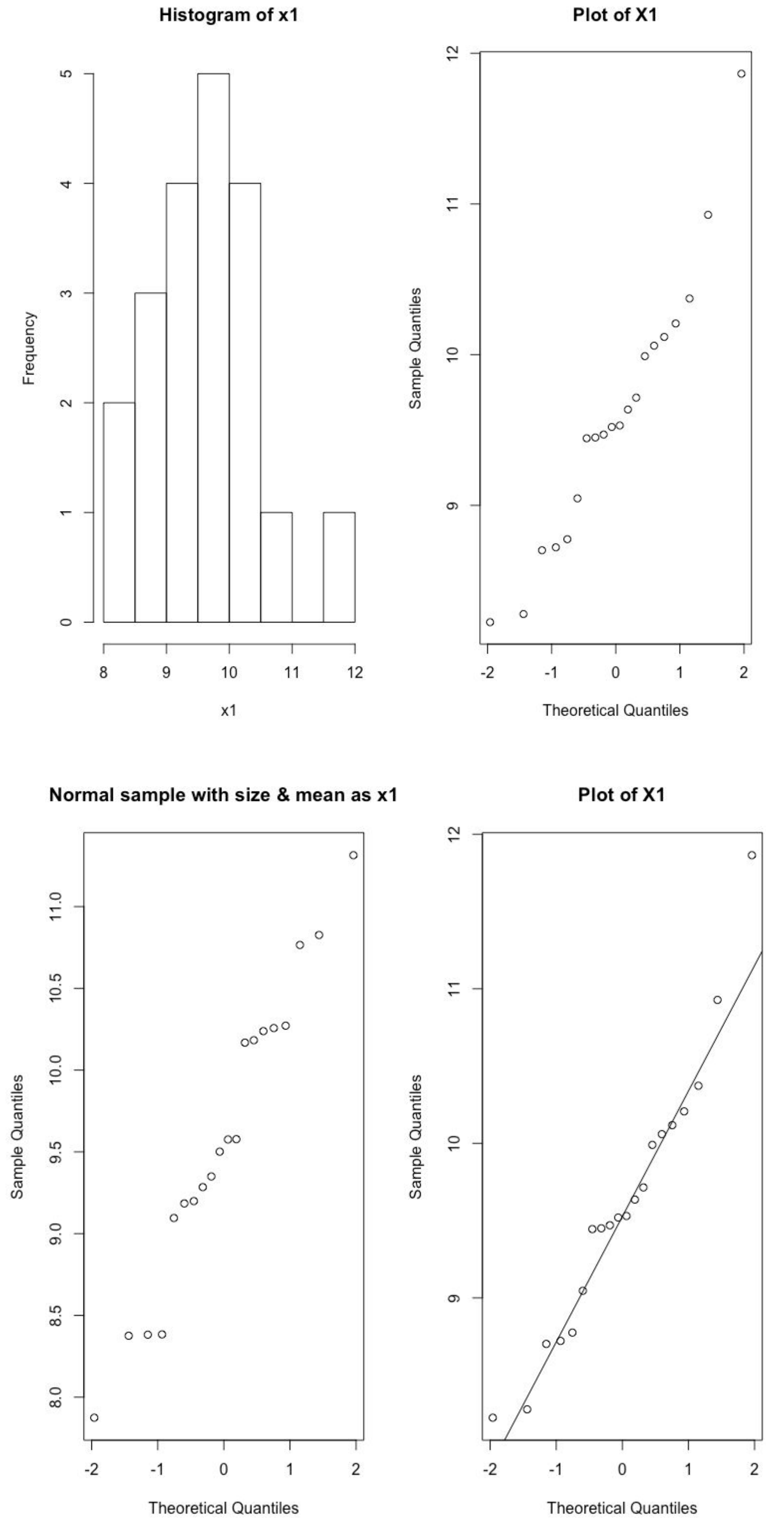
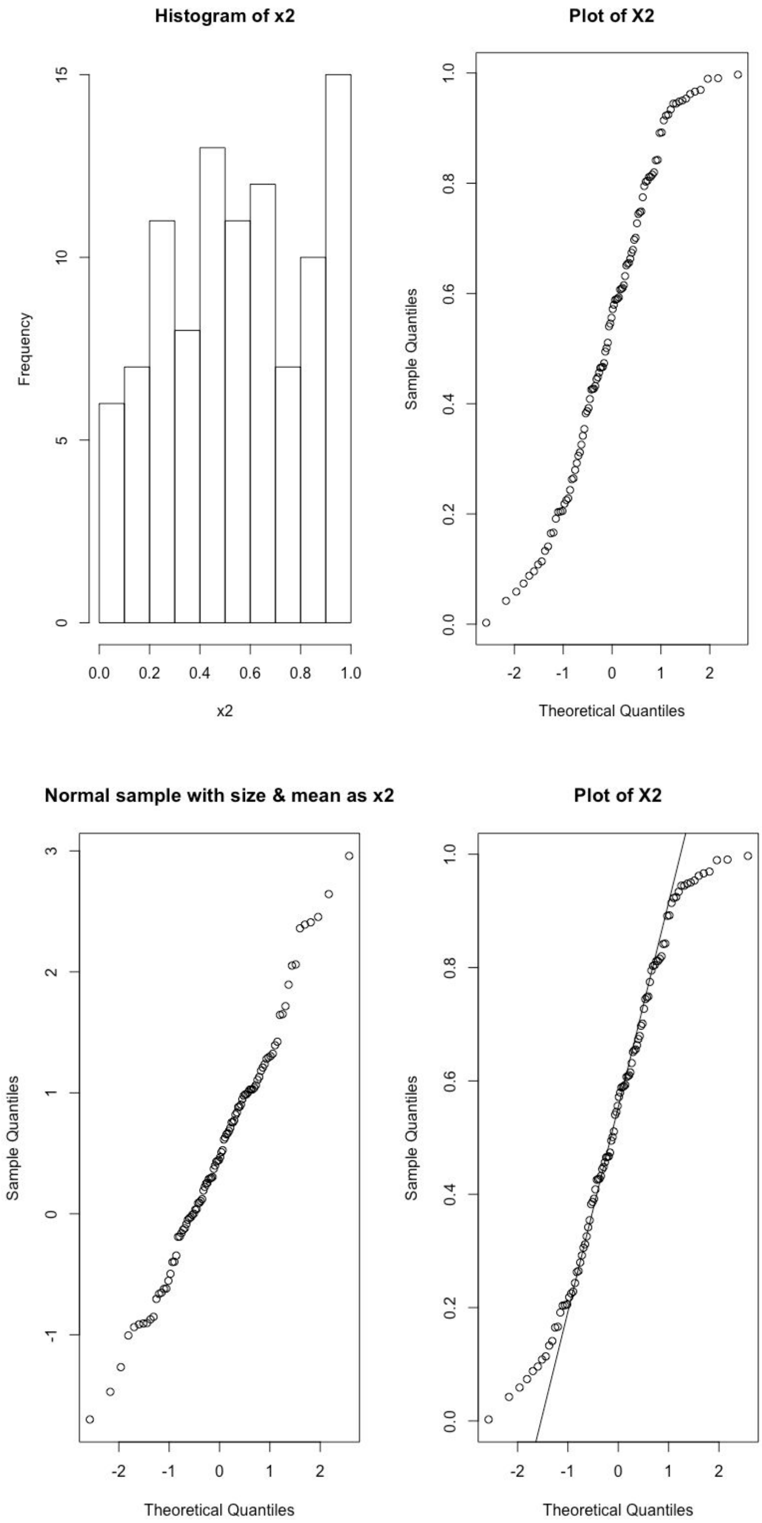
14 February 2018

