Assignment 1-Group11

Name: Jiamian Liu **student number:** 2632301
Name: Xiaoyu Yang **student number:** 2640948
Name: Fangzheng Lyu **student number:** 2644757

Exercise 1

The histogram and QQ-plot of x1, x2, … , x5 is shown in Fig.1 to Fig.5.

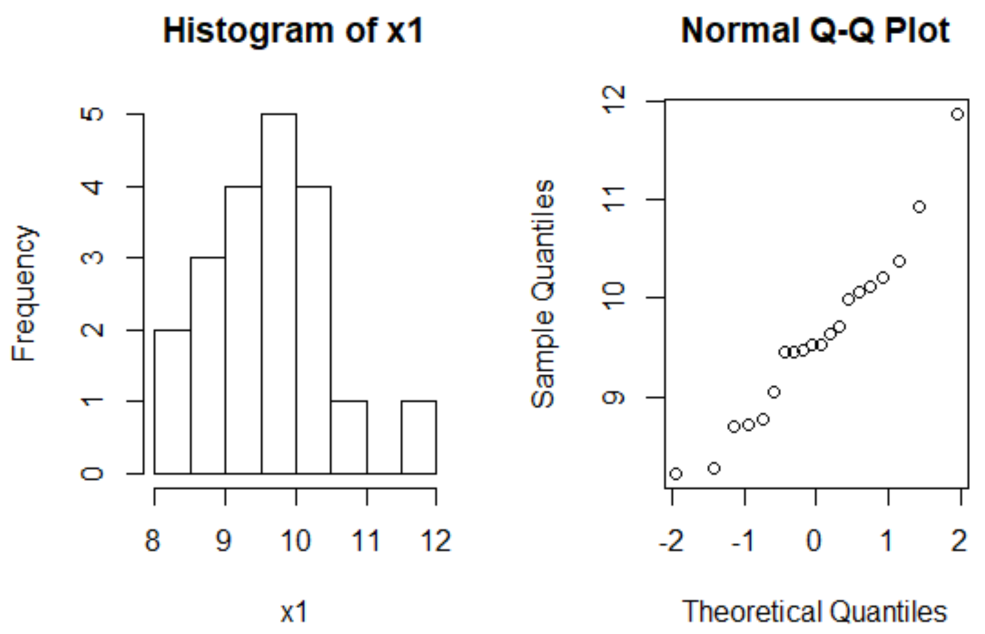

Figure 1: Histogram and QQ-plot of x1

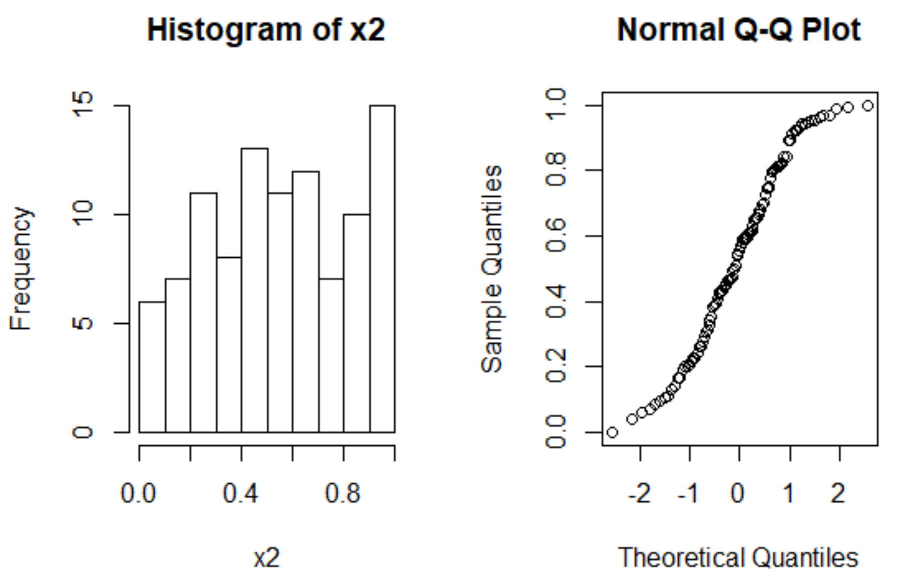

Figure 2: Histogram and QQ-plot of x2

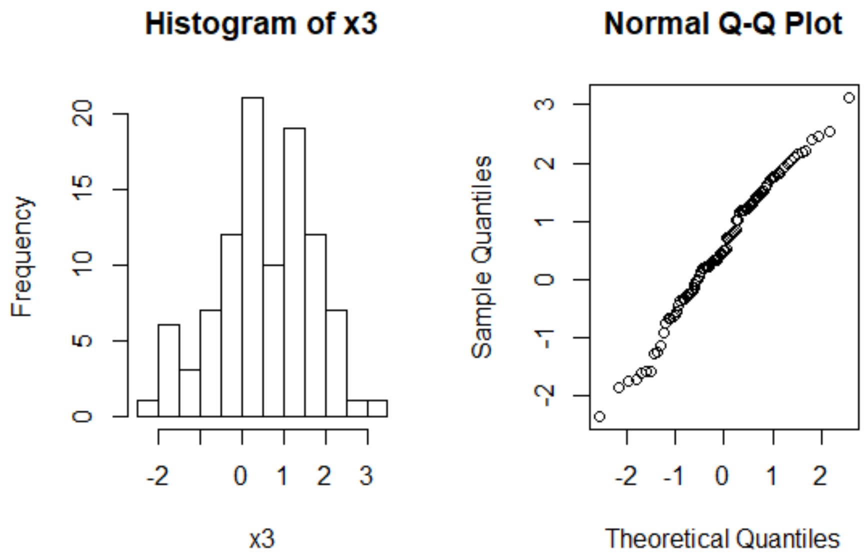

Figure 3: Histogram and QQ-plot of x3

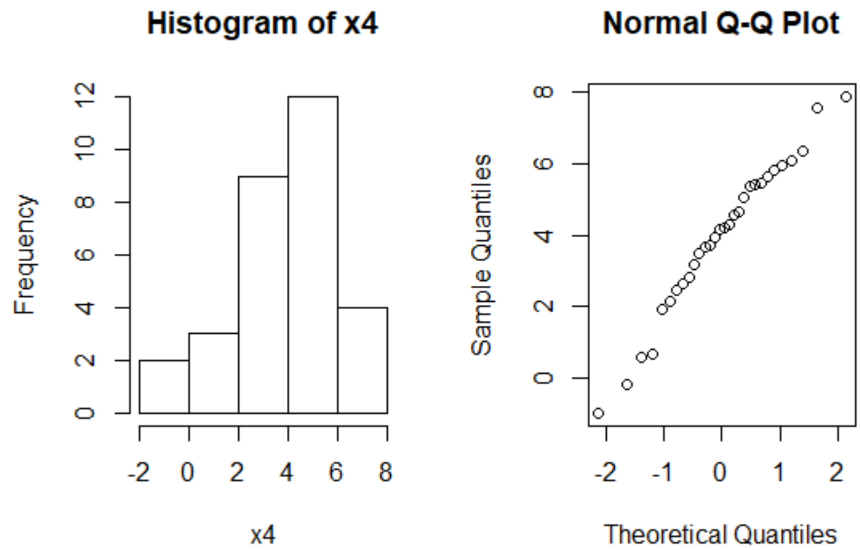

Figure 4: Histogram and QQ-plot of x4

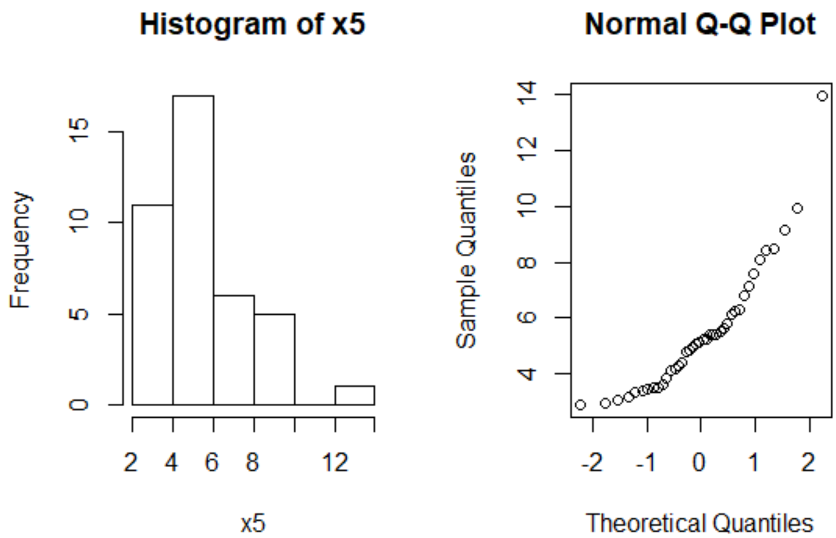

Figure 5: Histogram and QQ-plot of x5

```
> shapiro.test(x1)

        Shapiro-Wilk normality test

data:  x1
W = 0.95401, p-value = 0.432

> shapiro.test(x2)

        Shapiro-Wilk normality test

data:  x2
W = 0.95825, p-value = 0.003022

> shapiro.test(x3)

        Shapiro-Wilk normality test

data:  x3
W = 0.98461, p-value = 0.2975

> shapiro.test(x4)

        Shapiro-Wilk normality test

data:  x4
W = 0.97395, p-value = 0.6517

