

MATH 588

HW2

Md Ismail Hossain

2/08/2022

Question 1

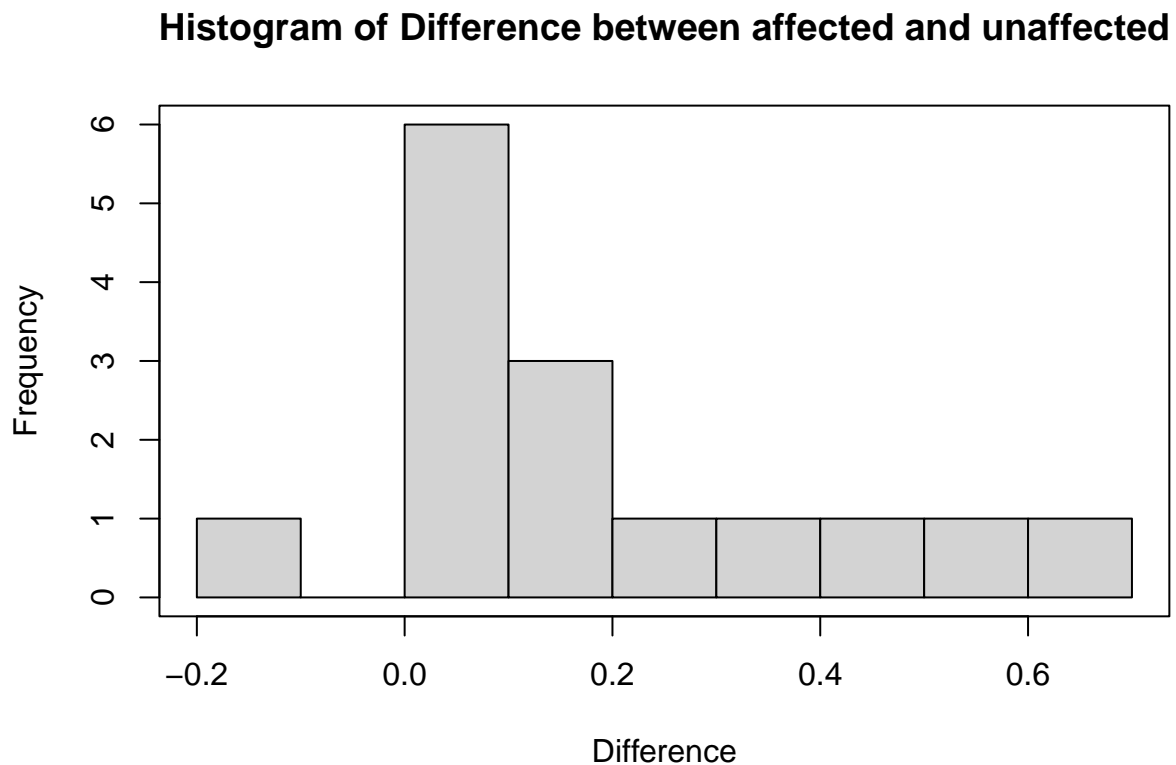
a and b

```
library(Sleuth3)
head(case0202)

##      Unaffected Affected
## 1         1.94      1.27
## 2         1.44      1.63
## 3         1.56      1.47
## 4         1.58      1.39
## 5         2.06      1.93
## 6         1.66      1.26

case0202$Difference <- case0202$Unaffected-case0202$Affected

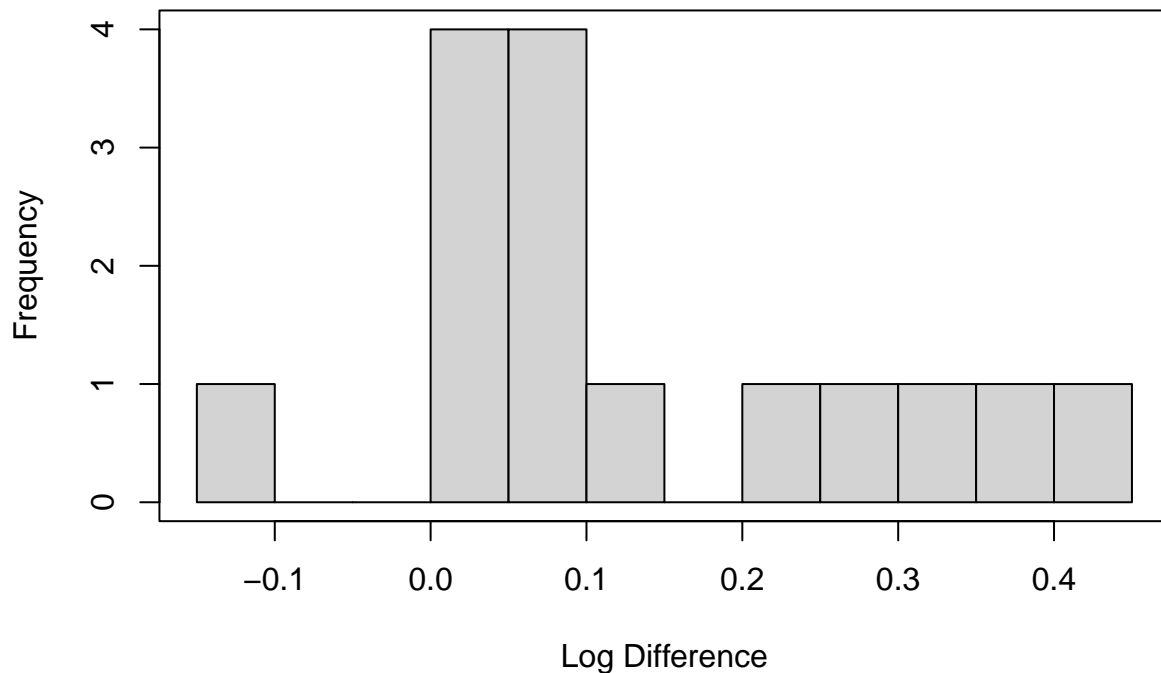
#a)
#par(mfrow=c(2,1))
hist(case0202$Difference,main="Histogram of Difference between affected and unaffected",
      xlab="Difference",breaks = 10)
box()
```



```
#b)
case0202$log_unaff <- log(case0202$Unaffected)
case0202$log_aff <- log(case0202$Affected)
case0202$diff_log <- case0202$log_unaff - case0202$log_aff
```

```
hist(case0202$diff_log,main="Histogram of log transformed Difference between affected and unaffected",
      xlab="Log Difference",breaks = 10)
box()
```

