Assignment 1, EDDA 2017

Martin de la Riva(11403799) and Kieran O'Driscoll(11426438), Group 23

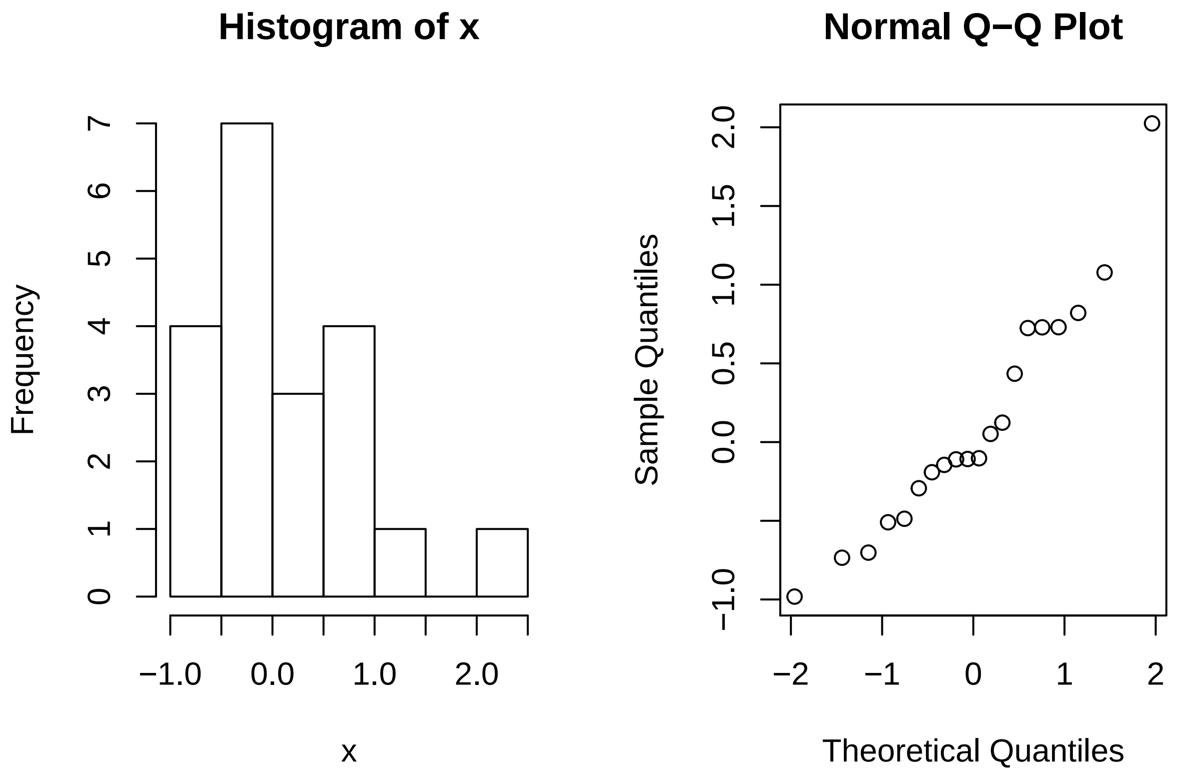
10 April 2017

Exercise 1

```
load(file="assign1.RData")
```

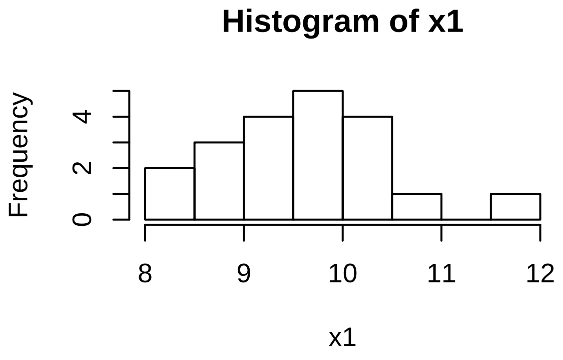
Firstly, we did some tests of normal distributions with similar sizes as x1. . . x5 to get a greater idea of how they could look like. E.g.: in the following code we test the look of Histograms and QQ-plot with a 20-size vector.