> shapiro.test(x5)

        Shapiro-Wilk normality test

data:  x5
W = 0.86828, p-value = 0.0002581
```

Figure 6: Shapiro.test

From the histogram we could know that x2 and x5 are not similar to the normal distribution. x2 may be generated from runif() and x5 may be generated from rchisq().
Besides, in the other figures, the peak of histogram is near the mean value of the distribution and the QQ plot is nearly a straight line. Also, we use the shapiro.test to verify the result. As the p-value of x1,x3,x4 are greater than 0.05, x1, x3, x4 could have been sampled from a normal distribution.

Exercise 2

Table 1 shows the p-values results of three situations.

Table 1 the p-values results of exercise 2

	p-values<5%	**p-values<10%**	**distribution**
mu=nu=180, m=n=30, sd=10	42	86	Figure 6
mu=nu=180, m=n=30, sd=1	51	98	Figure 7
mu=180, nu=175, m=n=30, sd=6	891	940	Figure 8

The null hypothesis is that there is no difference between the distribution of men and women heights. We can find that:

(1) For situation 1 and situation 2, we accept the null hypothesis, because most of p-values are larger than 5%. Therefore, men and women have similar distribution of heights.

(2) Situation 1 and situation 2 both accept the null hypothesis, because men and women have same parameters (numbers, means and standard deviations).

(3) For situation 3, we reject the null hypothesis, because most of p-values are smaller than 5%. The reason is that men and women have different means (mu=180, nu=175). Therefore, men and women have different distribution of heights.

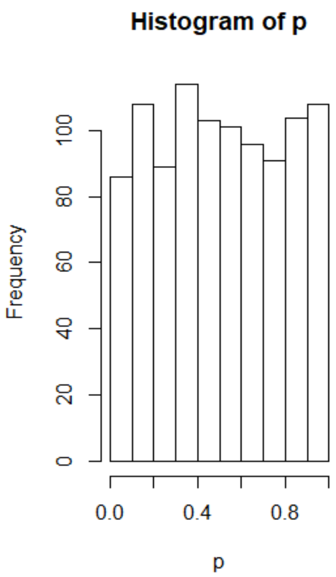

Figure 7: 180, 30, 10

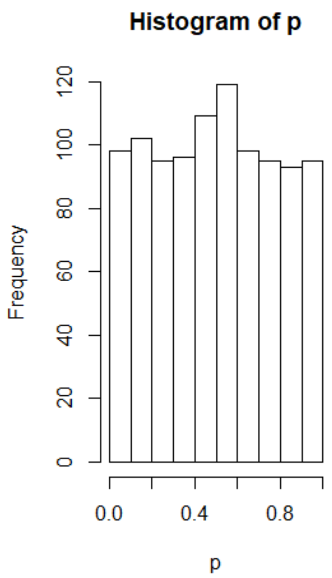

Figure 8: 180, 30, 1

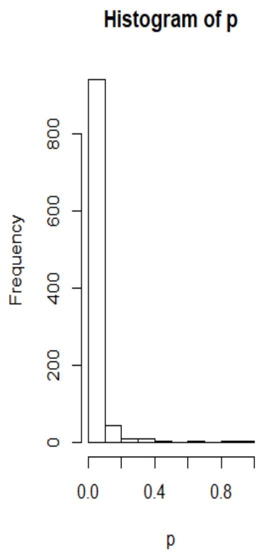

Figure 9: mu=180, nu=175, 30, 6

Scripts of exercise 2:

```
m=30
n=30
mu=180
nu=180
sd=10
x=rnorm(m,mu,sd)
y=rnorm(n,nu,sd)

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)
```

```
power
length(p)
length(p[p<0.05])
length(p[p<0.1])
hist(p)
```

Exercise 3

The plot generated from the exercise 3 script is shown in Fig. 9.