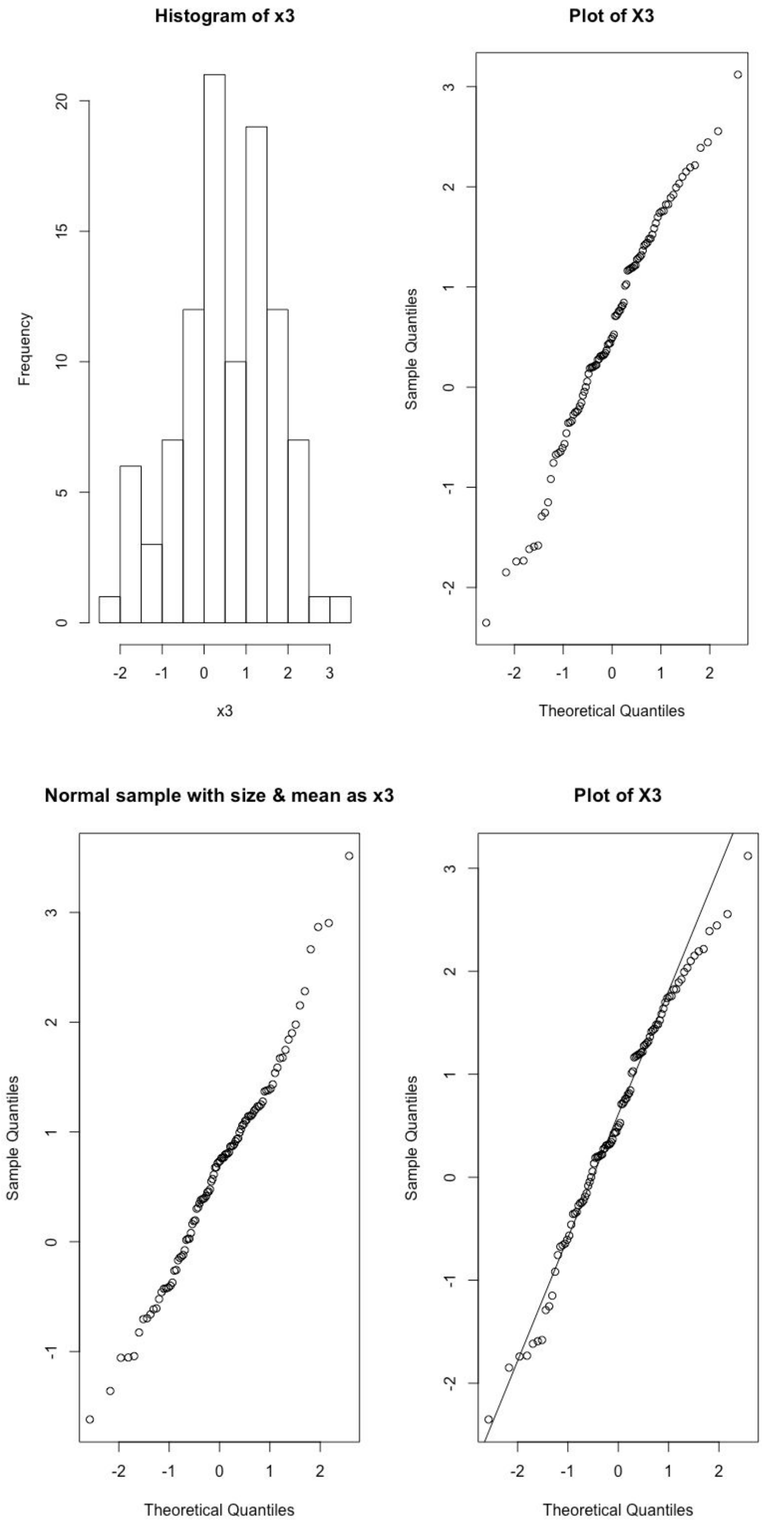
Assignment 1

Chao Zhang & Ibrahim Kanj
Group 13

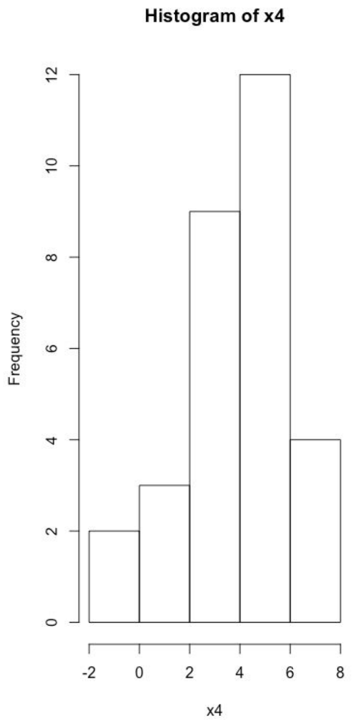
Exercise 1)

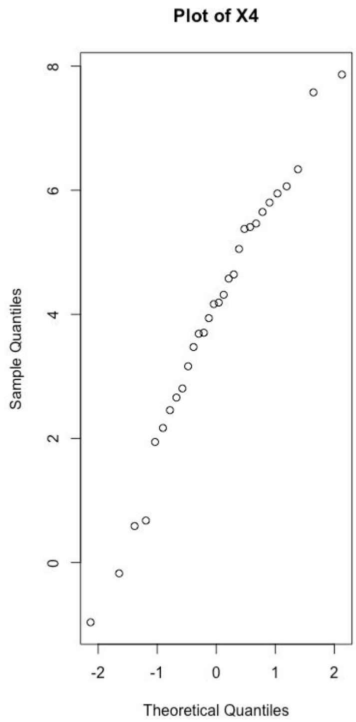


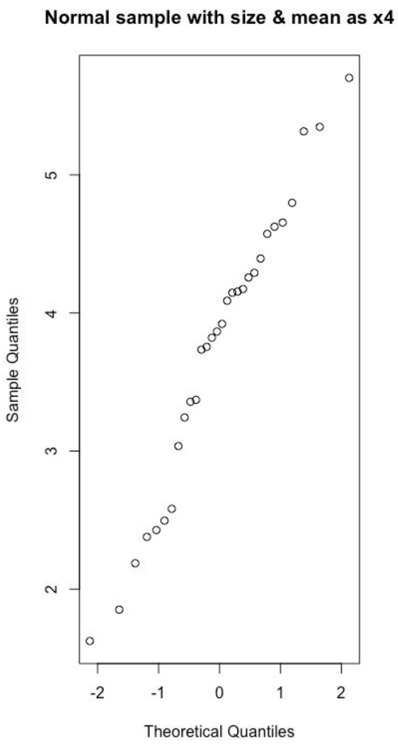

Histogram of x2
Frequency
x2
0
5
10
15
0.0
0.2
0.4
0.6
0.8
1.0
Plot of X2
Sample Quantiles
Theoretical Quantiles
Normal sample with size & mean as x2
Sample Quantiles
Theoretical Quantiles
Plot of X2
Sample Quantiles
Theoretical Quantiles