Histogram of log transformed Difference between affected and unaffected



The distribution of differences of log volumes looks more symmetric than the original scale.

c:

```
#c) Paired t-test

# original data
t.test(case0202$Unaffected,case0202$Affected,pair=TRUE)

##
## Paired t-test
##
## data: case0202$Unaffected and case0202$Affected
## t = 3.2289, df = 14, p-value = 0.006062
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.0667041 0.3306292
## sample estimates:
## mean of the differences
##                0.1986667
```

```
# log transformed data
t.test(case0202$log_unaff,case0202$log_aff,pair=TRUE)

##
## Paired t-test
##
## data: case0202$log_unaff and case0202$log_aff
## t = 3.1967, df = 14, p-value = 0.006463
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.04228227 0.21470597
## sample estimates:
## mean of the differences
## 0.1284941
```

We have found that the difference between affected and unaffected groups are statistically significant at 5% level of significance for both transformed and non transformed data sets. But because of normality assumption we should consider the log transformed data in testing the hypothesis.

d:

```
# d) 95% CI

# log-transformed 95% confidence interval of the log difference values.
t.test(case0202$diff_log)$conf.int

## [1] 0.04228227 0.21470597
## attr(,"conf.level")
## [1] 0.95

# 95% CI in original scale (differences)
exp(t.test(case0202$diff_log)$conf.int)

## [1] 1.043189 1.239497
## attr(,"conf.level")
## [1] 0.95
```

Question 2

a and b

```
# Q2

set.seed(123)

output_2_a <- vector()
for (i in 1:1000) {
  x=rep(c(0,9), each=15)
  y = 120 + 0.5*x + rnorm(30,0,8)
  p_value = coef(summary(lm(y~x)))[2,"Pr(>|t|)"]
  output_2_a[i] = p_value<=0.05
}
```

```
cat("(a) Power for half the plants at each of two fertilizer levels of 0 and 9 is ",mean(output_2_a),"\n")

## (a) Power for half the plants at each of two fertilizer levels of 0 and 9 is 0.307

output_2_b <- vector()
for (i in 1:1000) {
  x=rep(0:9, each=3)
  y = 120 + 0.5*x + rnorm(30,0,8)
  p_value = coef(summary(lm(y~x)))[2,"Pr(>|t|)"]
  output_2_b[i] = p_value<=0.05
}

cat("(b) Power for 3 plants each at fertilizer levels 0 through 9 at each whole number is",mean(output_2_b),"\n")
```

```
## (b) Power for 3 plants each at fertilizer levels 0 through 9 at each whole number is 0.162
```

What advantage is there to using all ten levels of fertilizer?

Although the power of the study reduced when we are using ten levels of the fertilizers, there is a advantage in the analysis of using too many levels. By using only two fertilizer levels we could not understand the residual vs fit plot, because using only two level gives us a linear line. On the other hand if there is any non linear relationship then we could understand using the residual vs fit plot for all ten levels of fertilizer.

