HW7

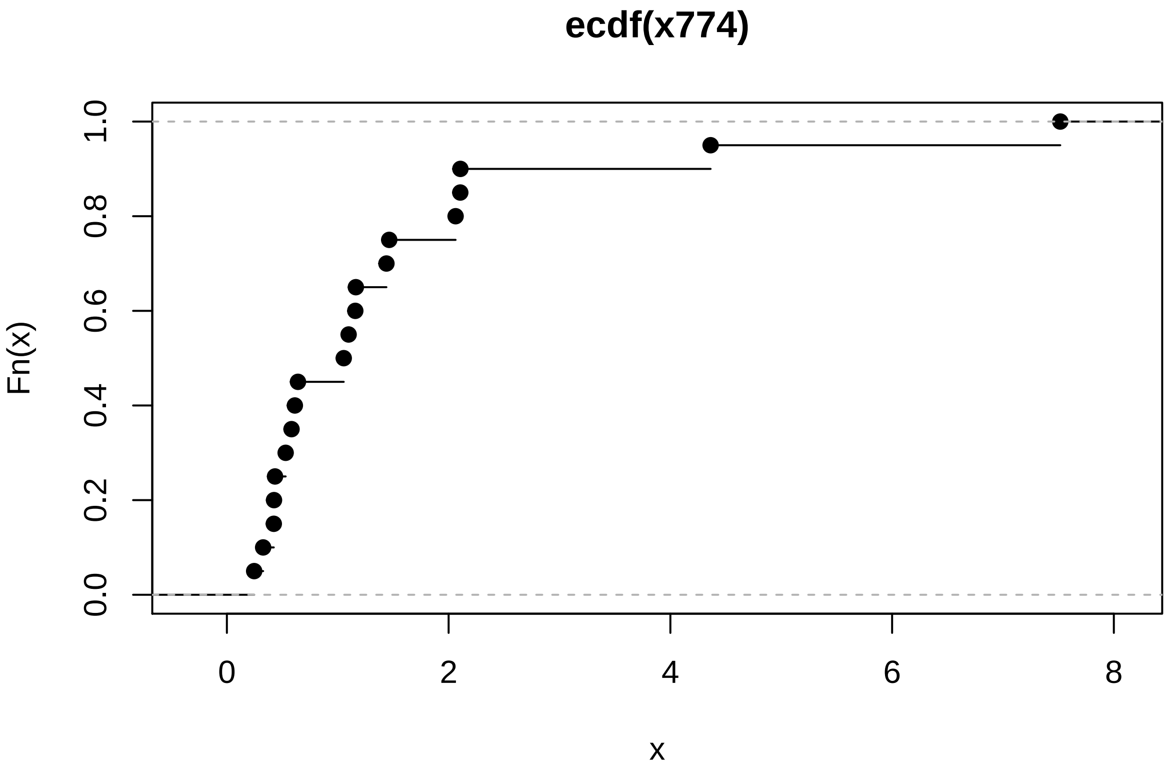
Sumeet Mishra

2/28/2019

```
x774 = scan("http://pages.iu.edu/~mtrosset/StatInfeR/Data/sample774.dat")
x776 = scan("http://pages.iu.edu/~mtrosset/StatInfeR/Data/test351.dat")
```

1.(a)

```
plot(ecdf(x774))
```



(b)

plug-in estimates of the mean

```
mean(x774)
```

```
## [1] 1.4876
```

variance

```r
var=mean(x774^2) - mean(x774)^2
var
```

```
## [1] 2.787554
```

median

```r
median(x774)
```

```
## [1] 1.076
```

interquartile range

```r
iq=quantile(x774,0.75,type=2)-quantile(x774,0.25,type=2)
iq
```

```
##    75%
## 1.2815
```

(c)

