Homework 04 Answer Sheet

Psych 10C
Due: Sunday, October 16th (by 11:59pm PT)

Zhenze

Submission Details

- Download *HW04AnswerSheet.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 *HW04AnswerSheet.Rmd* document.
- Once you have completed the assignment, click the *Knit* button to turn your completed answer document into a pdf file.
- Submit your HW04AnswerSheet.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 (4 points)

It is claimed that an industrial safety program is effective in reducing the loss of working hours due to factory accidents. The data below show the weekly loss of working hours due to accidents in six plants both before and after the safety program is initiated. Does the data substantiate the claim that the safety program has reduced the loss of working hours? Use RStudio to perform a repeated-sample t-test with $\alpha = .05$. Print your computed t-statistic, your critical t-value, and explain your conclusion.

PlantNum	Before	After
1	12	10
2	29	28
3	16	17
4	37	35
5	28	25
6	15	16

ANSWER

```r
# related samples t-test, one tail
# n=6, d=x1-x2,
Diffs1<- Data$Before-Data$After
sum(Diffs1)
```

```
## [1] 6
```

```r
# calculate the standard error
SD1<- sd(Diffs1)
n1<- nrow(Data)
SE1<- SD1/ sqrt(n1)
# find t-statistic
X1mean1<- mean(Data$Before)
X2mean1<- mean(Data$After)
tvalue1<- (X1mean1-X2mean1)/ SE1
# find critical t value
Criticalt1<- qt(0.95, n1-1)
cat("t-value:", tvalue1, "\ncritical t-value:", Criticalt1)
```

```
## t-value: 1.46385
## critical t-value: 2.015048
```

- Since the t-value we got is lower than the critical t value, we fail to reject the null hypothesis: there is not enough evidence to say that the safety program has reduced the loss of working hours.

Problem #2 (4 points)

Two methods of memorizing difficult material are being tested to determine if one produces better retention. Nine pairs of students are included in the study. The students in each pair are matched according to I.Q. and academic background and then are assigned to the two methods at random. A memorization test is given to all the students, and the scores below are obtained. Is there is a significant difference in effectiveness of the two methods. Use RStudio to perform a repeated-sample t-test with $\alpha = .05$. Print your computed t-statistic, your critical t-value, and explain your conclusion.

Pair	MethodA	MethodB
1	90	85
2	86	87
3	72	70
4	65	62
5	44	44
6	52	53
7	46	42
8	38	35
9	43	46

ANSWER:

```
# independent. two-tailed, n=9
# find t value
Data1<- Data$MethodA
Data2<- Data$MethodB
q2n1<- length(Data1)
q2n2<- length(Data2)
var1<- var(Data1)
var2<- var(Data2)
SE2<- sqrt(var1/q2n1+var2/q2n2)
tvalue2<- (mean(Data1)- mean(Data2))/ SE2

# find critical t value
critical2<- qt(0.975, q2n1+q2n2-2)

cat("t-value:", tvalue2, "\ncritical t-value:", critical2)
```

```
## t-value: 0.1471451
## critical t-value: 2.119905
```

- Since the t-value is lower than the critical t value, we fail to reject the null hypothesis. Therefore, there isn't any significant difference in effectiveness of the two methods

---

Problem #3 (2 points)

Using RStudio, construct a 95% confidence interval for the mean difference in scores for the two methods in Problem 2.

ANSWER:

```
MeanDiff<- mean(Data1)-mean(Data2)
CILow<- MeanDiff-critical2*SE2
CIHigh<- MeanDiff+critical2*SE2
cat("CI=[", CILow, ",", CIHigh, "]")
```

```
## CI=[ -17.87587 , 20.54253 ]
```

Problem #4 (4 points)

Psychologists have made extensive studies on the relationship between child abuse and later criminal behavior. A study was conducted that consisted of follow-ups of 52 boys who were abused in their preschool years and of 61 boys who were not abused. The data of the number of criminal offenses of those boys in their teens yielded the following summary statistics:

	Abused	Nonabused
Mean	2.52	1.63
Standard Deviation	1.84	1.22

Is the mean number of criminal offenses significantly higher for the abused group than that for the nonabused group? Use RStudio to perform an independent-sample t-test with $\alpha = .05$. Print the value for your pooled variance, your computed t-statistic, your critical t-value, and explain your conclusion.