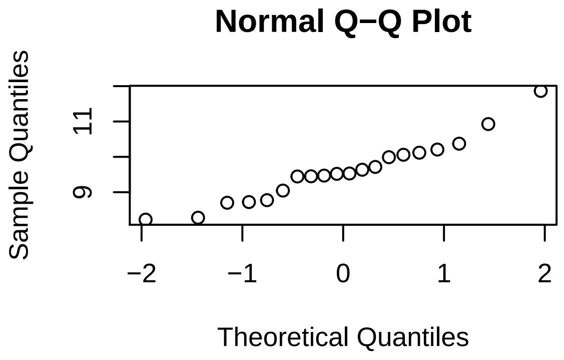
```
par(mfrow=c(1,2))
x=rnorm(20)
hist(x)
qqnorm(x)
```

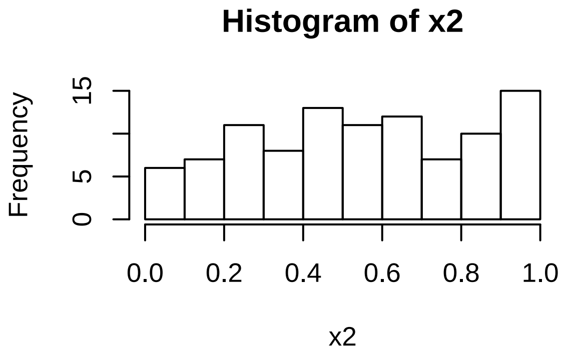


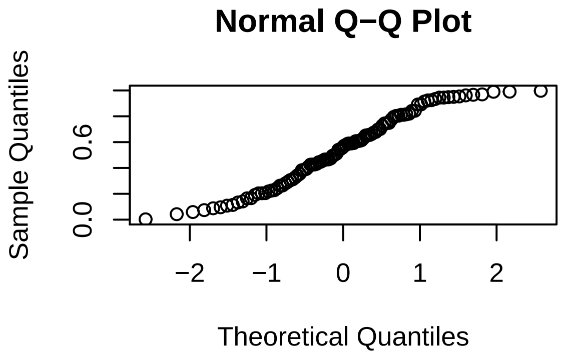
Then we plot the other distributions to see the results. We decided that x1, x3 and x4 are based on a normal distribution, while x2 and x5 are not.

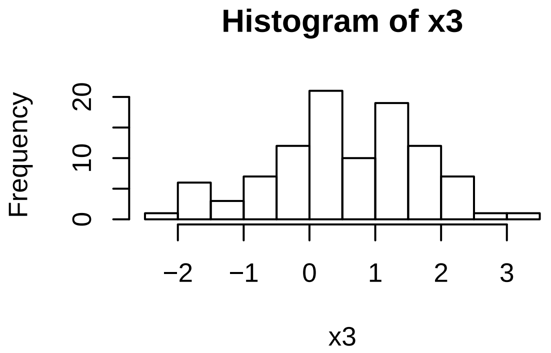
```
hist(x1)
qqnorm(x1)
# x1 normal distribution?: yes

hist(x2)
qqnorm(x2)
# x2 normal distribution?: no

hist(x3)
qqnorm(x3)
# x3 normal distribution?: yes

hist(x4)
qqnorm(x4)
# x4 normal distribution?: yes

hist(x5)
qqnorm(x5)
# x5 normal distribution?: no
```