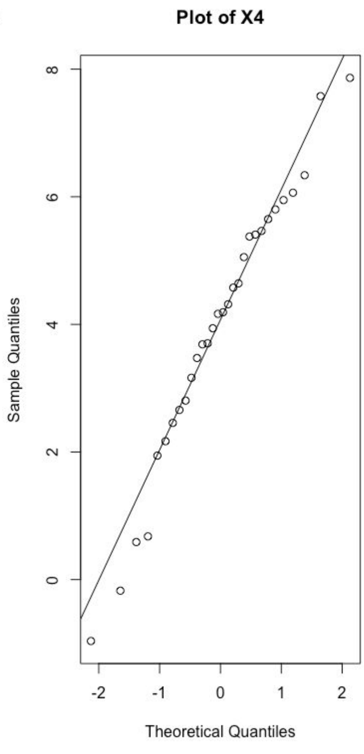


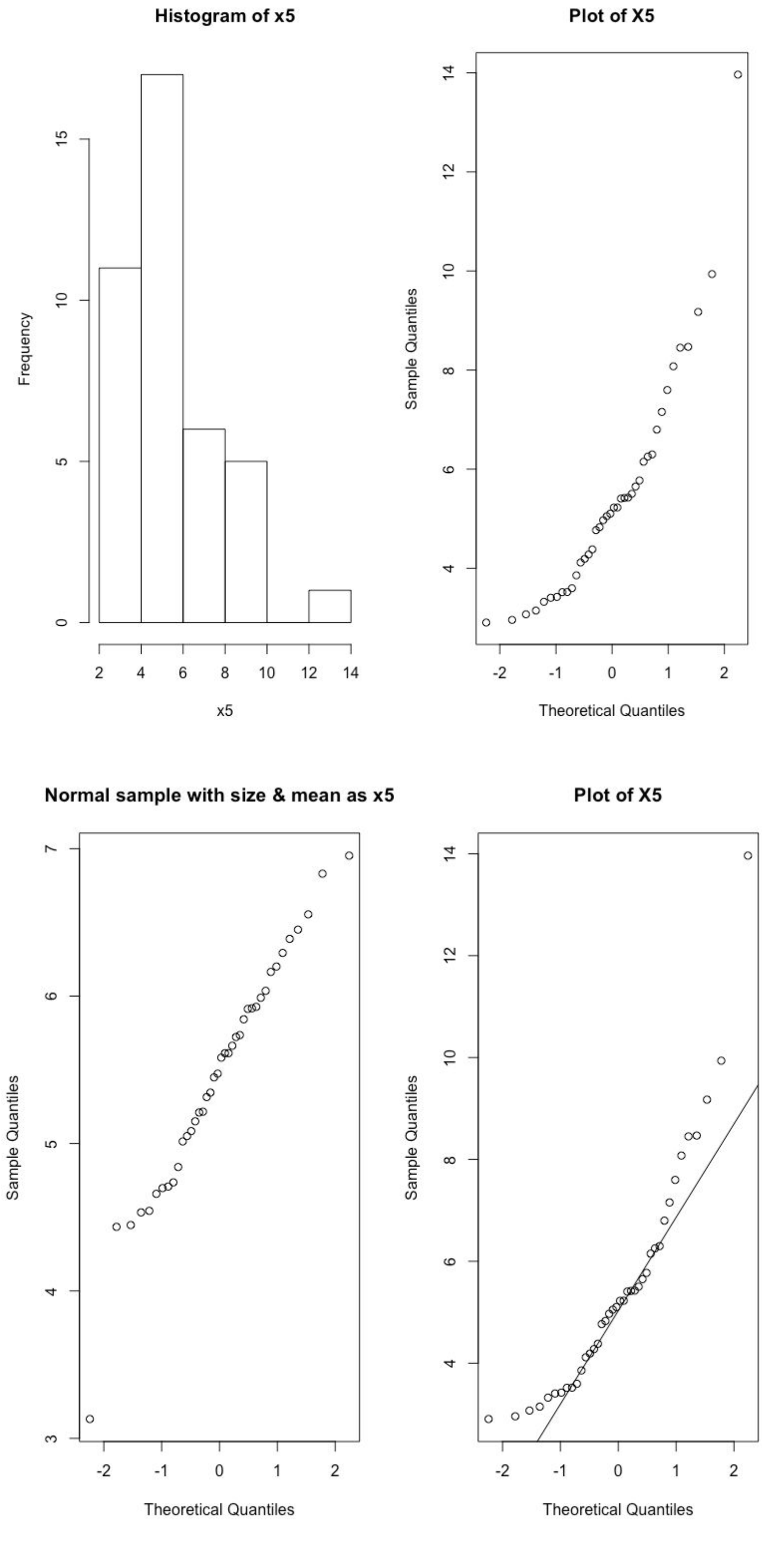
Histogram of x3
Frequency
x3
Plot of X3
Sample Quantiles
Theoretical Quantiles
Normal sample with size & mean as x3
Sample Quantiles
Theoretical Quantiles
Plot of X3
Sample Quantiles
Theoretical Quantiles


Histogram of x4
Frequency
x4
0
2
4
6
8
10
12
-2
0
2
4
6
8


Plot of X4
Sample Quantiles
Theoretical Quantiles
0
2
4
6
8
-2
-1
0
1
2


Normal sample with size & mean as x4
Sample Quantiles
Theoretical Quantiles
2
3
4
5
-2
-1
0
1
2