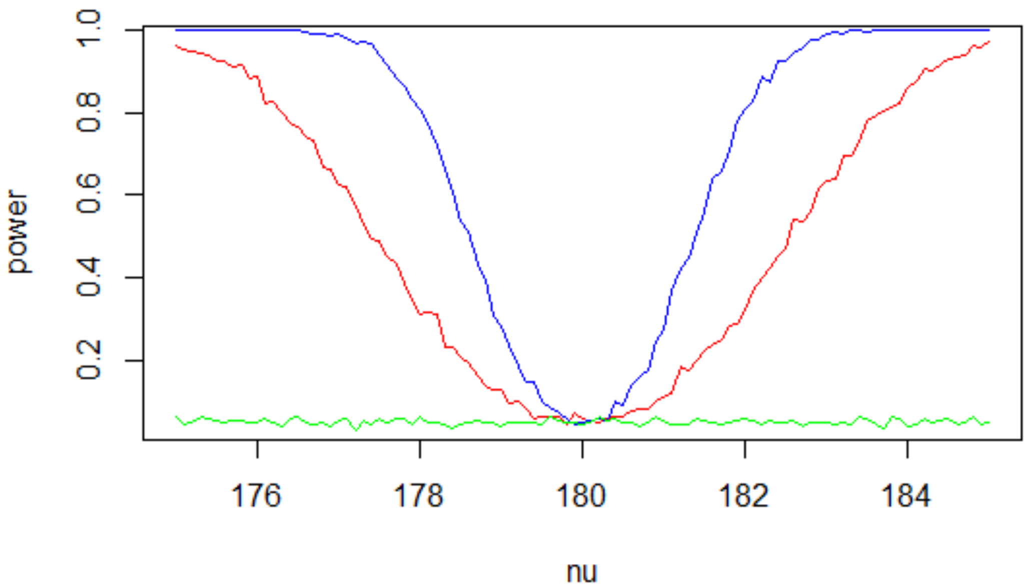

Figure 10: power as a function of nu

In Fig.10, the red line is the result of question 1(m=n=30 and sd=5), blue line is the result of question 2(m=n=100 and sd=5) and the green line is the result of question 3(m=n=30 and sd=100).

Findings:

From the plot we could know that in the proper standard deviations (sd = 5), when the absolute value of difference of mu and nu is larger, the power tends to be larger and the limit of the power is 1, which means for every given set of parameters (m; n; μ; υ; sd), the probability that the t-test rejects the null hypothesis increase. When the mean values are equal, the power is 0. In different value of nu, when the number of samples is larger, such as the blue line m=n=100 larger than the red line m=n=30, the power is larger, which means for larger mean value, for every given set of parameters (m; n; μ; υ; sd), the probability that the t-test rejects the null hypothesis increase. When the standard deviation is improper, which means too large for the test set, there is no big different power value for different nu values as the green line shows. In this situation, for every given set of parameters (m; n; μ; υ; sd), the probability that the t-test rejects the null hypothesis is nearly 0 for each value.

The scripts of the first three questions is listed as below:

Script of exercise 3:

```
nu=175
power1=numeric(101)
power2=numeric(101)
power3=numeric(101)
n=numeric(101)

for(i in 1:101){
   B=1000
   p=numeric(B)
   for (b in 1:B) {x=rnorm(30,180,5)
                   y=rnorm(30,nu,5)
                   p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
   power=mean(p<0.05)
   power1[i]=power

   for (b in 1:B) {x=rnorm(100,180,5)
                   y=rnorm(100,nu,5)
                   p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
   power=mean(p<0.05)
   power2[i]=power

   for (b in 1:B) {x=rnorm(30,180,100)
                   y=rnorm(30,nu,100)
                   p[b]=t.test(x,y,var.equal=TRUE)[[3]]}
   power=mean(p<0.05)
   power3[i]=power


   n[i]=nu
   nu=nu+0.1
}

plot(n,power1,xlab="nu",ylab="power",type="l",col="red")
lines(n,power2,col="blue")
lines(n,power3,col="green")
```