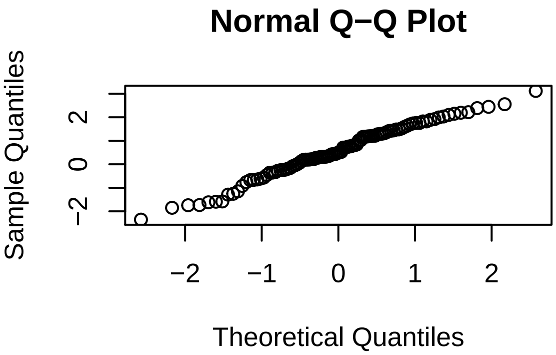


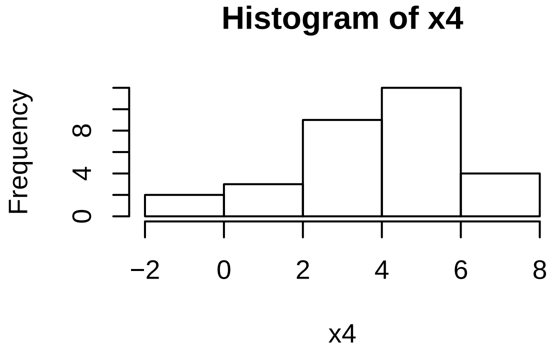


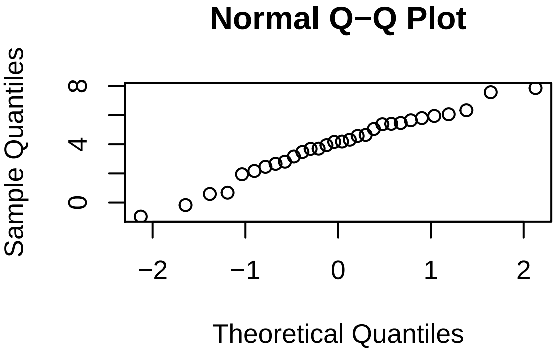


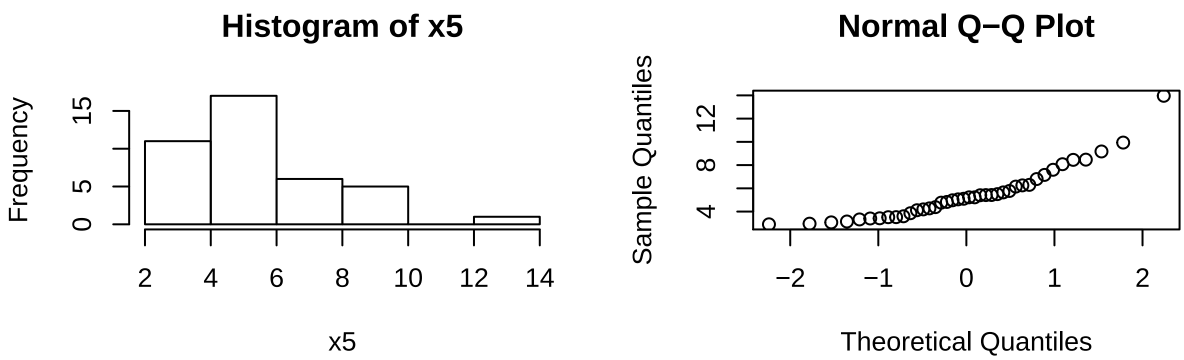



Histogram of x3
Frequency
0
10
20
−2
−1
0
1
2
3
x3


Normal Q−Q Plot
Sample Quantiles
−2
0
2
−2
−1
0
1
2
Theoretical Quantiles


Histogram of x4
Frequency
0
4
8
−2
0
2
4
6
8
x4


Normal Q−Q Plot
Sample Quantiles
0
4
8
−2
−1
0
1
2
Theoretical Quantiles



Exercise 2

```r
par(mfrow=c(1,2))
calculatePvalues <- function(m,n,mu,nu,sd,B){
    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]]
    }
    return(p)
}
```

In the script above we create a function to calculate P-values given the needed attributes.

