HW9

106022113

Helped by 106022103 in Question 1

Question 1

Parameters: diff = 0.3, sd =2.9, n = 50, alpha = 0.05

i. Will this scenario create systematic or random error?

ii. Which part of the t-statistic would be affected?

iii. Will it increase or decrease our power to reject the null hypothesis?

iv. Which kind of error becomes more likely (Type I or Type II)?

a. Target the general population of Taiwanese users, but missed older customers whom might use much less.

i. ANSWER: The main error here is the coverage error. Since in sampling from the population, the sample missed the population in older customers. Hence, this will neither create systematic nor random error because the error is not from measurement.

ii. ANSWER: Since we missed a segement of the population which uses much less, the difference and the standard deviation will alter.

iii. ANSWER: The older customers use much less of the product, so the mistake will increase the power to reject the null hypothesis.

iv. ANSWER: Hence, Type I error becomes more likely

b.20 people wears the wrong device and should be removed from data.

i. ANSWER: The sample size will change from 50 to 30 after removing the data. Hence, the standard error will increase after dividing the square root of the sample size, and it will create random errors.

ii. ANSWER: The sample size n changes

iii. ANSWER: The decrease of the sample size in the simulation plot indicates that the purple area being reduced, so the power to reject the null hypothesis will also decrease.

iv. ANSWER: Type II error will be more likely to occur.

c. Annoying professor suggested the alpha be changed to 0.9

i. ANSWER: No error will be generated here since we just tuned the parameter.

ii. ANSWER: Alpha will be changed because of the annoying professor.

iii. ANSWER: In the simulation, the purple area increases, so the power to reject the null hypothesis will also increase.

iv. ANSWER: Type I error will be more likely to occur.

d. Missed data for teenagers in weekend, may exaggerate older people's data.

i. ANSWER: Since there will be bias on satisficaton measurement, there will be error that teenagers scored lower. And this will create systematic error.

ii. ANSWER: The difference and standard deviation will alter.

iii. ANSWER: In the simulation, with the standard deviation becoming larger, the purple area decreases, indicating the power to reject the null hypothesis reduces.

iv. ANSWER: Type II error will be more likely to occur.

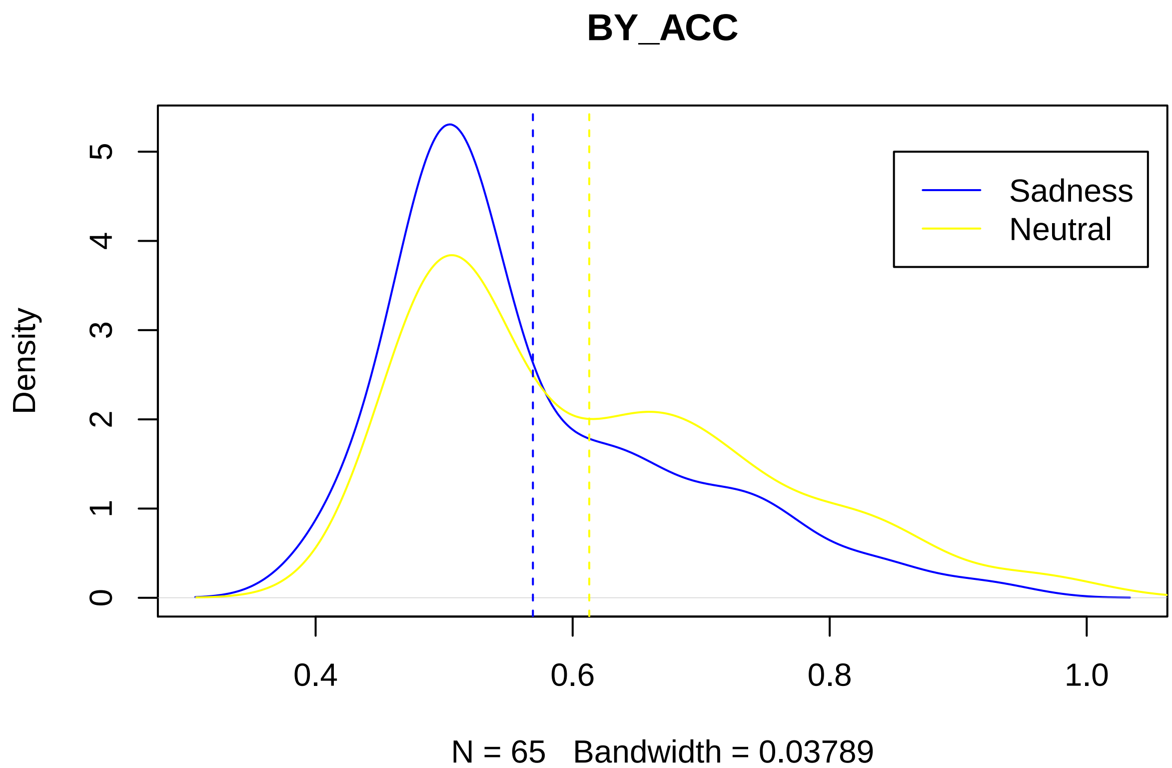
Question 2.

```
experiment <- read.csv('study2Data.csv', header=TRUE)
BY_data <- with(experiment, data.frame(Subject, Axis='BY', Emotion_Condition, ACC=BY_ACC, SAD_ESRI))
RG_data <- with(experiment, data.frame(Subject, Axis='RG', Emotion_Condition, ACC=RG_ACC, SAD_ESRI))
```

