Homework 07 Answer Sheet

Psych 10C

Due: Sunday, October 30th (by 11:59pm PT)

Zhenze Zhang

Submission Details

- Download *HW07AnswerSheet.Rmd* from the Canvas course space and open it RStudio.
- Enter your name in the *author* field at the top of the document.
- Complete the assignment by entering your answers in your *HW07AnswerSheet.Rmd* document.
- Once you have completed the assignment, click the *Knit* button to turn your completed answer document into a pdf file.
- Submit your HW07AnswerSheet.pdf file only (no other formats are acceptable) before the assignment's deadline.

Problems

For each problem, show/describe all of your work.

Problem #1 (9 points)

A new treatment meant to help those with chronic arthritis pain was developed and tested for its long-term effectiveness. Participants in the experiment rated their level of pain on a 0 (no pain) to 9 (extreme pain) scale at three-month intervals. Was the treatment effective? Use RStudio to perform the following steps of a repeated-measures ANOVA on the data. Perform your analysis step-by-step using RStudio, but without using RStudio's *aov()* function. Use an $\alpha = .05$.

Load the data from our course website (HW07_ArthritisData.csv) into a data frame named *ArthritisData*. The pain scores will be contained in the *Pain* variable of the data frame (ArthritisData$Pain) and the different points in time at which measurements were made (i.e. the different conditions) will be contained in the *Time* variable (ArthritisData$Time). The four "Time" levels are "Before", "3mo", "6mo", and "9mo".

```
ArthritisData<- read.csv("/Users/zzze/Downloads/HW07_ArthritisData.csv")
```

(a) (2 points) Use RStudio to compute and print all degrees of freedom for your repeated-measures analysis.

ANSWER:

```
N<- nrow(ArthritisData)
k<- length(table(ArthritisData$Time))
n<- sum(ArthritisData$Time == "Before")
```

```
dfTotal<- N-1
dfTotal
```

```
## [1] 15
```

```
dfBetCond<- k-1
dfBetCond
```

```
## [1] 3
```

```
dfWithinCond<- dfTotal-dfBetCond
dfWithinCond
```

```
## [1] 12
```

```
dfBetSub<- n-1
dfBetSub
```

```
## [1] 3
```

```
dfError<- (n-1)*(k-1)
dfError
```

```
## [1] 9
```

(b) (3 points) Use RStudio to compute and print all sums of squares terms for your repeated-measures analysis

ANSWER:

```
sumx<- sum(ArthritisData$Pain)
sumxsq<- sum(ArthritisData$Pain ^2)
SSTotal<- sumxsq- sumx^2/N
SSTotal
```

```
## [1] 19.9375
```

```
TimeTotal<- xtabs(Pain~ Time, ArthritisData) # the totals for each condition
SumCondTotalsq<-sum(TimeTotal^2)
SSBetCond<- SumCondTotalsq/n - sumx^2/ N
SSBetCond
```

```
## [1] 6.6875
```

```r
SSWithinCond<- SSTotal-SSBetCond
SSWithinCond
```

```
## [1] 13.25
```

```r
SubjectTotals<- xtabs(Pain~ Subject, ArthritisData)
SumSubTotalsq<- sum(SubjectTotals^2)
n.perSubject<- sum(ArthritisData$Subject== "Sub1")
SSBetSub<- SumSubTotalsq/ n.perSubject - sumx ^2/ N
SSBetSub
```

```
## [1] 12.1875
```

```r
SSError<- SSWithinCond- SSBetSub
SSError
```

```
## [1] 1.0625
```

(c) (2 points) Use RStudio to compute and print the MS terms, the observed F statistic, and the critical F value for your analysis.

ANSWER:

```r
MSBetCond<- SSBetCond/dfBetCond
MSBetCond
```

```
## [1] 2.229167
```

```r
MSError<- SSError/ dfError
MSError
```

```
## [1] 0.1180556
```

```r
F<- MSBetCond/ MSError
F
```

```
## [1] 18.88235
```

```r
FCritical<- qf(0.05, dfBetCond, dfError, lower.tail = FALSE)
FCritical
```

```
## [1] 3.862548
```

(d) (2 points) Given the results of your analysis, should you reject or accept the null hypothesis? Explain/interpret your decision.

ANSWER:

- Since our F value of 18.88235 is greater than our critical F-value of 3.862548 we can reject the null hypothesis and claim that the treatment have differing effects on the pain relief.

Problem #2 (5 points)

Perform a repeated-measures analysis of variance on the data from Problem #1 using RStudio's *aov()* function.

(a) (2 points) Run the *aov()* function to perform your repeated-measures analysis. Save the output to your own variable and display it using the summary() function.