1

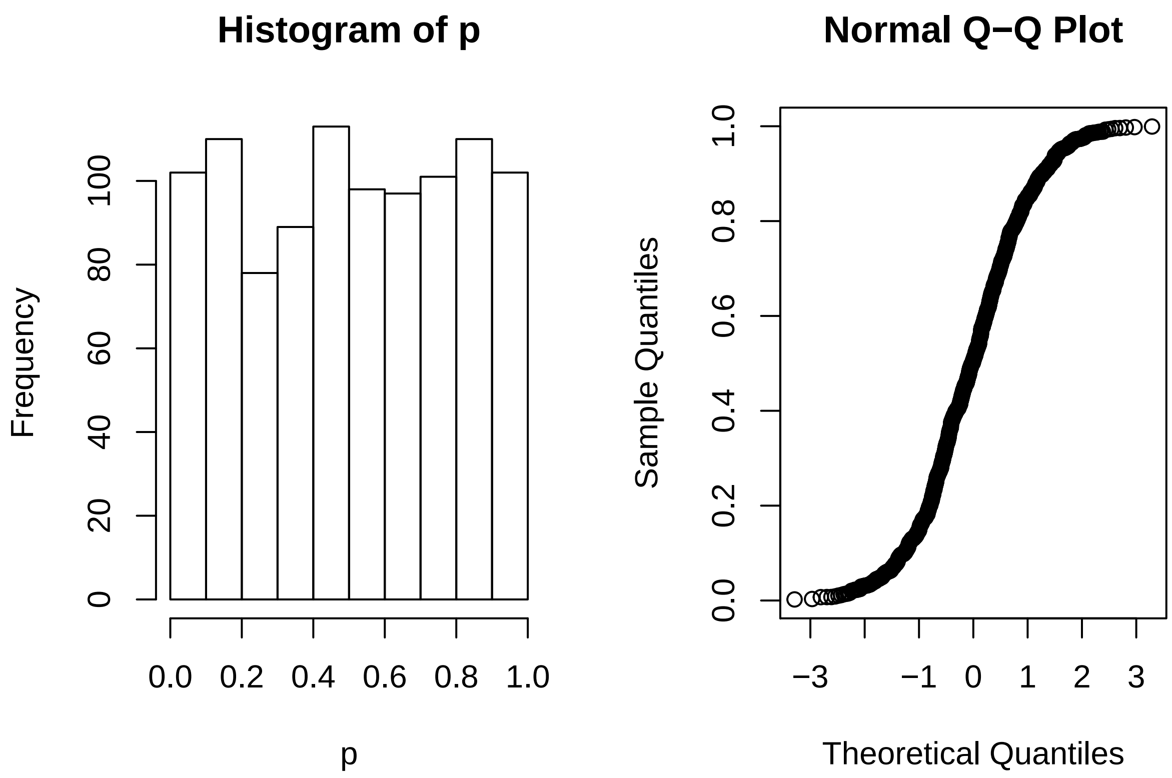
```r
mu=nu=180; m=n=30; sd=10; B=1000
p=calculatePvalues(m,n,mu,nu,sd,B)
# number of pvalues smaller than 5% (49)
length(p[p<0.05])
```

```
## [1] 47
```

```r
# number of pvalues smaller than 10% (98)
length(p[p<0.1])
```

```
## [1] 102
```

```r
# distribution of p-values: uniform distribution
par(mfrow=c(1,2))
hist(p)
qqnorm(p)
```



We got around 50 number of p-values smaller than 5%, and around 100 smaller than 10%, which results in 5% and 10% of the trials respectively. The p-values follow a uniform distribution.