a. Visualize the differences between BY, RG and it's emotional conditions

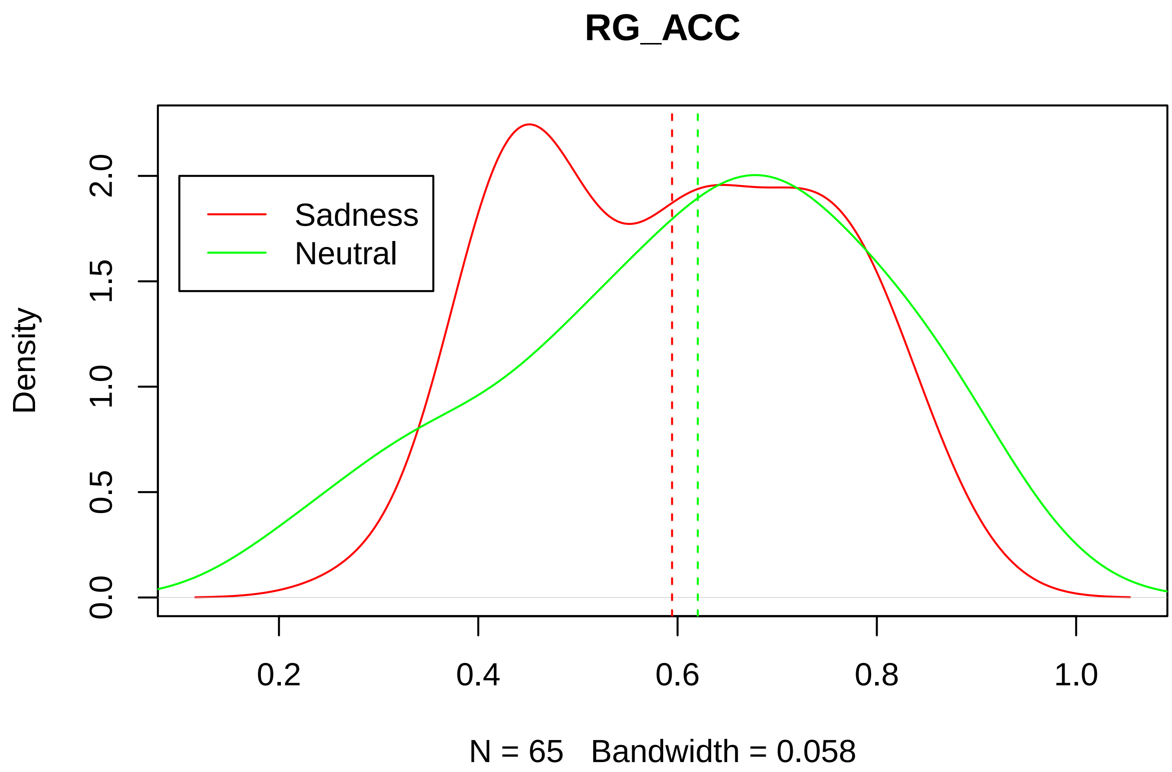
```
sad_by <- BY_data[BY_data$Emotion_Condition=="Sadness",]
neu_by <- BY_data[BY_data$Emotion_Condition=="Neutral",]
plot(density(sad_by$ACC),main = "BY_ACC", col = "blue")
lines(density(neu_by$ACC),col = "yellow")
legend(0.85,5,c("Sadness","Neutral"),col = c("blue","yellow"),lty = c(1,1))
abline(v = mean(sad_by$ACC),col = "blue",lty = 2)
abline(v = mean(neu_by$ACC),col = "yellow",lty = 2)
```