ANSWER

```
my.anova<- aov(Pain~Time+ Error(Subject), ArthritisData)
summary(my.anova)
```

```
##
## Error: Subject
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  3  12.19   4.062
##
## Error: Within
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Time       3  6.687  2.2292   18.88 0.000319 ***
## Residuals  9  1.062  0.1181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(b) (1 point) Verify that your conclusion is the same as in Problem #1.

ANSWER:
Since the p value of 0.000319 is less than 0.05, we can rejct the null hypothesis. The answer is the same as in Problem 1.

(c) (2 points) Use RStudio's aov() function to perform an *independent-measures* anova on the same data. Would we make the same decision regarding our null hypothesis if this data came from DIFFERENT subjects in all conditions? Why?

ANSWER:

```
my.anovaInd<- aov(Pain~Time, ArthritisData)
summary(my.anovaInd)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Time         3  6.687   2.229   2.019  0.165
## Residuals   12 13.250   1.104
```

- Since the p value of 0.165 is greater than 0.05, we would not reject the null hypothesis, thus not make the same decision. Since the repeated-measures ANOVA allows us to remove the "between subjects" sum of square from the ""within conditions" sum of square, we can reduce error term.

Problem #3 (6 points)

An investigator is interested in comparing the cardiovascular fitness of elite runners on three different training courses, each of which covers 10 miles. The courses differ in terms of terrain: Course 1 is flat, Course 2 has graded inclines, and Course 3 includes steep inclines. Each runner's heart rate is monitored at mile 5 of the run on each course. Ten runners are involved, and their heart rates measured on each course can be downloaded on the course website ("HW07_RunnerData.csv").

Load the data from our course website (HW07_RunnerData.csv) into a data frame named *RunnerData*. The heart rate measurements will be contained in the *HeartRate* variable of the data frame (RunnerData$HeartRate) and the conditions (i.e. different courses) will be contained in the *Course* variable (TapeData$Course).

```
RunnerData<- read.csv("/Users/zzze/Downloads/HW07_RunnerData.csv")
```

(a) (2 points) Use RStudio to perform a repeated measures ANOVA on the data. Perform your ANOVA using RStudio's *aov()* function. Use an $\alpha = .05$. Use the *summary()* function to show the results of the ANOVA. Report whether or not the null hypothesis should be rejected and explain your decision.

ANSWER:

```
my.anova2<- aov(HeartRate ~ Course+ Error(Subject), RunnerData)
summary(my.anova2)
```

```
##
## Error: Subject
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  9   2224   247.2
##
## Error: Within
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Course     2  476.5  238.23    15.6 0.000117 ***
## Residuals 18  274.9   15.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Since our p value of 0.000117 is less than 0.05, we can reject the null hypothesis and claim that different courses have differing effects on the runners' heartrate.

(b) (4 points) If the F-statistic calculated in part (a) is significant, perform a Tukey test to compare each pair of conditions. Use $\alpha = .05$. Report HSD, all pairwise mean differences, and interpret the results.

ANSWER:

```r
k2<- length(table(RunnerData$Course)) #number of conditions
n2<- sum(RunnerData$Course == "Course1") # number of samples in each condition
dfError2<- (n2-1)*(k2-1)
FCritical2<- qf(0.95, 2, dfError2, lower.tail = FALSE)
FCritical2
```

```
## [1] 0.05143974
```

- Since the F value of 15.6 is greater than critical F-value of 0.05143974, we can reject the null hypothesis.

```r
MSError<- 15.27
q<- qtukey(0.95, k2, dfError2)
HSD<- q* sqrt(MSError/n2)
HSD
```

```
## [1] 4.460083
```

```r
Course1Mean<- mean(RunnerData[RunnerData$Course == "Course1", "HeartRate"])
Course2Mean<- mean(RunnerData[RunnerData$Course == "Course2", "HeartRate"])
Course3Mean<- mean(RunnerData[RunnerData$Course == "Course3", "HeartRate"])
```

```r
Course1.2<- Course1Mean- Course2Mean
Course1.3<- Course1Mean- Course3Mean
Course2.3<- Course2Mean- Course3Mean
cat("Course1-Course2:",Course1.2,"\nCourse1-Course3:",Course1.3,
    "\nCourse2-Course3:",Course2.3)
```

```
## Course1-Course2: -5.8
## Course1-Course3: -9.7
## Course2-Course3: -3.9
```

- Comparisons among the means for each course using Tukey's HSD (HSD= 4.460083) showed that significantly different effects were yielded by Course 1 & 2 and by Course 1 & 3 (at the 0.05 level).