2

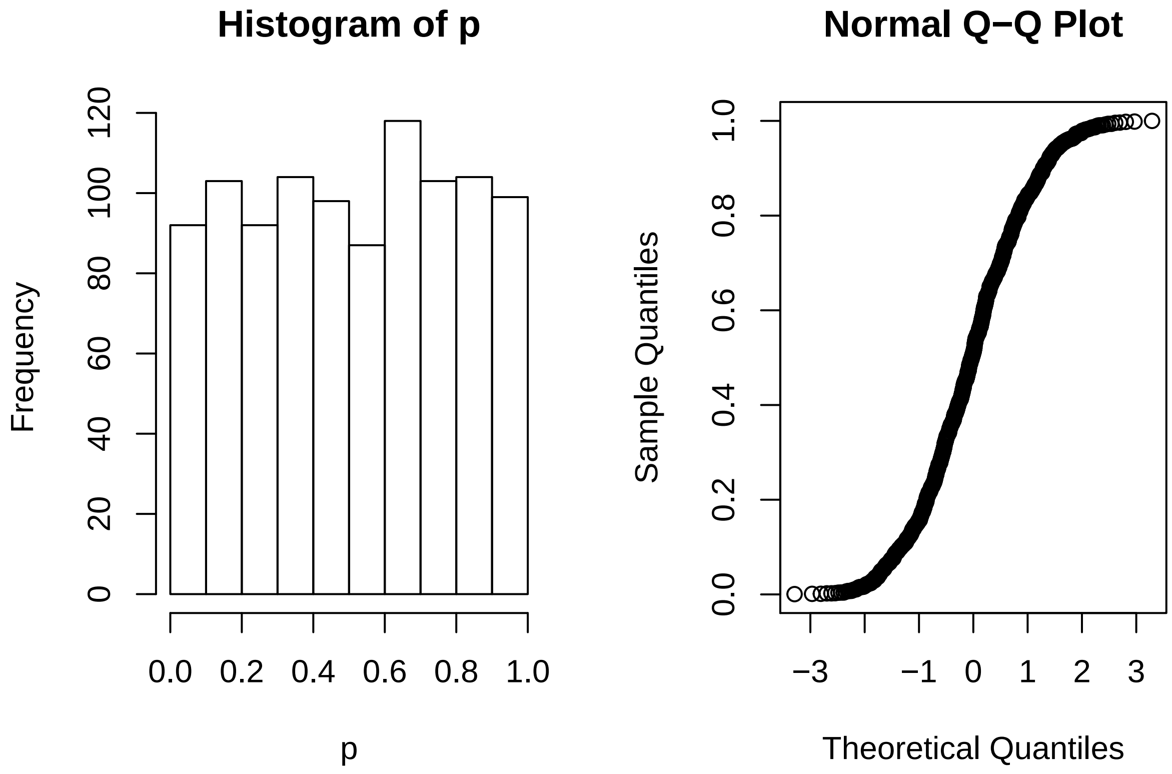
```r
mu=nu=180; m=n=30; sd=1; B=1000
p=calculatePvalues(m,n,mu,nu,sd,B)
# number of pvalues smaller than 5% (51)
length(p[p<0.05])
```

```
## [1] 46
```

```
# number of pvalues smaller than 10% (104)
length(p[p<0.1])
```

```
## [1] 92
```

```
# distribution of p-values: uniform distribution
par(mfrow=c(1,2))
hist(p)
qqnorm(p)
```



We got really similar results as in 2.1.