Plot of X4
Sample Quantiles
Theoretical Quantiles
0
2
4
6
8
-2
-1
0
1
2



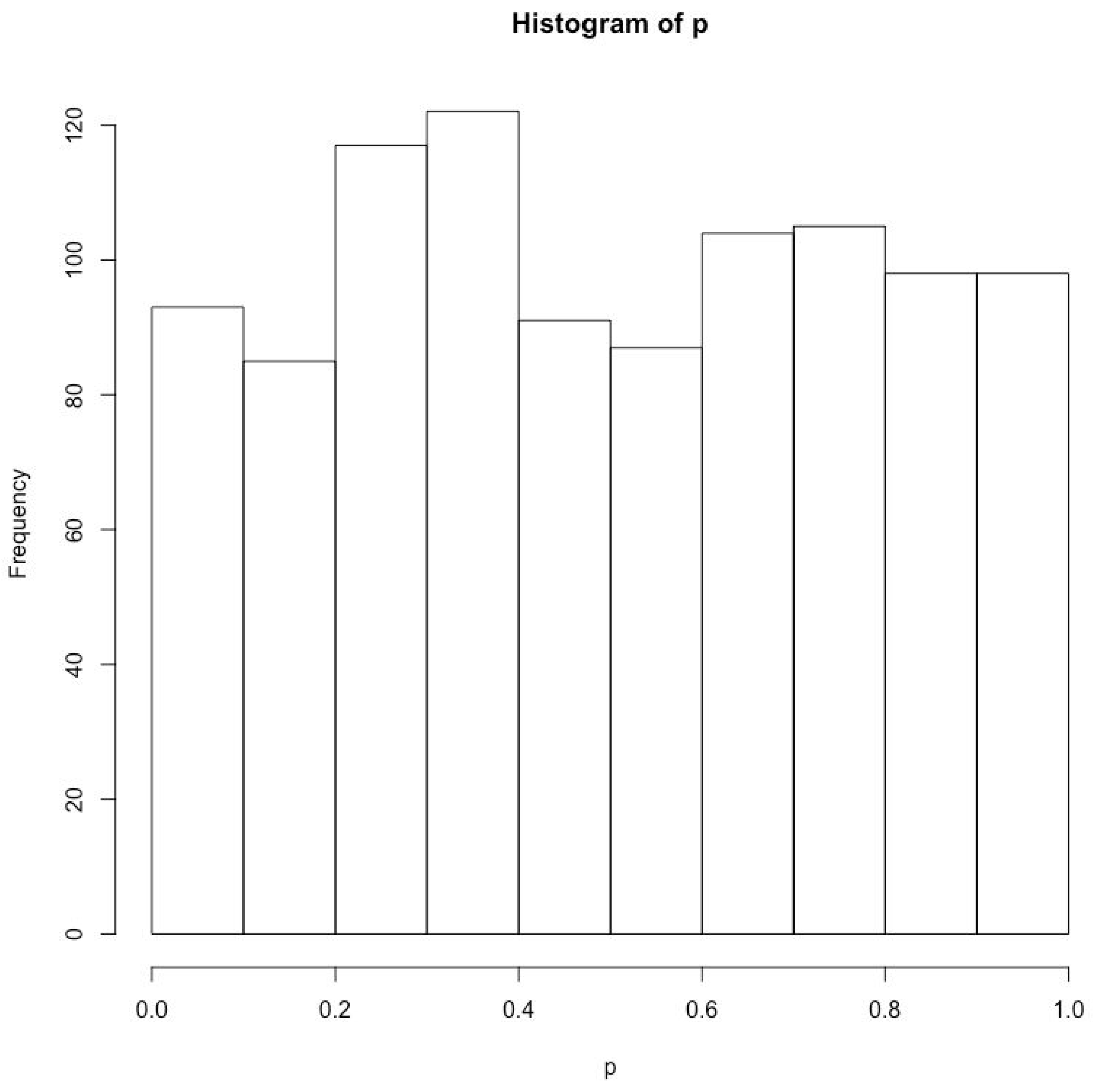
- For each data vector (x1….x5) we are displaying its histogram and its plot. We then compare this plot to a normal sample having the same size and mean of the vector to see whether the points could have been sampled of a normal distribution..