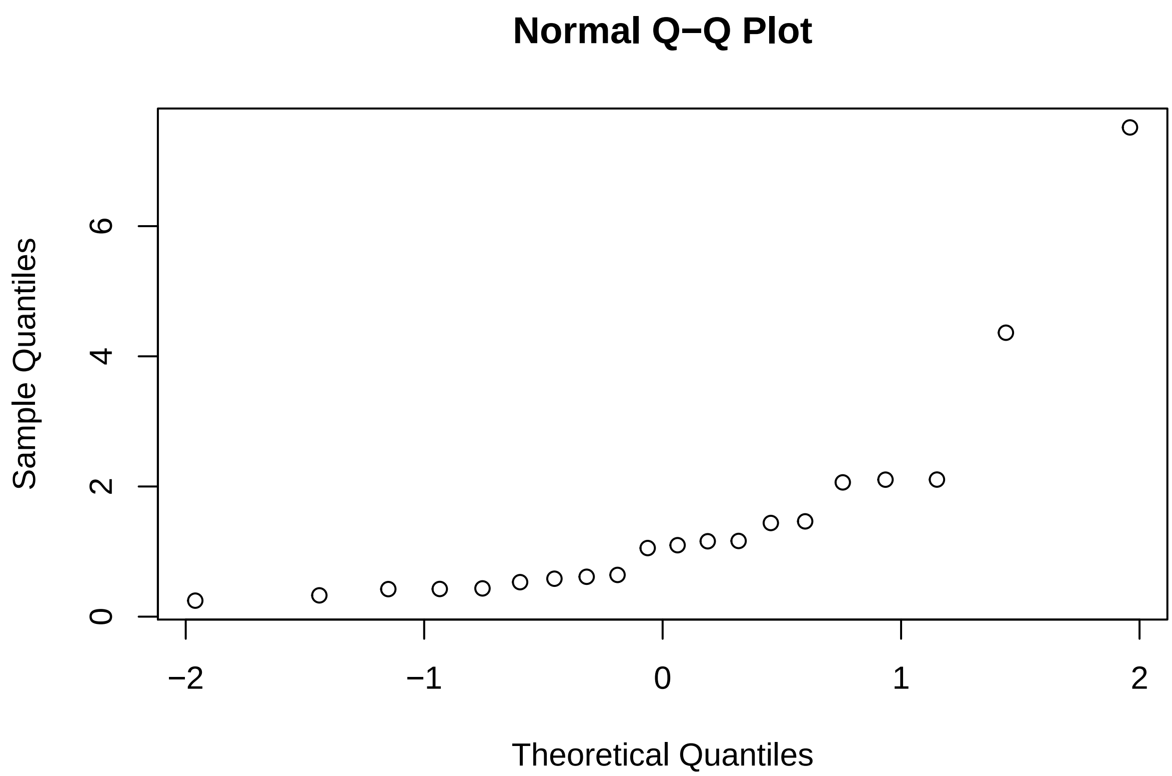
```r
iq/sqrt(var)
```

```
##       75%
## 0.7675505
```

Since the ratio of sd to interquantile range is not ~1.35,we can say that it is not taken from normal distribution.

(d)