3

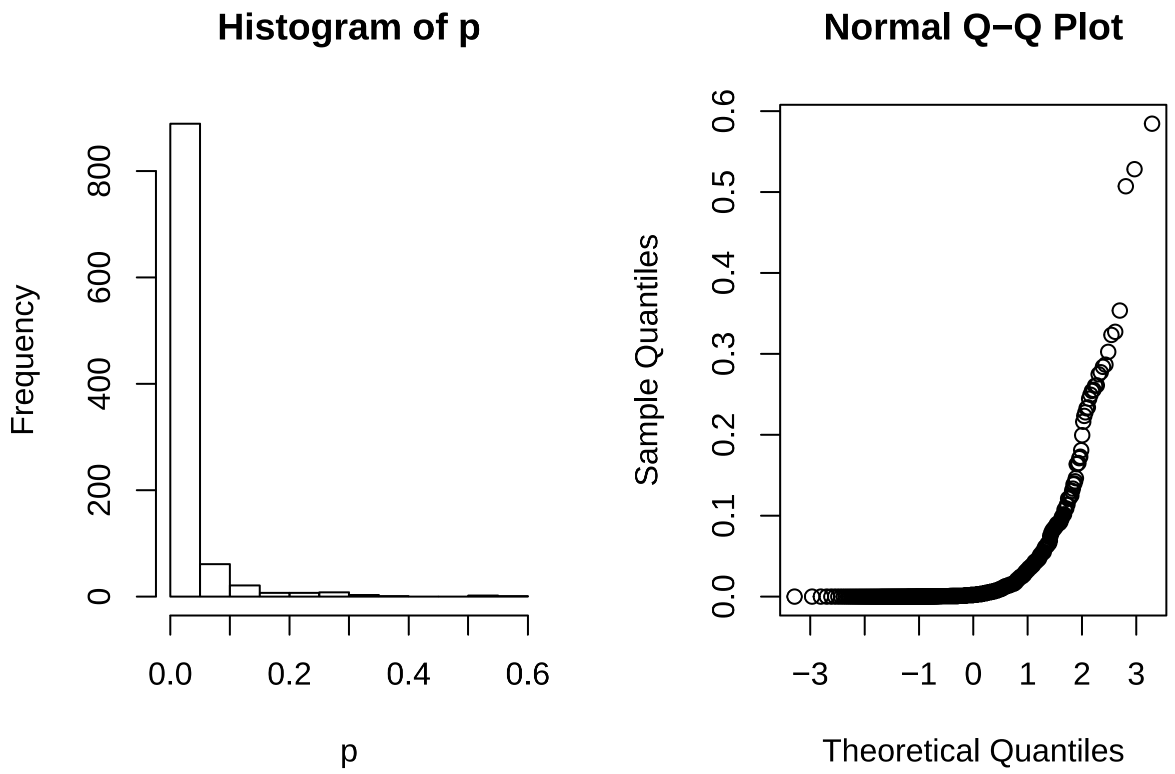
```
mu=180; nu=175; m=n=30; sd=6; B=1000
p=calculatePvalues(m,n,mu,nu,sd,B)
# number of pvalues smaller than 5% (903)
length(p[p<0.05])
```

```
## [1] 889
```

```
# number of pvalues smaller than 10% (943)
length(p[p<0.1])
```

```
## [1] 950
```

```
# distribution of p-values: chi-square distribution with low degrees of freedom
par(mfrow=c(1,2))
hist(p)
qqnorm(p)
```



We get completely different results, getting to 90% and 95% repectively. In this case, the p-values seem to follow a chi-square distribution with low degrees of freedom.

4

(H_0: difference in (m,n) means of population=0)

2.1 and 2.2: we don't reject H_0, as the means are not significantly big. We expected to see a lower percentage of p-values less than 5% and 10% in 2.2 due to the lower standard deviation, although no big differences were noticed.

2.3: difference in means is really big, so H_0 is usually rejected.