From the above histograms and plots, we were able to see that x2 and x5 are not normal where there is a lot of difference between the normal sample plot and the actual plot. As for the other data vectors, looking at their histograms they appear to be normal.

Exercise 2)

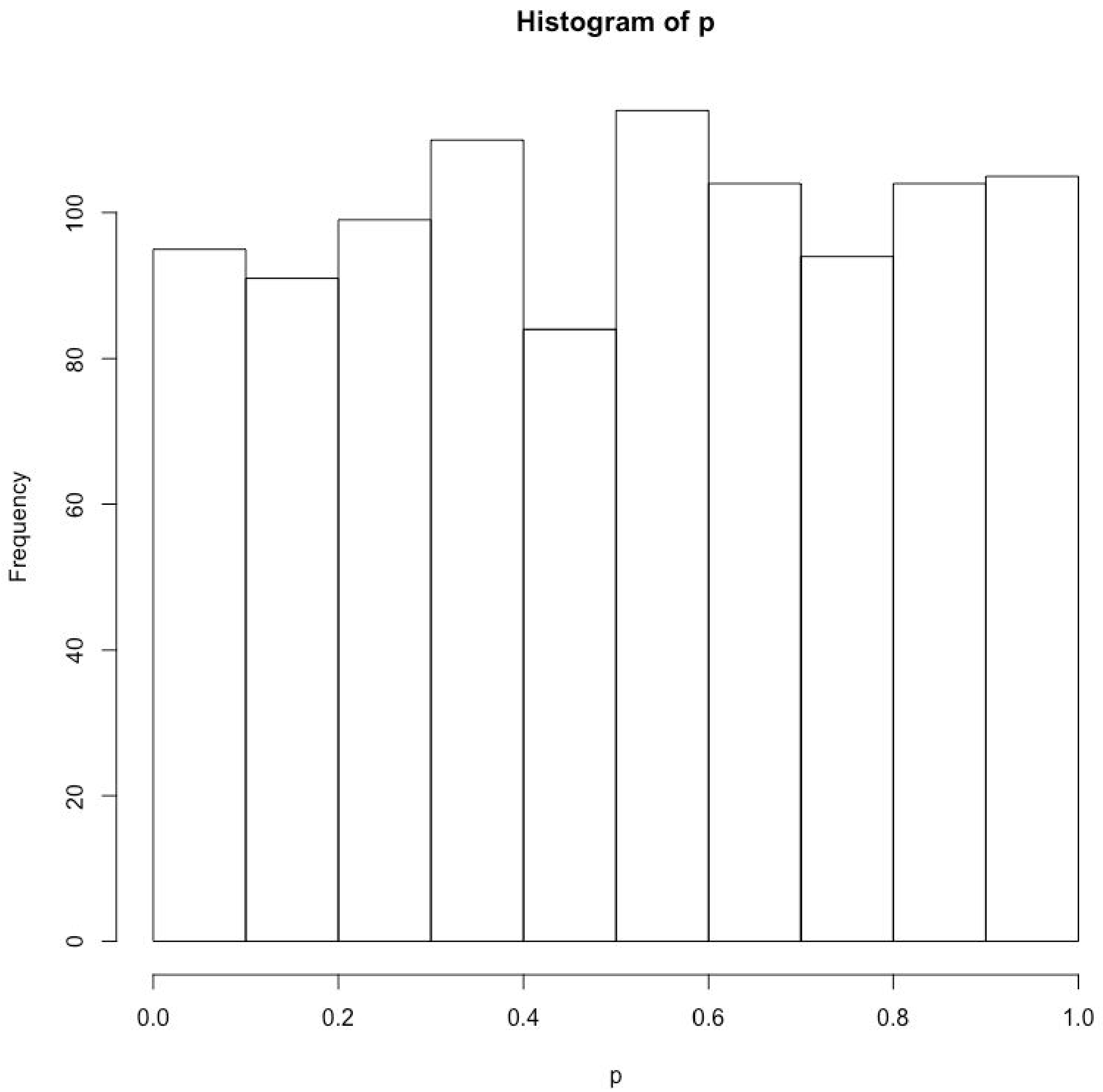
1)