Compare between Sad and Neutral for BY_ACC



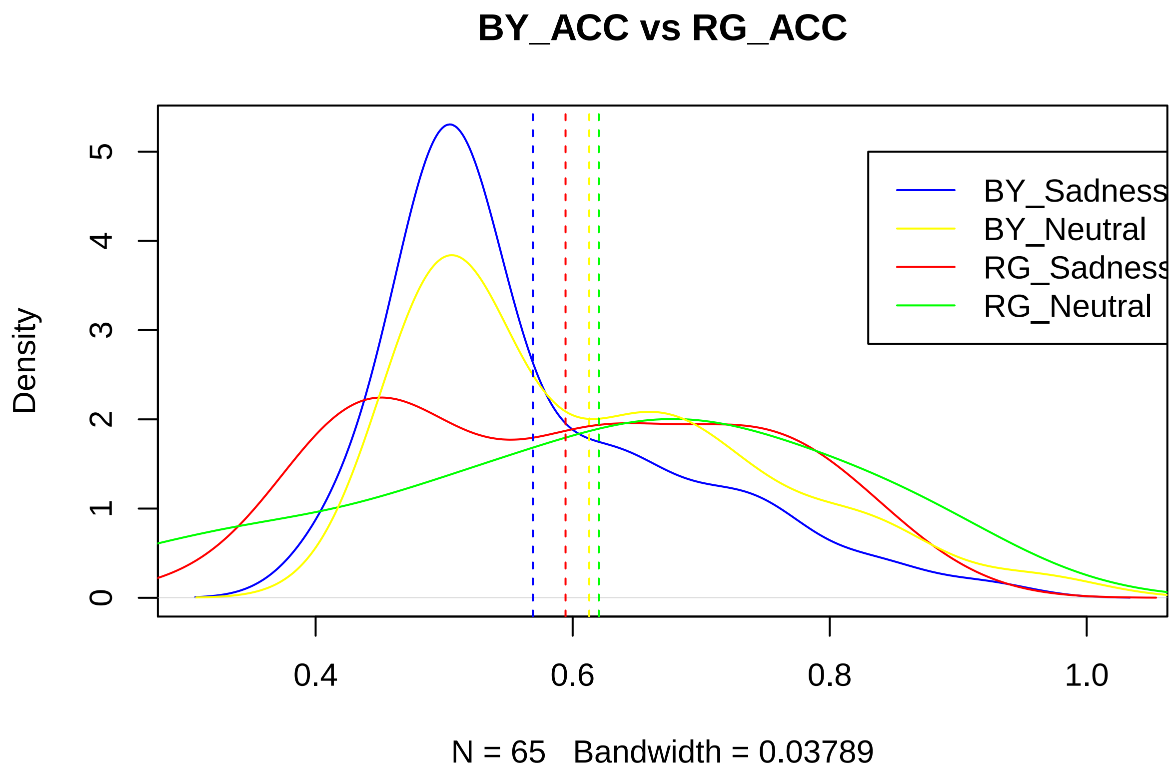

```
sad_rg <- RG_data[RG_data$Emotion_Condition=="Sadness",]
neu_rg <- RG_data[RG_data$Emotion_Condition=="Neutral",]
plot(density(sad_rg$ACC),main = "RG_ACC", col = "red")
lines(density(neu_rg$ACC),col = "green")
legend(0.1,2,c("Sadness","Neutral"),col = c("red","green"),lty = c(1,1))
abline(v = mean(sad_rg$ACC),col = "red",lty = 2)
abline(v = mean(neu_rg$ACC),col = "green",lty = 2)
```

