td>
<td>19.88</td>
<td>-3.95</td>
<td>-1.17</td>
</tr>
<tr>
<td>stEftSze</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\text{contrasts}$

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>[2,]</td>
<td>-0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>[3,]</td>
<td>0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>[4,]</td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

$\text{group.means.sds}$

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>20.82</td>
<td>24.80</td>
<td>22.86</td>
</tr>
<tr>
<td>S.D.s</td>
<td>3.22</td>
<td>0.84</td>
<td>4.90</td>
</tr>
</tbody>
</table>
This function gives following descriptive statistics and some contrast analysis outputs.

- summary of residuals
- contrast coefficient estimates
- summary of pooled group’s mean according to the contrasts
- assigned contrasts
- means and standard deviations of each groups

In Figure 3 (p. 10), blue dotted lines are connected to the group means. In first graph, temperatures measured from patients who were fed placebo or drug A is scattered in left and temperatures measured from patients who were fed drug B or drug C is scattered in right as per the first contrast. Similarly other two following graphs generated according to the specified contrasts. Without considering the R console output of contrast analysis, a non statistician can come to a conclusion about the relationship among the groups very easily by looking at the above graphs, even without having much knowledge in theory behind contrast analysis.
Figure 3: Graphic Display of Contrast Effect of ANOVA
2.4 Granova for Display of Dependent Sample Data

This function produces interactive two dimensional scatter plot which explains more details about difference between two dependent quantitative samples.

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

The function itself has many parameters and those parameters can be adjusted according to the requirement and finally produce a useful representation of the data to understand the difference between two dependent quantitative samples. Although it produces some statistics for conclude the relationship between two samples.

Example 4: Comparing 2 dependent quantitative samples; here it compares a measurement of patients before treatment and after treatment. Fake data created to illustrate the function

```r
> #Measurement after the treatment
> PostMeasure<-rnorm(25,20,3)
> #Measurement before the treatment
> PreMeasure<-rnorm(25,27,5)
> Data3<-data.frame(PreMeasure,PostMeasure)
> granova.ds(Data3, \text{revc} = \text{FALSE}, \text{main} = "Assessment Plot", + \text{ident} = \text{TRUE})
```

Summary Stats

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>25.000</td>
</tr>
<tr>
<td>mean(x)</td>
<td>27.214</td>
</tr>
<tr>
<td>mean(y)</td>
<td>20.798</td>
</tr>
<tr>
<td>mean(D=x-y)</td>
<td>6.416</td>
</tr>
<tr>
<td>SD(D)</td>
<td>6.395</td>
</tr>
<tr>
<td>ES(D)</td>
<td>1.003</td>
</tr>
<tr>
<td>r(x,y)</td>
<td>-0.080</td>
</tr>
<tr>
<td>r(x+y,d)</td>
<td>0.500</td>
</tr>
<tr>
<td>LL 95%CI</td>
<td>3.776</td>
</tr>
<tr>
<td>UL 95%CI</td>
<td>9.056</td>
</tr>
<tr>
<td>t(D-bar)</td>
<td>5.016</td>
</tr>
<tr>
<td>df.t</td>
<td>24.000</td>
</tr>
<tr>
<td>pval.t</td>
<td>0.000</td>
</tr>
</tbody>
</table>

This function gives an interactive graph (Figure 4) and some descriptive statistics such as

- sample size
- means of two dependent samples
Figure 4: Granova for Display of Dependent Sample Data

- mean of difference
- standard deviation of difference
- correlation between two variables
- correlation between sum of variables and difference of variables
- 95% confidence interval for difference
- t-statistic for the sample difference
- degrees of freedom
- p-value for t-statistics
In Figure 4 (p. 12), points are plotted like normal scatter plot but more than scatter plot there are some details about difference between two variables. Red dotted horizontal line is representing the mean of the variable which is labeled in vertical axis. Red dotted vertical line is representing the mean of the variable which is labeled in horizontal axis. There are two black lines one is $y=x$ line and other one is drawn perpendicular to $y=x$ line. As mentioned in the graph green line indicates the 95% confidence interval for difference and thin red dotted line indicates the mean difference. And blue crosses are the projections of the point on the line which is perpendicular to $y=x$ line. Lengths between blue crosses and $y=x$ line indicate the different between two observations of corresponding individuals.

This function itself has a parameter called `ident` that allows to identify the corresponding individual’s index by clicking on the scattered point. And this assessment plot shows the corresponding t-statistic for the difference among two dependent variables.