Conditions: mu = nu = 180, m = n = 30 and sd = 10



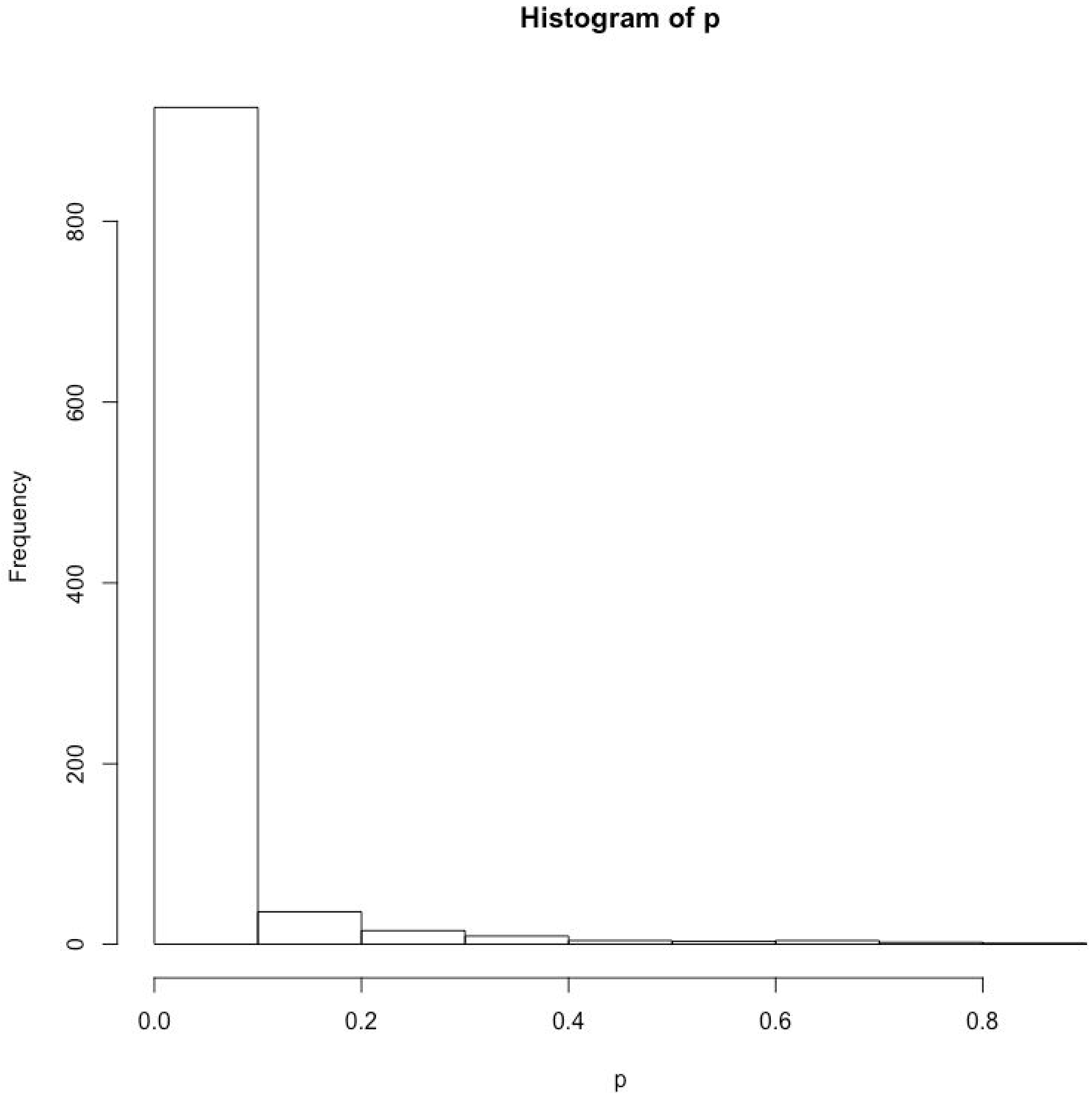
The frequency of p values smaller than 5 percent is 5.3% (53 out of 1000)
The frequency of p values smaller than 10 percent is 9.4% (94 out of 1000)
From the histogram we could see that there is no big of a difference between the frequencies of the different p-values where the frequencies are between 82 and and 122.