ANSWER:

```
# alpha= 0.05, n1(abused)= 52, n2(non-abused)= 61, one-tailed test
n1Ab<- 52
n2Non<-61
meanAb<- 2.52
meanNon<- 1.63
varAb<- 1.84^2
varNon<- 1.22^2
PooledVar<- ((n1Ab-1)*varAb+ (n2Non-1)*varNon)/(n1Ab+n2Non-2)
SE4<- sqrt(PooledVar*(1/n1Ab+ 1/n2Non))
tvalue4<- (meanAb-meanNon)/SE4
critical4<- qt(0.95, n1Ab+n2Non-2)
cat("pooled variance:", PooledVar, "\nt-value:", tvalue4, "\ncritical t-value:", critical4)
```

```
## pooled variance: 2.360086
## t-value: 3.069399
## critical t-value: 1.658697
```

- Since the t value is larger than the critical t-value, we can reject the null hypothesis. Therefore, the mean number of criminal offenses is significantly higher for the abused group than that for the non-abused group.

Problem #5 (6 points)

The peak oxygen intake per unit of body weight, called the "aerobic capacity," on an individual performing a strenuous activity is a measure of work capacity. For a comparative study, measurements of aerobic capacities are recorded for a group of 20 Peruvian Highland natives and for a group of 10 U.S. Lowlanders acclimatized as adults in high altitudes. Load the workspace *HW04_AerobicCapacity.Rdata* from the course website. It contains two vectors, *HighlanderData* and *LowlanderData*, which contain the corresponding data. Use RStudio to perform the following.

(a) (3 points) Construct a 98% confidence interval for the mean difference in aerobic capacity between the two groups.

ANSWER:

```
load("/Users/zzze/Downloads/HW04_AerobicData.RData")
DataHigh<-HighlanderData
DataLow<- LowlanderData

nHigh<- length(DataHigh)
nLow<- length(DataLow)
varHigh<- var(DataHigh)
varLow<- var(DataLow)

MeanDiff5<- mean(DataHigh)-mean(DataLow)
CICritical<- qt(0.99, nHigh+nLow-2)
q5PooledVar<- ((nHigh-1)*varHigh+ (nLow-1)*varLow)/ (nHigh+nLow-2)
SE5<- sqrt(q5PooledVar*(1/nHigh + 1/nLow))

CILow5<- MeanDiff5- CICritical*SE5
CIHigh5<- MeanDiff5+ CICritical*SE5
cat("CI=[",CILow5, CIHigh5, "]" )
```

```
## CI=[ 2.530134 12.96297 ]
```

(b) (3 points) Do the data provide a strong indication of a difference in mean aerobic capacity between the highland natives and the acclimatized lowlanders? Use RStudio to perform an independent-sample t-test with $\alpha = .05$. Print the value for your computed t-statistic, your critical t-value, and explain your conclusion.

ANSWER:

```
# variance of highlander data and lowlander data
varHigh<- var(DataHigh)
varLow<- var(DataLow)

# Sample size for each data
nHigh<- length(DataHigh)
nLow<- length(DataLow)

#pooled variance and standard error
q5PooledVar<- ((nHigh-1)*varHigh+ (nLow-1)*varLow)/ (nHigh+nLow-2)
SE5<- sqrt(q5PooledVar*(1/nHigh + 1/nLow))

#t statistics
tvalue5<- (mean(HighlanderData)-mean(LowlanderData))/ SE5

# two tailed test, alpha=0.05, df=nHigh+nLow-2
critical5<- qt(0.975, nHigh+nLow-2 )

cat("pooled variance:", q5PooledVar, "\nt-value:", tvalue5, "\ncritical t-value:", critical5)
```

```
## pooled variance: 29.80338
## t-value: 3.663785
## critical t-value: 2.048407
```

- Since the t-value is larger than the critical t value, we can reject the null hypothesis. In conclusion, the data provide a strong indication of a difference in mean aerobic capacity between the highland natives and the acclimatized lowlanders.