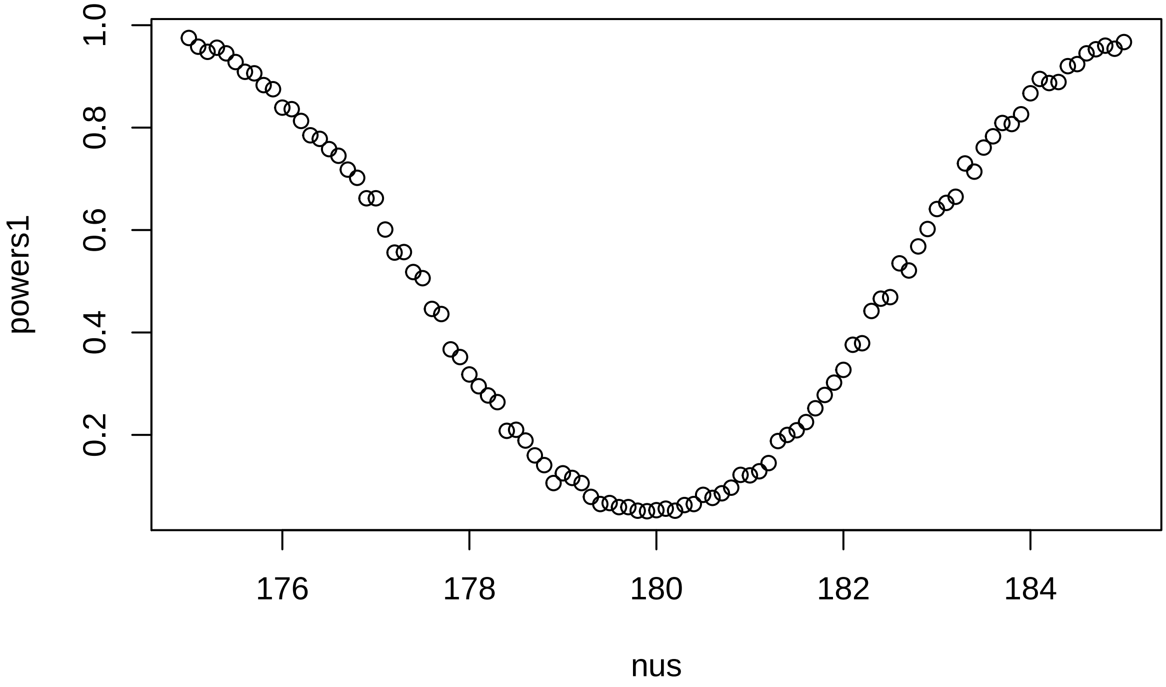
Exercise 3

```
par(mfrow=c(1,1))
calculatePvaluesNus <- function(m,n,mu,nus,sd,B){
    powers=numeric(length(nus))
    for (i in 1:length(nus)) {
        nu=nus[i]
        p=calculatePvalues(m,n,mu,nu,sd,B)
        powers[i]=mean(p<0.05)
    }
    return(powers)
}
```

We start creating a function to calculate p-values over different values of nu.

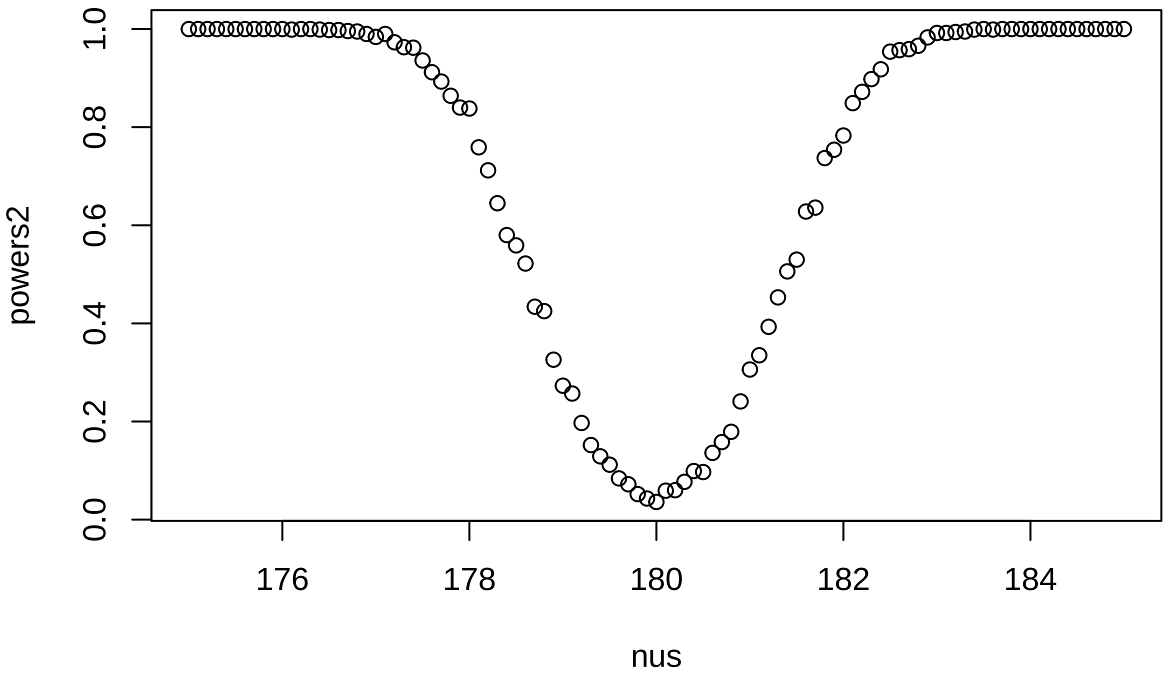
1

```
    mu=180; m=n=30; sd=5; B=1000
    nus=seq(175,185,by=0.1)
    powers1=calculatePvaluesNus(m,n,mu,nus,sd,B)
    plot(nus, powers1)
```



