One-way ANOVA Example

PSYCH 710

This handout provides an example of how to conduct a one-way ANOVAs in R. The functions in this handout assume that R is being run on OS-X, although I believe that most, if not all, of the commands also should work on the Windows version of R.

initialize R

1. Create the folder Rlab2 inside the PSY710 folder located in your home directory. Make sure Rlab2 is empty.
2. Launch R and navigate to Rlab2 using the following command:
   ```
   > setwd("~/PSY710/Rlab2") # set working directory
   ```
3. Before starting this lab, please enter the following **VERY IMPORTANT COMMAND** command at R’s prompt:
   ```
   > options(contrasts=c("contr.sum","contr.poly"))
   ```
   This command sets up R to use the sum-to-zero definition for effects.

analyze Table 1 data

1. Table 1 shows data from an imaginary experiment that used a one-way, between-subjects experimental design: Each subject was assigned randomly to one of three experimental conditions, with the constraint that there had to be 6 subjects per condition. The numbers in the table show the dependent variable from each subject. I used Excel to create an ascii (csv) file that represents those numbers in a long-format file using the variable names group and score. The variable group represents group membership with three values: g1, g2, and g3. A third variable/column, named g, also indicates group membership using the integers 1, 2, and 3. Read the data file command into R using the following command:
   ```
   > myData <- read.csv("http://psycserv.mcmaster.ca/bennett/psy710/datasets/rlab2data1.csv")
   ```
2. Calculate the mean and standard deviation of score for each group using the following commands
   ```
   > # tapply(score,group,mean) # this does not work. why not?
   > with(myData,tapply(score,group,mean)) # this does work. why?
   ```
   ```
   g1   g2   g3
   49.6667 50.8333 43.5000
   ```
3. The data from Table 1 are displayed in Figure 1. What do the various parts of the boxplots represent? Do the boxplots suggest that the assumptions made by the ANOVA are reasonable for these data?
4. We want to evaluate the effect of group on score. Use the command lm() to fit a full linear model that includes score as the dependent variable, and group and an intercept as predictor variables. Then use anova() to use the fitted model to create an ANOVA table. What do the quantities in the two rows of the ANOVA table mean?
5. What are null and alternative hypotheses are being investigated by your ANOVA?

6. Compare the anova tables that you get when you use `group` and `g` as the grouping variables. How do they differ? Why do they differ? (Hint: inspect the contents of `myData`.)

7. Calculate the association strength between `group` and `score`.

 ```r
> summary(lm.01) # print regression table

Call:
  lm(formula = score ~ 1 + group, data = myData)

Residuals:
   Min     1Q Median     3Q    Max
-4.833  -3.083  -0.500   2.958   5.500

Coefficients:

> lm.01 <- lm(score~1+group,data=myData) # linear model
> anova(lm.01) # anova table

Analysis of Variance Table

  Response: score
            Df   Sum Sq  Mean Sq   F value    Pr(>F)  
group       2 186.3333  93.16702  7.446670 0.005674 **
Residuals   15 187.6667  12.51111             
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 1: Boxplots of data in Table 1.
Bennett, PJ

Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.0000  0.8337  57.574  <2e-16 ***
group1      1.6667  1.1790   1.414   0.1779
group2      2.8333  1.1790   2.403   0.0296 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.537 on 15 degrees of freedom
Multiple R-squared: 0.4982, Adjusted R-squared: 0.4313
F-statistic: 7.447 on 2 and 15 DF, p-value: 0.005674

> adj.r.squared <- 0.4313 # from regression table
> (omega.squared <- adj.r.squared) # association strength

[1] 0.4313

8. Evaluate the homogeneity of variance assumption.

> with(myData, bartlett.test(score, group))

Bartlett test of homogeneity of variances

data: score and group
Bartlett's K-squared = 0.37526, df = 2, p-value = 0.8289

> # bartlett.test(myData$score, myData$group) # another way
> # bartlett.test(score~group, data=myData) # and another way!

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>46</td>
<td>49</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>56</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>49</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>54</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>47</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>53</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Data from an imaginary experiment that assigned subjects randomly to one of three experimental conditions.