Compare between RG_ACC




```
plot(density(sad_by$ACC),main = "BY_ACC vs RG_ACC", col = "blue")
lines(density(sad_rg$ACC), col = "red")
lines(density(neu_rg$ACC),col = "green")
lines(density(neu_by$ACC),col = "yellow")
legend(0.83,5,c("BY_Sadness","BY_Neutral","RG_Sadness","RG_Neutral"),col = c("blue","yellow","red","gre
abline(v = mean(sad_by$ACC),col = "blue",lty = 2)
abline(v = mean(neu_by$ACC),col = "yellow",lty = 2)
abline(v = mean(sad_rg$ACC),col = "red",lty = 2)
abline(v = mean(neu_rg$ACC),col = "green",lty = 2)
```

BY_data vs RG_data



We can see that the means are really close to each other at around 0.6.

b. Run a t-test check if there is a significant difference between sad and neutral participants at 95% confidence (BY_data)

```
t.test(neu_by$ACC,sad_by$ACC)
```

```
##
##  Welch Two Sample t-test
##
## data:  neu_by$ACC and sad_by$ACC
## t = 2.0435, df = 125.61, p-value = 0.04309
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.001384159 0.086308149
## sample estimates:
## mean of x mean of y
## 0.6129231 0.5690769
```

ANSWER: Since the p-value is under 0.05, we can reject the null hypothesis that there isn't different between the means of these two values. So there is a significant difference between the sad and neutral accuracy.

c.Run a t-test check if there is a significant difference between sad and neutral participants at 95% confidence (RG_data)

```
t.test(neu_rg$ACC,sad_rg$ACC)
```

```
##
##  Welch Two Sample t-test
##
## data:  neu_rg$ACC and sad_rg$ACC
## t = 0.87491, df = 121.98, p-value = 0.3833
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.03263405  0.08432635
## sample estimates:
## mean of x mean of y
## 0.6203077 0.5944615
```

ANSWER: Since the p-value is above 0.05, we do not have enough evidence to reject the null hypothesis that there isn't different between the means of these two values. So there isn't a significant difference between the sad and neutral accuracy of the RG data.

d. Do the above t-tests support claim that there is interaction between emotion and color axis?

ANSWER: The conducted t tests indicates that there isn't significant difference for the outcomes for the discrimination tests between color and emotions. So the emotions did not affect the discrimination between colors.

e. Run factorial ANOVA where color perception accuracy is determined by emotion, color, and interaction

```
all_data <- rbind(BY_data, RG_data)
final <- aov(formula = ACC ~ Axis + Emotion_Condition + Axis:Emotion_Condition, data=all_data)
summary(final)
```

```
##                         Df Sum Sq Mean Sq F value Pr(>F)
## Axis                     1  0.017 0.01745   0.806 0.3703
## Emotion_Condition        1  0.079 0.07893   3.644 0.0574 .
## Axis:Emotion_Condition   1  0.005 0.00526   0.243 0.6224
## Residuals              256  5.545 0.02166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANSWER: In the summary of ANOVA, the *Emotion_Condition* produced a significant f value than the color axis or interaction. However,it's p-value is larger than 5%. If it's smaller than 5%, than we have evidence that the emotions may influence the color discrimination with 95% confidence interval.

```r
layout(matrix(c(1,2,3,4),2,2))
plot(final)
```

```
## hat values (leverages) are all = 0.01538462
##  and there are no factor predictors; no plot no. 5
```