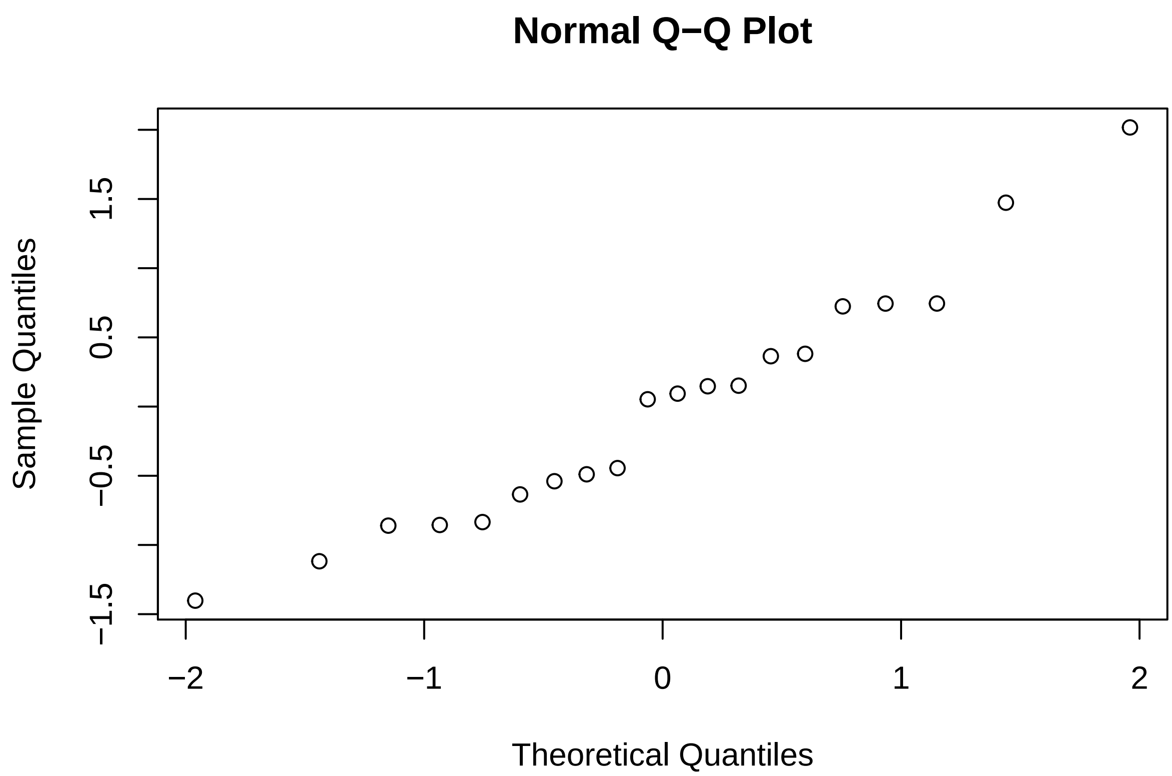
```r
qqnorm(x774)
```



Since the points on the graph together does not make a line from bottom left to top right(linear), we can say that it is not drawn from normal distribution.

