Homework 1

Md Muhtasim Billah, Jugal Marfatia, Namrata Ray

9/17/2019

Question 1.

```
df = read.csv("hw1.csv")

df$r = (df$PM1 - df$MM1)/(df$PM1+ df$MM1)

df$z = abs(df$r- 0.15)

df$rank = rank(df$r)

df$pos_or_not = ifelse(df$r >= 0, 1 , 0)

print(df)
```

```
##        MM1     PM1          r         z rank pos_or_not
## 1  2916.22 40154.0  0.8645830 0.7145830   13          1
## 2  6069.49  3356.9 -0.2877655 0.4377655    1          0
## 3  1089.75 10972.0  0.8193048 0.6693048   11          1
## 4  4705.19 10380.1  0.3761883 0.2261883    4          1
## 5   956.33  9379.7  0.8149522 0.6649522   10          1
## 6  3179.57  9231.1  0.4876070 0.3376070    5          1
## 7   214.22 49444.2  0.9913723 0.8413723   16          1
## 8  1555.74 35269.3  0.9155064 0.7655064   15          1
## 9   666.33  5727.2  0.7915612 0.6415612    9          1
## 10 3689.52 50360.0  0.8634763 0.7134763   12          1
## 11 2296.42  1531.5 -0.1998265 0.3498265    3          0
## 12 9452.72  5598.9 -0.2560402 0.4060402    2          0
## 13 4263.36 63598.8  0.8743524 0.7243524   14          1
## 14 3310.51 13614.1  0.6087933 0.4587933    6          1
## 15 1765.97  8586.3  0.6588246 0.5088246    7          1
## 16 1840.13 15106.5  0.7828323 0.6328323    8          1
```

The W is:

```
w = sum(df$rank*df$pos_or_not)
print(w)
```

```
## [1] 130
```

Conducting the Wilcoxon Test.

```r
E = length(df) *(length(df)+ 1)/ 4
SE = sqrt(length(df) *(length(df)+ 1) * (2*length(df)+ 1)/24)

Z_Statstic = (w-E)/ SE
print(Z_Statstic)
```

```
## [1] 25.05401
```

The P-Value is:

```r
2*pnorm(-abs(Z_Statstic))
```

```
## [1] 1.578828e-138
```

Hypothesis testing:

Null-Hypothesis: Mean difference $= 0$
Alternate-Hypothesis: Mean difference $\neq 0$
P-value ≤ 0.05
Therefore, reject the null. (Keeping $\alpha = 0.05$)

Checking with the built-in Pacakge.

```r
wilcox.test(df$PM1, df$MM1,paired=TRUE)
```

```
##
##  Wilcoxon signed rank test
##
## data:  df$PM1 and df$MM1
## V = 130, p-value = 0.0004272
## alternative hypothesis: true location shift is not equal to 0
```

###Tukey Biweight method

```r
y= log2(df$PM1) - log2(df$MM1)

md<-median(y)

z<-abs(y-md)

mz<-median(z)

u<-(y-md)/(5*mz+.0001)
w<-ifelse(abs(u)<=1,(1-u^2)^2,0)
table.data=cbind(y,md,z,mz,u,w)
tuk_avg=sum(y*w)/sum(w)
tuk_avg
```

```
## [1] 2.921347
```

Question 2.

###a.

```
summary(df$MM1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   214.2  1439.2  2606.3  2998.2  3833.0  9452.7
```

```
summary(df$PM1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1532    7872   10676   20769   36490   63599
```

###b.

```
mean(df$PM1, na.rm = FALSE)
```

```
## [1] 20769.41
```

```
sd(df$PM1, na.rm = FALSE)
```