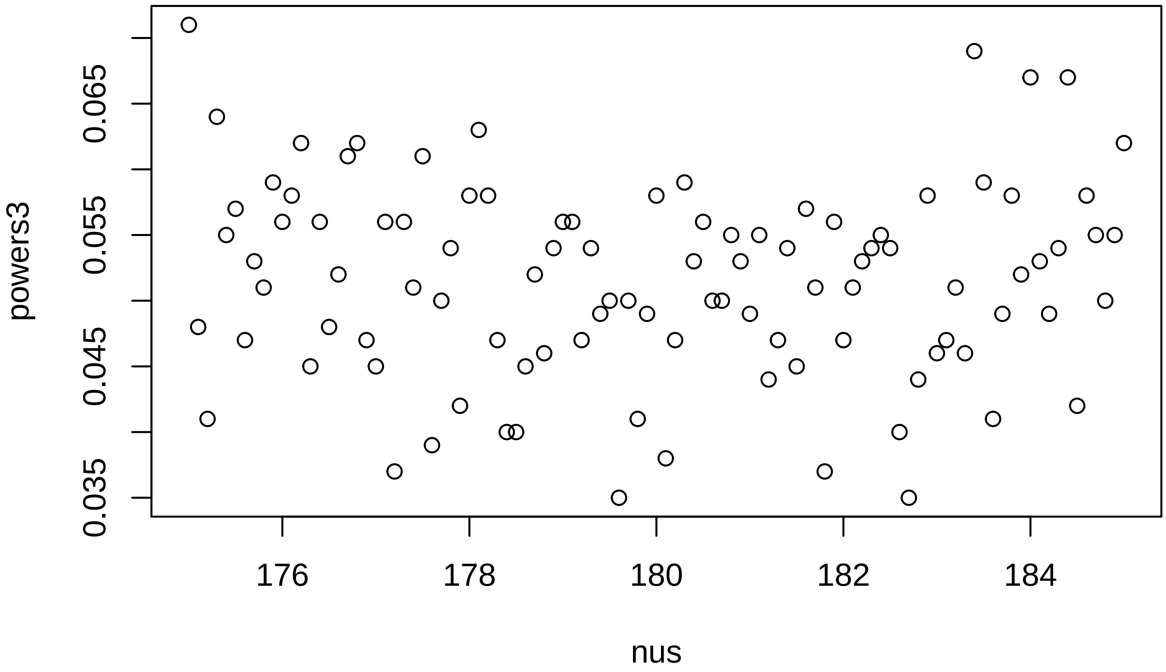
2

```
mu=180; m=n=100; sd=5; B=1000
nus=seq(175,185,by=0.1)
powers2=calculatePvaluesNus(m,n,mu,nus,sd,B)
plot(nus, powers2)
```



3

```
mu=180; m=n=30; sd=100; B=1000
nus=seq(175,185,by=0.1)
powers3=calculatePvaluesNus(m,n,mu,nus,sd,B)
plot(nus, powers3)
```



4

3.1 and 3.2: when sample is bigger, there is less uncertainty about the population, therefore increasing in our example the amount of rejected hypothesis. It can be seen how with 100 size samples (3.2) H_0 starts not to get rejected in 178 cm, while for 30 size samples (3.1) H_0 starts not to get rejected at 176 cm.

3.3: In this case the standard deviation is too high, and therefore it is more difficult to make assumptions about the population. This leads to random results, even when nu == 180 cm (only a subtle difference can be detected).