Math189 HW3

Group Members:

Yaqi Chen (PID: A15742547; Section: A03)

Yuetong Lyu (PID: A13779993; Section: A04)

Siyi He (PID: A13400569; Section: A03)

Zhenyuan Xu (PID: A92067995; Section: A02)

Jiawei Chao (PID: A13818001; Section: A02)

Problem 1

Part a

```
Water <- read.table("water.txt", header = TRUE)
t.test(Water$bottom, Water$surface, aternative = "t", paired = T)
```

```
t.test(Water$bottom, Water$surface, aternative = "t", paired = T)

  Paired t-test

data:  Water$bottom and Water$surface
t = 4.8638, df = 9, p-value = 0.0008911
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.043006 0.117794
sample estimates:
mean of the differences
                 0.0804
```

Here we treated the dataset as a paired sample and got a small p-value that is fair enough to reject our null hypothesis. Thus, we conclude that the mean zinc concentration in bottom water is the same as that in surface.

Part b

The independence of the two samples can influence our t-test result. If two groups are dependent, we must use a paired t-test instead of an independent two sample t-test. Otherwise, our answer would be wrong since there are two different formulas to get the standard error in our t-test. If two samples are related, the errors will be related as well. They will form different t-distribution compared to the situation when these two samples are independent. Violation of this assumption will result in inaccurate p-value.

Problem 2

Part a

Setup

```
install.packages('HSAUR3')

library("HSAUR3")

data("pottery")

pottery

site1 <- subset(pottery, kiln == 1)

site2 <- subset(pottery, kiln == 2)

site4 <- subset(pottery, kiln == 4)

site5 <- subset(pottery, kiln == 5)

len1 <- length(site1$Al2O3)

len2 <- length(site2$Al2O3)

len2

len4 <- length(site4$Al2O3)

len5 <- length(site5$Al2O3)

len <- c(len1,len2,0,len4,len5)

c <- c(1,2,4,5)

mean(pottery[,1])
```

```
mean(subset(pottery, kiln == 1)[,1])

mean(subset(pottery, kiln == 5)[,1])

chemicals <- c('Al2O3', 'Fe2O3', 'Mgo', 'CaO', 'Na2O', 'K2O', 'TiO2', 'MnO',
'BaO')
```

#f stat

```
Fstatistic <- c()
for (k in 1:9){

  error_sum <- 0

  treat_sum <- 0

  whole_mean <- mean(pottery[,k])

  for (i in c){

    mean <- mean(subset(pottery, kiln == i)[,k])

    n <- len[i]

    treat_sum <- treat_sum + n * ((mean-whole_mean)^2)

    subpottery <- subset(pottery, kiln == i)

    for (j in 1:n){

      difference <- (subpottery[j,k] - mean)^2

      error_sum <- error_sum + difference

    }

  }

  treat <- treat_sum / 3

  error <- error_sum / 39

  \#print (paste(treat,error))

  ans = treat/error
```

```
  print (paste(chemicals[k], ans))

  Fstatistic <- c(Fstatistic, ans)

}
```

```
[1] "Al2O3 26.2034498265741"
[1] "Fe2O3 154.326776387811"
[1] "Mgo 97.9157560767963"
[1] "CaO 53.5350307231462"
[1] "Na2O 10.5266992234656"
[1] "K2O 82.3564640243008"
[1] "TiO2 14.8485513138411"
[1] "MnO 52.8034239690174"
[1] "BaO 0.480014498907605"
```

Part b

#Bonferroni

```
lower <- qf(0.05/9,3,39) #0.02537717
upper <- qf(1-0.05/9,3,39) #4.892638
for (i in 1:9){
  if ((Fstatistic[i] < upper) & (Fstatistic[i] > lower)){
      print(paste('We fail to reject the null hypothesis for', chemicals[i]))
    }
  else{print(paste('We reject the null hypothesis for', chemicals[i]))}
}
```

```
[1] "We reject the null hypothesis for Al2O3"
[1] "We reject the null hypothesis for Fe2O3"
[1] "We reject the null hypothesis for Mgo"
[1] "We reject the null hypothesis for CaO"
[1] "We reject the null hypothesis for Na2O"
[1] "We reject the null hypothesis for K2O"
[1] "We reject the null hypothesis for TiO2"
[1] "We reject the null hypothesis for MnO"
[1] "We fail to reject the null hypothesis for BaO"
```

Given only the F statistics of BaO is between 0.0254 and 4.893, we will still reject Ho and conclude that the first eight chemical concentrations are different among the four sites.

Part c

#BH

```
sort_F=sort(Fstatistic,decreasing=T)
rejection<-numeric(9)
for(i in 1:9){
  lower = qf(i*0.05/9,3,39) #0.1163315
  upper=qf(1-i*0.05/9,3,39) #2.845068
  if((sort_F[i]>upper) & (sort_F[i]<lower)){
    print(paste('We reject the null hypothesis for', chemicals[i]))
    }else{
    print(paste('We fail to reject the null hypothesis for', chemicals[i]))
    }
}
```

```
[1] "We reject the null hypothesis for Al2O3"
[1] "We reject the null hypothesis for Fe2O3"
[1] "We reject the null hypothesis for Mgo"
[1] "We reject the null hypothesis for CaO"
[1] "We reject the null hypothesis for Na2O"
[1] "We reject the null hypothesis for K2O"
[1] "We reject the null hypothesis for TiO2"
[1] "We reject the null hypothesis for MnO"
[1] "We fail to reject the null hypothesis for BaO"
```

Still only the F statistics of BaO lies outside the critical values. Therefore, we reach the same conclusion as Bonferroni one. We reject the first eight hypothesis and Ho.