HW 3

Q1.1

```r
data = read.delim("multiple.txt", sep = " ",header=FALSE)
```

```r
sample_mean = colMeans(data)
sample_sd = sapply(data, sd)
hnull = hnull = integer(50)

n = nrow(data)
t_statistic = (sample_mean - hnull)/(sample_sd/sqrt(n))

pval = 2*apply(rbind(pt(t_statistic, df = n-1, lower.tail = F),
                     pt(t_statistic, df = n-1, lower.tail = T)), 2, min)
rejection_p <- pval < 0.1
print(pval)
```

```
##           V1           V2           V3           V4           V5           V6
## 7.477468e-34 1.030644e-32 1.613953e-33 5.107798e-37 4.299681e-35 4.181053e-33
##           V7           V8           V9          V10          V11          V12
## 1.782424e-39 2.281744e-34 4.767758e-34 9.020768e-42 8.501800e-01 3.387655e-01
##          V13          V14          V15          V16          V17          V18
## 8.433591e-01 2.353081e-01 8.346443e-01 3.410084e-01 4.495285e-01 8.988573e-01
##          V19          V20          V21          V22          V23          V24
## 4.855360e-01 8.298508e-02 7.435778e-01 4.344605e-01 5.521855e-01 4.775194e-01
##          V25          V26          V27          V28          V29          V30
## 7.573900e-01 1.354197e-01 9.927262e-01 2.944327e-01 6.783256e-01 1.898939e-01
##          V31          V32          V33          V34          V35          V36
## 4.716828e-01 9.302681e-02 4.736660e-01 9.189993e-01 2.138515e-01 6.519921e-01
##          V37          V38          V39          V40          V41          V42
## 3.408716e-01 2.211123e-01 8.806238e-01 7.315268e-02 7.171567e-01 6.472996e-01
##          V43          V44          V45          V46          V47          V48
## 6.915055e-01 2.475067e-02 2.191321e-01 7.678119e-01 3.489369e-01 1.367531e-01
##          V49          V50
## 8.886784e-01 7.316609e-01
```

```r
print(rejection_p)
```

```
##    V1    V2    V3    V4    V5    V6    V7    V8    V9   V10   V11   V12   V13
##  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE
##   V14   V15   V16   V17   V18   V19   V20   V21   V22   V23   V24   V25   V26
## FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
##   V27   V28   V29   V30   V31   V32   V33   V34   V35   V36   V37   V38   V39
## FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   V40   V41   V42   V43   V44   V45   V46   V47   V48   V49   V50
##  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
```

Q1.2

```
alpha = 0.1
FWER = 1 - (1 - alpha)^50
R = sum(rejection_p)
V = sum(rejection_p[11:50])
print(FWER)
```

```
## [1] 0.9948462
```

```
print(V/R)
```

```
## [1] 0.2857143
```

FWER: 0.9948462 FDP: 0.2857143

Q1.3

```
pval.bon = p.adjust(pval, method = "bonferroni")
rejection_bon = pval.bon < 0.1
print(rejection_bon)
```

```
##    V1    V2    V3    V4    V5    V6    V7    V8    V9   V10   V11   V12   V13
##  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE FALSE
##   V14   V15   V16   V17   V18   V19   V20   V21   V22   V23   V24   V25   V26
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   V27   V28   V29   V30   V31   V32   V33   V34   V35   V36   V37   V38   V39
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
##   V40   V41   V42   V43   V44   V45   V46   V47   V48   V49   V50
## FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
FWER.bon = 1 - (1 - alpha/50)^50
print(FWER.bon)
```

```
## [1] 0.09525318
```

FWER: 0.09525318

Q1.4

```
pval.bh = p.adjust(pval, method = "BH")
rejection_bh = pval.bh < 0.1
alphas = 0.1 * seq(1/50, 1, by = 1/50)
FWER.bh = 1 - prod(1 - alphas)
FDP_bh = sum(rejection_bh[11:50])/sum(rejection_bh)
print(FWER.bh)
```

```
## [1] 0.928672
```

```
print(FDP_bh)
```

```
## [1] 0
```

BH's FWER is slightly lower than step1 and way higher than step2 but BH's FDP is 0, which is way lower than step1 and 3.

Q2

The null hypothesis is that the population means of all six variables are the same for the genuine and counterfeit bank notes. The alternative hypothesis is that they are different. We can assume that they have the same population variance-covariance matrices.

```
library(glue)
banknotes <- read.delim("SwissBankNotes.txt", sep = "")
df <- banknotes
genuine <- df[0:100,]
counterfeit <- df[101:200,]
n <- 100
delta <- colMeans(genuine) - colMeans(counterfeit)
p <- 6
Sg <- cov(genuine)
Sc <- cov(counterfeit)
S_pooled <- ((n-1)*Sg + (n-1)*Sc) / (n+n-2)
t_squared <- (n*n)/(n+n) * t(delta) %*% solve(S_pooled) %*% (delta)
f_statistic <- t_squared * (n+n-p-1)/(p*(n+n-2))
cv <- qf(0.05, p, n+n-p-1, lower.tail = FALSE)
print(glue("The F statistic is {f_statistic} and the critical value is {cv}"))
```

```
## The F statistic is 391.921702277771 and the critical value is 2.14580146767029
```

Since the F statistic is larger than the critical value, we are able to reject the null hypothesis. We have evidence that the counterfeits are distinguishable from the genuine bank notes on at least one of the variables.