2)
Conditions: mu = nu = 180, m = n = 30 and sd = 1



The frequency of p values smaller than 5 percent is 5% (50 out of 1000)
The frequency of p values smaller than 10 percent is 9.6% (96 out of 1000)
Similar to the above histogram, we can see that the frequency of the p-values is also in the range of 82 to 120.

3)
Conditions: mu = 180, nu = 175, m = n = 30 and sd = 6

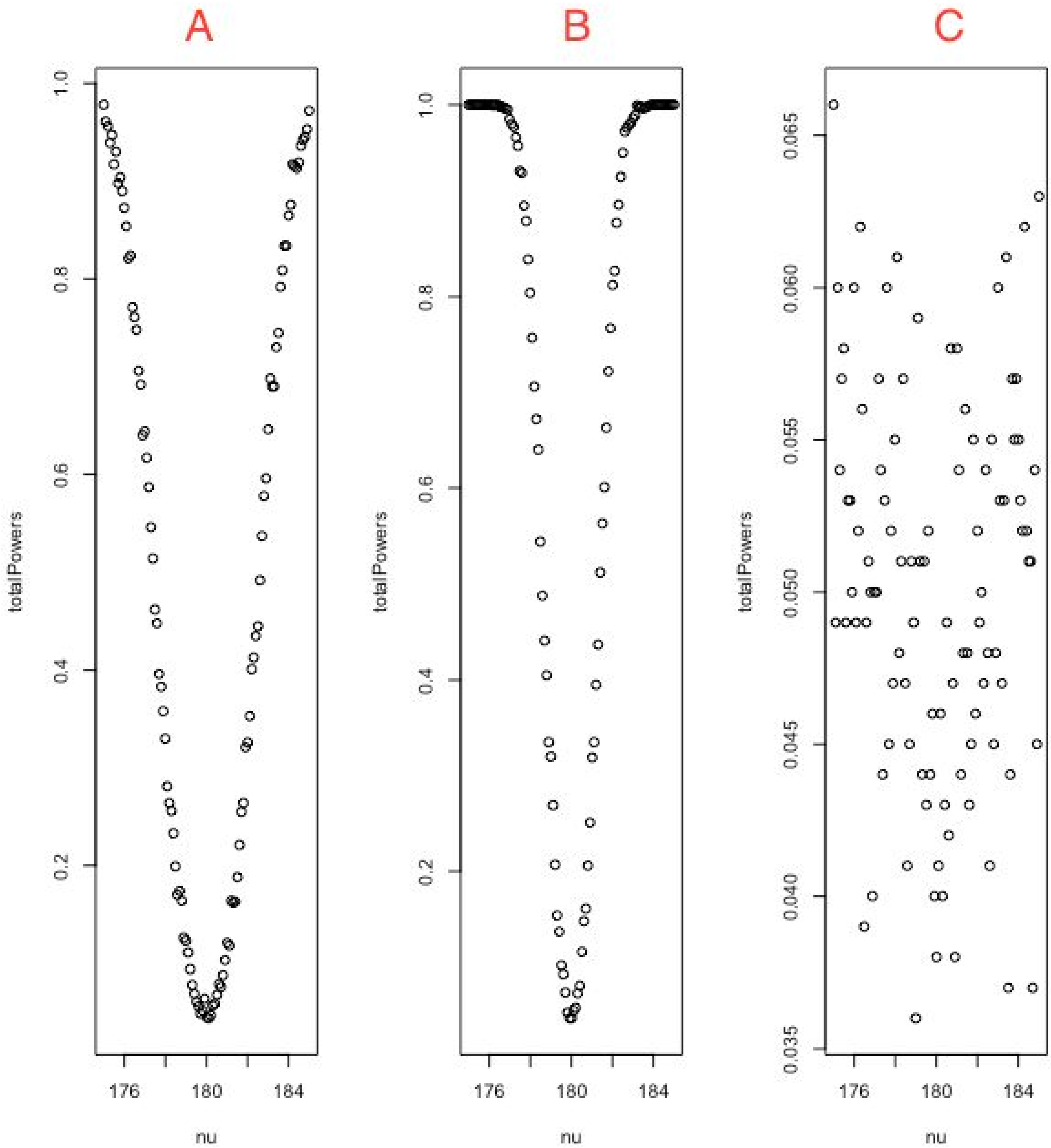


The frequency of p values smaller than 5 percent is 86.3% (863 out of 1000)
The frequency of p values smaller than 10 percent is 92.8% (928 out of 1000)
This histogram really differs than the above two where we can see that the majority of the p-values were less than 0.1.

4) From the above histograms we have seen that when we have the same mean (mu=nu= 180) changing the standard deviation from 10 to 1 did not really have an effect on the frequency of the p-values. Especially if we look at the frequency of p-values below 0.05 is still at 5% thus we can not reject the null hypothesis (the means are equal). However when we changed the means (mu = 180 and nu = 175) and having a standard deviation of 6, we

have seen that the frequency of p-values below 0.05 shooted toward 86.3 % thus we could definitely reject the null hypothesis in this case and say that the means are different.