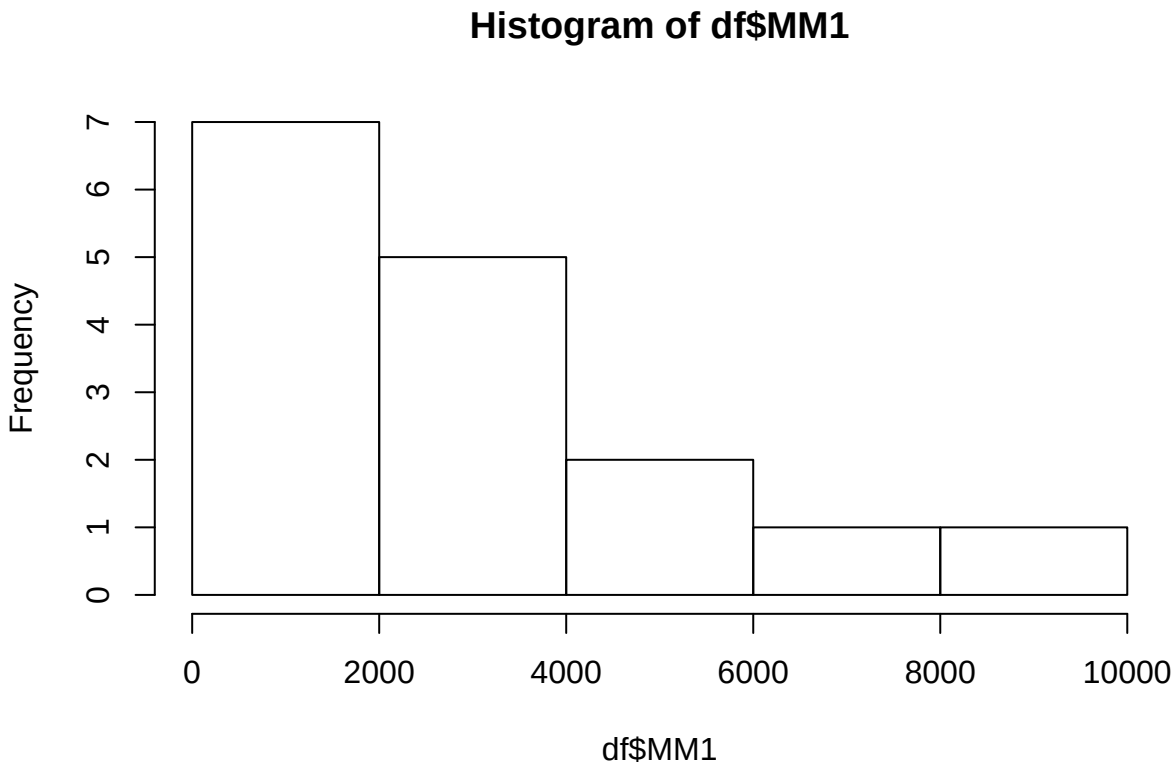
```
## [1] 19910.46
```

```
IQR(df$PM1, na.rm = FALSE)
```

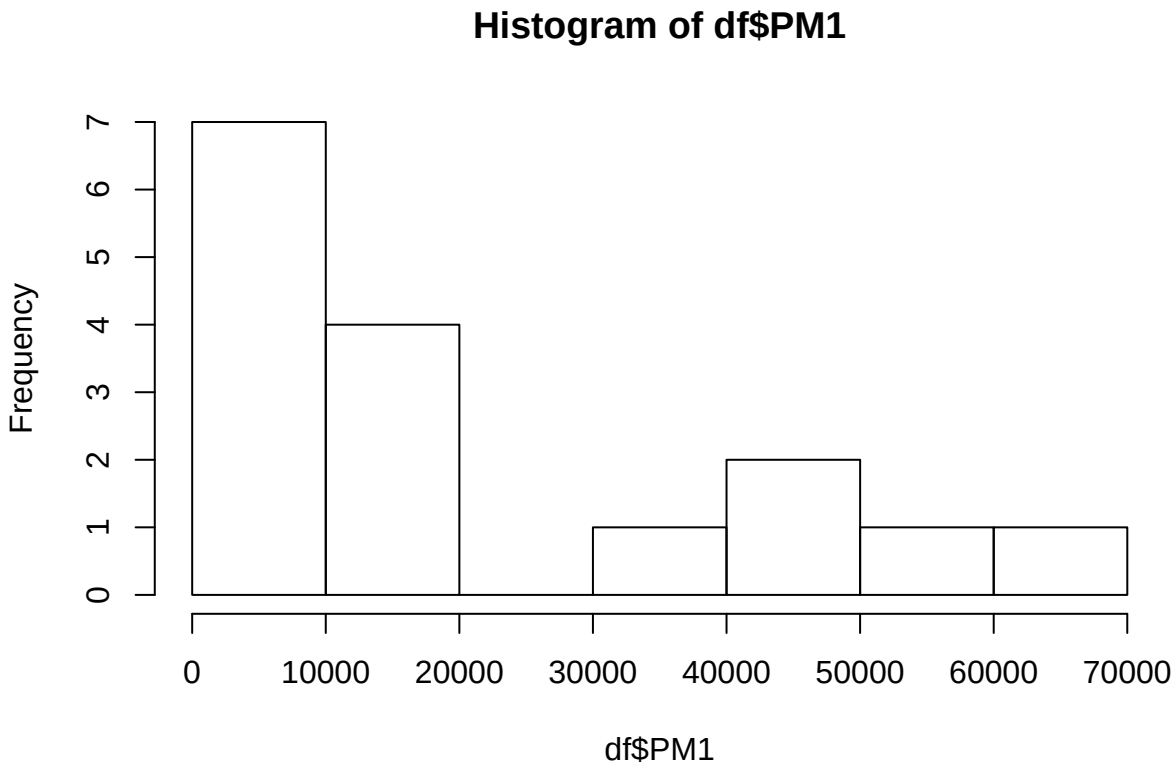
```
## [1] 28618.95
```

###c.

```
hist(df$MM1)
```


Histogram of df$MM1
Frequency
0
1
2
3
4
5
6
7
0
2000
4000
6000
8000
10000
df$MM1

```
hist(df$PM1)
```


Histogram of df$PM1
Frequency
0
1
2
3
4
5
6
7
0
10000
20000
30000
40000
50000
60000
70000
df$PM1

###

d.

```r
plot(df$PM1~df$MM1, data=df)
```