Statistics Lab #2

PSYCH 710

2 Lab 2: One-way ANOVA

2.1 initialize R

Before answering the questions in this handout, you should read the handout that provides an example of how to conduct a one-way ANOVAs in R, which can be obtained here:

http://www.psychology.mcmaster.ca/bennett/psy710/problems/ps1/Anova_example.pdf

Create the folder Rlab2 inside the PSY710 folder located in your home directory. Make sure Rlab2 is empty. Then launch R and enter the following commands:

```r
> setwd("~/PSY710/Rlab2") # set working directory
> options(contrasts=c("contr.sum","contr.poly")) # sum-to-zero effects
```

2.2 Answer all parts of the following 2 questions.

1. An experiment measured the performance of human subjects in three different test conditions. Thirty subjects were assigned randomly to three groups with the constraint that each group had 10 subjects. Each subject was tested only once. The independent (i.e., grouping) variable is condition; the dependent variable is score.

   (a) Read the data file into R and with the following command:
   ```r
   > myData <- read.csv("http://psycserv.mcmaster.ca/bennett/psy710/datasets/dataset1_Sep25.csv")
   ```

   (b) Calculate the group means and standard deviations.
   ```r
   > (theMeans <- with(myData,tapply(score,condition,mean)) )
   > (theSDs <- with(myData,tapply(score,condition,sd)) )
   ```

   (c) Evaluate the differences among the group means using a one-way ANOVA. Print the ANOVA table, state the null hypothesis that is being evaluated by each p-value, and state your conclusion regarding the null hypothesis.

   ```r
   > myData.aov.01 <- aov(score~1+condition,data=myData)
   > anova(myData.aov.01)
   ```

Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<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>condition</td>
<td>2</td>
<td>371.35</td>
<td>185.677</td>
<td>1.8582</td>
<td>0.1754</td>
</tr>
<tr>
<td>Residuals</td>
<td>27</td>
<td>2697.93</td>
<td>99.923</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
> # summary(myData.aov.01) # this command also prints the anova table
> # the next 2 lines use the lm command:
> # myData.lm.01 <- lm(score~1+condition,data=myData)
> # anova(myData.lm.01)

(d) Our ANOVA assumes that the data in each group are distributed normally. This assumption is equivalent to saying that the residuals of the best-fitting model are distributed normally. Use a statistical test to evaluate the normality assumption for the current data.

> the.residuals <- residuals(myData.aov.01)
> shapiro.test(the.residuals)

Shapiro-Wilk normality test
data: the.residuals
W = 0.9714, p-value = 0.5792

> # you can combine the previous 2 lines into 1 command:
> # shapiro.test(residuals(myData.aov.01))

2. An experiment was conducted to evaluate the effectiveness of four Grade-5 math textbooks. A total of 80 students were assigned randomly to one of four classes, and one textbook was assigned randomly to each class. Each class was taught by the same teacher. At the end of the year, knowledge of math was assessed using a standardized test.

(a) Read the data file into R and with the following command:

> mathbooks <- read.csv("http://psycserv.mcmaster.ca/bennett/psy710/datasets/dataset2_Sep25.csv")

(b) Identify the names and classes (i.e., types) of the variables in the data frame mathbooks.

> names(mathbooks)
[1] "book" "testscore"
> class(mathbooks$book)
[1] "factor"
> class(mathbooks$testscore)
[1] "numeric"

(c) Use a one-way ANOVA to evaluate the null hypothesis of no difference among group means.

> books.aov.01 <- aov(testscore~book,data=mathbooks)
> anova(books.aov.01)

Analysis of Variance Table

Response: testscore

<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>book</td>
<td>3</td>
<td>5664</td>
<td>1887.93</td>
<td>3.8503</td>
</tr>
<tr>
<td>Residuals</td>
<td>76</td>
<td>37265</td>
<td>490.33</td>
<td></td>
</tr>
</tbody>
</table>

---

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

(d) Our ANOVA assumes that the variance is constant across groups. Perform a statistical test to evaluate the homogeneity of variance assumption.

> bartlett.test(testscore~book,data=mathbooks)
Bartlett test of homogeneity of variances

data: testscore by book
Bartlett's K-squared = 17.7851, df = 3, p-value = 0.0004871

Answer: The Bartlett test evaluates the null hypothesis that the variance does not vary across groups. The test was significant ($K^2 = 17.8$, df=3, $p = 0.00048$), so we reject the null hypothesis in favor of the alternative (i.e., that the variance differed across groups). Given the results of this test, we might want evaluate the group means with \texttt{oneway.test}.

```r
> oneway.test(testscore~book,data=mathbooks)
```

One-way analysis of means (not assuming equal variances)

data: testscore and book
F = 2.1311, num df = 3.000, denom df = 41.047, p-value = 0.111

(e) Suppose the data in \texttt{mathbooks} did not come from an experiment that assigned students randomly to groups/textbooks. Instead, suppose they came from an experiment in which textbooks were assigned randomly to four existing classes in different schools. Would this change in the experimental design influence your interpretation of ANOVA? Explain.

Answer: The numerical results would remain the same, but our interpretation of the effectiveness of the textbooks would change. Our original experimental design assigned students to each class/textbook \textit{randomly}, and therefore it is unlikely (though not impossible) that the differences among groups were due to differences in the students assigned to each class. However, in this new design, using pre-existing classes means that textbook is confounded with class (and, if different teachers are used, teachers). Therefore, we are less certain about what caused the differences among group means. The point of this question is to illustrate that interpretations of the statistical analyses can depend heavily on the way in which the data were collected.