Question 3:

Null Hypothesis: Strength in left and right arm is similar for ambidextrous.

Alternative Hypothesis: Strength are not equal in left and right arm for ambidextrous .

Let's find the probability of getting a strong on right hand person out of 52 individual and it's $20/52 = 0.38$ approximately.

Now if we consider the normal approximation of the binomial then we need to consider two different types and for our case it's strong right hand person and strong left hand person. So, total number of person here is $n = 47$ and if we consider that selection of right hand person is a success then the probability could be considered as $p = 0.38$.

So, the $mean = np = 47(0.38) = 17.86$ and $variance = npq = 47(.38)(0.62) = 11.07$. The calculated z-statistic could be $Z = (20 - 17.86)/\sqrt{11.07} = 0.643$

```
2*pnorm(q=0.643, lower.tail=F)
```

```
## [1] 0.5202241
```

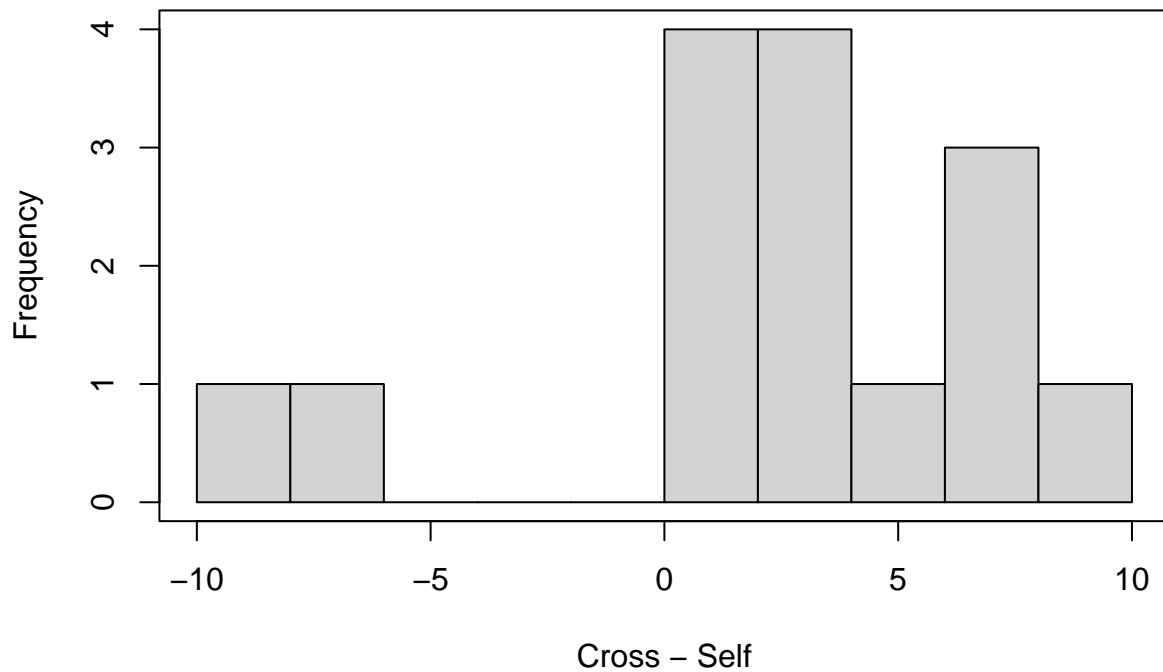
At 5% level of significance we could not reject the null hypothesis. So, we do not have enough data to show that there is not statistically significant difference of strength between left and right arm of the ambidextrous persons or strength is equal in both hand.

Question 4

a

```
with(ex0428, hist(Cross-Self, breaks=10))
box()
```

Histogram of Cross – Self



b

```
with(ex0428, t.test(Cross,Self,paired=T))
```

```
##  
## Paired t-test  
##  
## data: Cross and Self  
## t = 2.1474, df = 14, p-value = 0.04976  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.003165044 5.230168289  
## sample estimates:  
## mean of the differences  
## 2.616667
```

The p-value 0.04976, looks smaller than 5% level of significance. So, there should be statistically significant difference between Cross and Self group.

c

95% confidence interval is = [0.003, 5.230]

d

The histogram of the difference between Cross and self does not look normally distributed. So, before applying t-test need to transform the data or should apply nonparametric test.

e

```
with(ex0428, wilcox.test(Cross,Self,paired=T))
```

```
##
```

```
## Wilcoxon signed rank exact test
```

```
##
```

```
## data: Cross and Self
```

```
## V = 96, p-value = 0.04126
```

```
## alternative hypothesis: true location shift is not equal to 0
```

The p-value is 0.041, which is less than 0.05. So, at 5% level of significance we can conclude that there is significant difference between Cross fertilized plants and Self fertilized plants in-terms of their heights.