Exercise 3)



The conditions for each plot:

A) mu= 180, nu = seq(175, 185, by=0.1) , m=n=30 and sd=5
B) mu= 180, nu = seq(175, 185, by=0.1) , m=n=100 and sd=5
C) mu= 180, nu = seq(175, 185, by=0.1) , m=n=30 and sd=100

By looking at plot A we can see that when nu gets far away from 180 (where mu = 180), the power of the p-value start increasing drastically which is in accordance with the t-test and the rejection of the null hypothesis. By increasing the number of observations (n,m) to 100 we recognize more points around nu = 176, 177, 178, 182, 183, 184 and less at nu = 180. Finally, when we increase the standard deviation to 100, the values of the heights will fluctuate more and thus the frequency of p-value to be less than 0.05 becomes really small.

Appendix

Full R code

```
remove(list = ls())

load(file="assign1.RData")

par(mfrow=c(1,2))
hist(x1)
qqnorm(x1, main = "Plot of X1")
hist(x2)
qqnorm(x2, main = "Plot of X2")
hist(x3)
qqnorm(x3, main = "Plot of X3")
hist(x4)
qqnorm(x4, main = "Plot of X4")
hist(x5)
qqnorm(x5, main = "Plot of X5")

normalTwenty = rnorm(20,mean = mean(x1))
qqnorm(normalTwenty,main = "Normal sample with size & mean as x1")
qqnorm(x1, main = "Plot of X1")
qqline(x1)
normalHundred = rnorm(100,mean= mean(x2))
qqnorm(normalHundred,main = "Normal sample with size & mean as x2")
qqnorm(x2, main = "Plot of X2")
qqline(x2)
normalHundred = rnorm(100,mean= mean(x3))
qqnorm(normalHundred,main = "Normal sample with size & mean as x3")
qqnorm(x3, main = "Plot of X3")
qqline(x3)
normalThirty = rnorm(30,mean= mean(x4))
qqnorm(normalThirty,main = "Normal sample with size & mean as x4")
qqnorm(x4, main = "Plot of X4")
qqline(x4)
normalFourty = rnorm(40,mean= mean(x5))
qqnorm(normalFourty,main = "Normal sample with size & mean as x5")
qqnorm(x5, main = "Plot of X5")
qqline(x5)

##### x2 & x5 are not normal

remove(list = ls())
m=30
```

```
n=30
mu=180
nu=175
sd=10
x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
t.test(x,y,var.equal=TRUE)
t.test(x,y,var.equal=TRUE)[[3]]

B=1000
p=numeric(B)
for (b in 1:B) {x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
power=mean(p<0.05)


###Excercise 2)

#1)
par(mfrow=c(1,1))
m=30
n=30
mu=180
nu=180
sd=10
x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
t.test(x,y,var.equal=TRUE)
t.test(x,y,var.equal=TRUE)[[3]]

B=1000
p=numeric(B)
for (b in 1:B) {x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
power=mean(p<0.05)
powerTen = mean(p<0.10)
hist(p)


#2)
m=30
n=30
mu=180
nu=180
```

```
sd=1
x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
t.test(x,y,var.equal=TRUE)
t.test(x,y,var.equal=TRUE)[[3]]

B=1000
p=numeric(B)
for (b in 1:B) {x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
power=mean(p<0.05)
powerTen = mean(p<0.10)
hist(p)

#3)

m=30
n=30
mu=180
nu=175
sd=6
x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
t.test(x,y,var.equal=TRUE)
t.test(x,y,var.equal=TRUE)[[3]]

B=1000
p=numeric(B)
for (b in 1:B) {x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
power=mean(p<0.05)
powerTen = mean(p<0.10)
hist(p)


#Exersice 3)

#1)
par(mfrow=c(1,3))
m=30
n=30
mu=180
nu=seq(175,185,by=0.1)
sd=5
```

```
x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)
t.test(x,y,var.equal=TRUE)
t.test(x,y,var.equal=TRUE)[[3]]

B=1000
p=numeric(B)
totalPowers = length(nu)
for (anNu in 1:totalPowers){
for (b in 1:B) {x=rnorm(m,mu,sd)
y=rnorm(n,nu[anNu],sd)
p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
power=mean(p<0.05)
totalPowers[anNu] = power
powerTen = mean(p<0.10)
}
plot(nu,totalPowers)

#2)

sd=5
m=100
n=100
B=1000
p=numeric(B)
totalPowers = length(nu)
for (anNu in 1:totalPowers){
  for (b in 1:B) {x=rnorm(m,mu,sd)
  y=rnorm(n,nu[anNu],sd)
  p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
  power=mean(p<0.05)
  totalPowers[anNu] = power
  powerTen = mean(p<0.10)
}
plot(nu,totalPowers)

#3)
sd = 100
m=30
n=30
B=1000
p=numeric(B)
totalPowers = length(nu)
for (anNu in 1:totalPowers){
  for (b in 1:B) {x=rnorm(m,mu,sd)
  y=rnorm(n,nu[anNu],sd)
```

```
  p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
  power=mean(p<0.05)
  totalPowers[anNu] = power
  powerTen = mean(p<0.10)
}
plot(nu,totalPowers)
```