BACS HW (Week9)

106070038

2021-04-25

Question 1 Install the "manipulate" package

i. Would this scenario create systematic or random error (or both or neither)?
ii. Which part of the t-statistic or significance (diff, sd, n, alpha) would be affected?
iii. Will it increase or decrease our power to reject the null hypothesis?
iv. Which kind of error (Type I or Type II) becomes more likely because of this scenario?

```
#install.packages('manipulate')
#library(manipulate)
```

a. You discover that your colleague wanted to target the general population of Taiwanese users of the product. However, he only collected data from a pool of young consumers, and missed many older customers who you suspect might use the product much less every day.

i. Systematic error
ii. diff, sd
iii. decrease
iv. Type I

b. You find that 20 of the respondents are reporting data from the wrong wearable device, so they should be removed from the data. These 20 people are just like the others in every other respect.

i. Random error
ii. n
iii. 1-β decrease, decrease
iv. Type II

c. A very annoying professor visiting your company has criticized your colleague's "95% confidence" criteria, and has suggested relaxing it to just 90%.

i. Neither random error nor systematic error
ii. alpha
iii. It will not effect
iv. Type I

d. Your colleague has measured usage times on five weekdays and taken a daily average. But you feel this will underreport usage for younger people who are very active on weekends, whereas it over-reports usage of older users.

i. Systematic error

ii. diff, sd

iii. $1-\beta$ decrease, decrease

iv. Type II

Question 2

a.

```r
library(readr)
experiment <- read.csv(file = "study2Data.csv", header=TRUE)
# getwd()
# setwd("/Users/weiwei/Desktop/2021Spring_Courses/BACS/HW7")
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))

BY_data_density <- density(BY_data$ACC)
RG_data_density <- density(RG_data$ACC)
plot(BY_data_density, main = "Density plot of BY_ACC and RG_ACC", lwd=2, col = "blue")
lines(RG_data_density, lwd=2, col = "red")
```
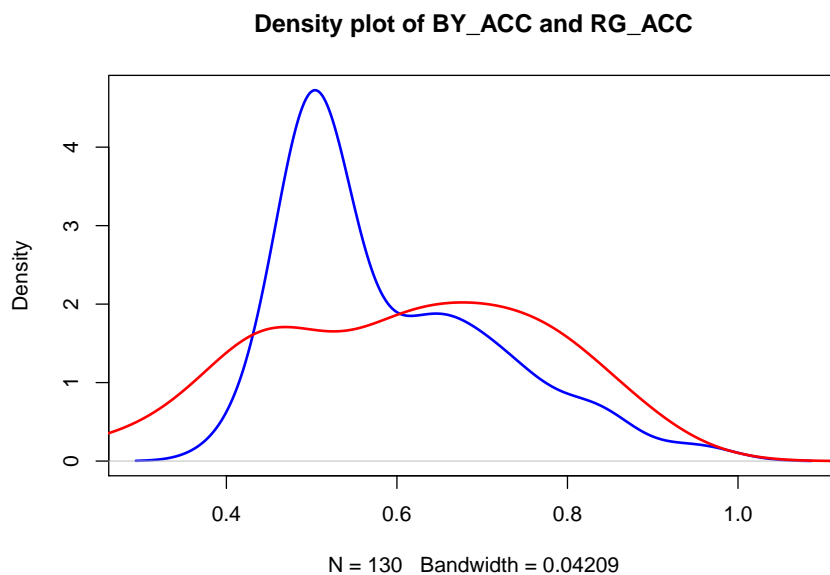


```r
summary(BY_data$ACC)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.420   0.500   0.540   0.591   0.670   0.960
```

```r
summary(RG_data$ACC)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2100  0.4600  0.6300  0.6074  0.7500  0.9600
```

```r
sd(BY_data$ACC)
```

```
## [1] 0.1238144
```

```r
sd(RG_data$ACC)
```

```
## [1] 0.1682585
```

- Answer: According to the plot and descriptive statistics

1. standard deviation of BY_data < standard deviation of RG_data

2. mean of BY_data < mean of RG_data

3. range of BY_data < range of RG_data

4. the distribution of RG_data is more **disperse(separate)** than BY_data

b. blue-yellow accuracy

```r
filter_Sadness_BY <- BY_data[BY_data$Emotion_Condition == "Sadness",]
filter_Neutral_BY <- BY_data[BY_data$Emotion_Condition == "Neutral",]
?t.test
t.test(filter_Sadness_BY$ACC, filter_Neutral_BY$ACC, var.equal = FALSE, alt = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  filter_Sadness_BY$ACC and filter_Neutral_BY$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:
##  -0.086308149 -0.001384159
## sample estimates:
## mean of x mean of y
## 0.5690769 0.6129231
```

- H0: the accuracy of sad and neutral participants are equal.
- H1: the accuracy of sad and neutral participants are not equal.
- p-value = 0.04309 < 0.05(95% confidence)
- Reject H0, there is a significant difference in blue-yellow accuracy between sad and neutral participants.

c. red-green accuracy

```
filter_Sadness_RG <- RG_data[RG_data$Emotion_Condition == "Sadness",]
filter_Neutral_RG <- RG_data[RG_data$Emotion_Condition == "Neutral",]
t.test(filter_Sadness_RG$ACC, filter_Neutral_RG$ACC, var.equal = FALSE, alt = "two.sided")
```

```
##
##  Welch Two Sample t-test
##
## data:  filter_Sadness_RG$ACC and filter_Neutral_RG$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:
##  -0.08432635  0.03263405
## sample estimates:
## mean of x mean of y
## 0.5944615 0.6203077
```

- H0: the accuracy of sad and neutral participants are equal.
- H1: the accuracy of sad and neutral participants are not equal.
- p-value = 0.3833 > 0.05(95% confidence)
- Do not reject H0, there is not a significant difference in red-green
  accuracy between sad and neutral participants.

d. (not graded)Do the above t-tests support a claim that there is an interaction between emotion and color axis?

- Answer: maybe yes, but it depends.
- Based on (b) and (c), blue-yellow accuracy are significant difference
  in sad and neutral participants and this result support the claim.
  However, the red-green accuracy do not.

e. Are any of these three factors (emotion/color-axis/interaction) possibly influencing color perception accuracy at any meaningful level of confidence?

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

```
## Analysis of Variance Table
##
## Response: ACC
##                      Df Sum Sq  Mean Sq F value  Pr(>F)
## Axis                  1 0.0174 0.017450  0.8055 0.37029
## Emotion_Condition     1 0.0789 0.078927  3.6435 0.05741 .
```

```
## Axis:Emotion_Condition   1 0.0053 0.005265   0.2431 0.62243
## Residuals              256 5.5455 0.021662
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Answer: the Signif. code of **Emotion_Condition** is '', which means 0.05 ~ 0.1.