(e)

```
y <- log(x774)
qqnorm(y)
```

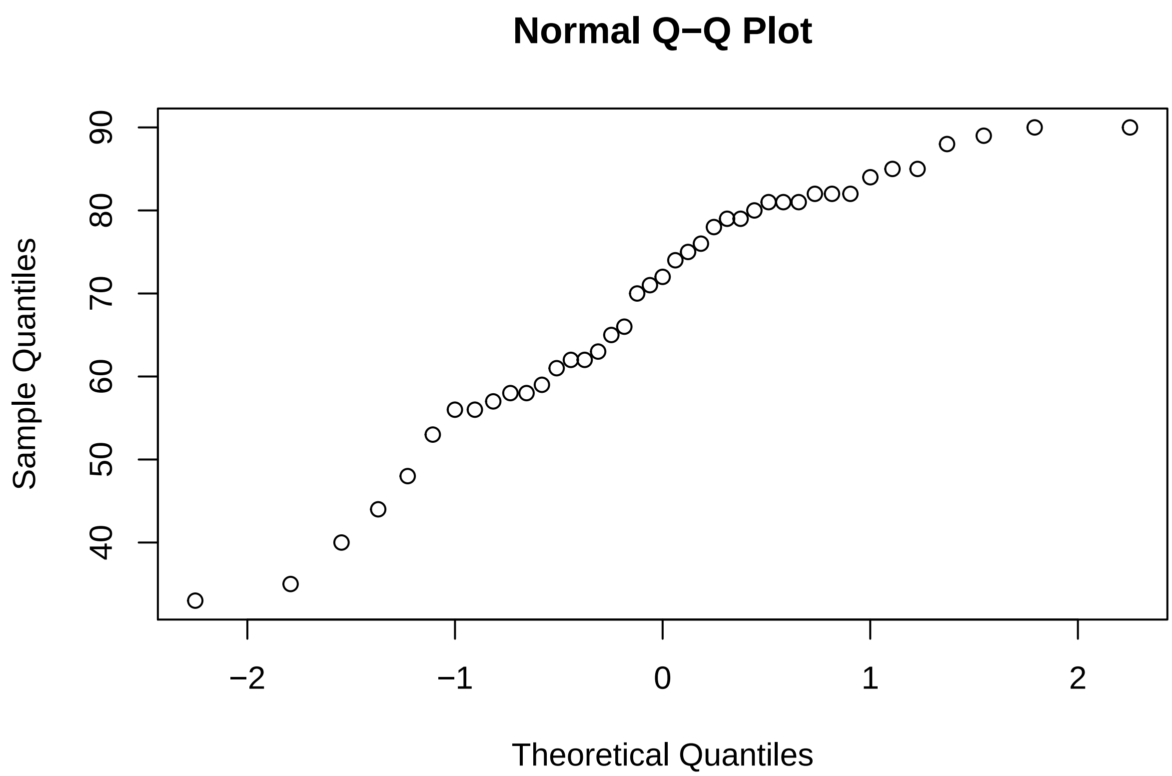


It is still not a normal distribution as the points on the graph together does not make a line from bottom left to top right(linear).

2.Forty-one students taking Math 351 (Applied Statistics) at the College of William & Mary were administered a test.

(a) Do these numbers appear to be a random sample from a normal distribution?

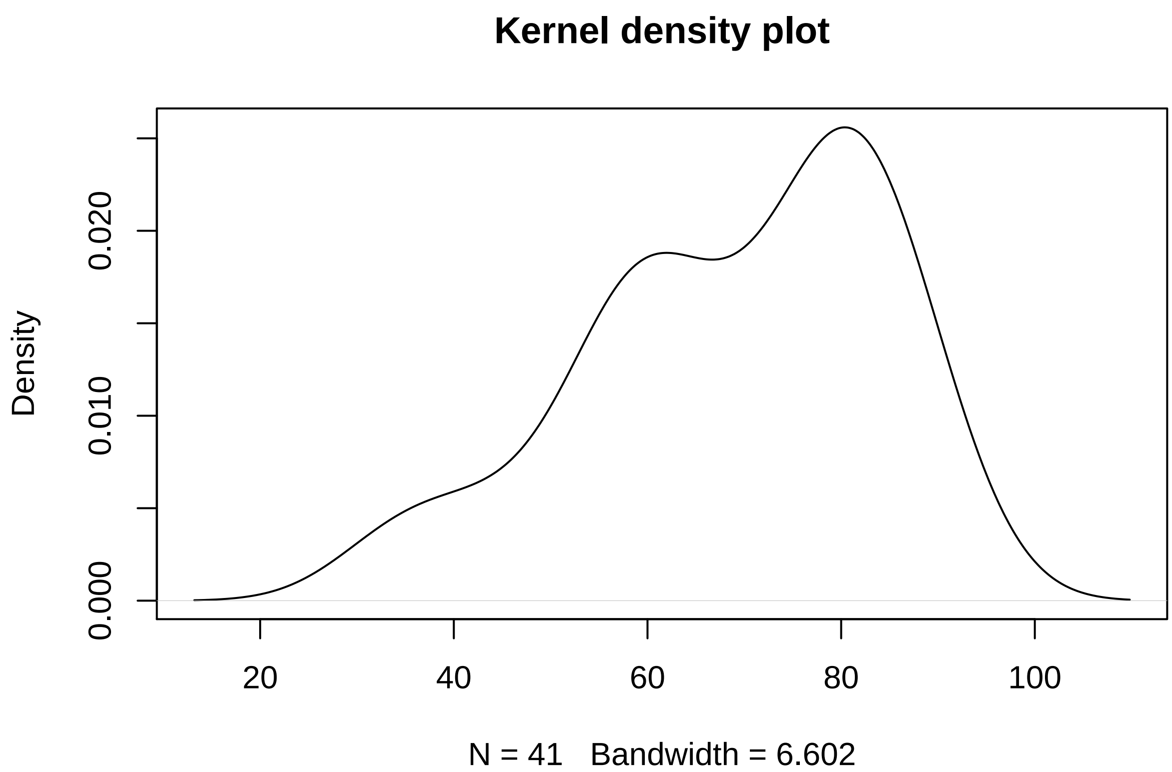
```
qqnorm(x776)
```



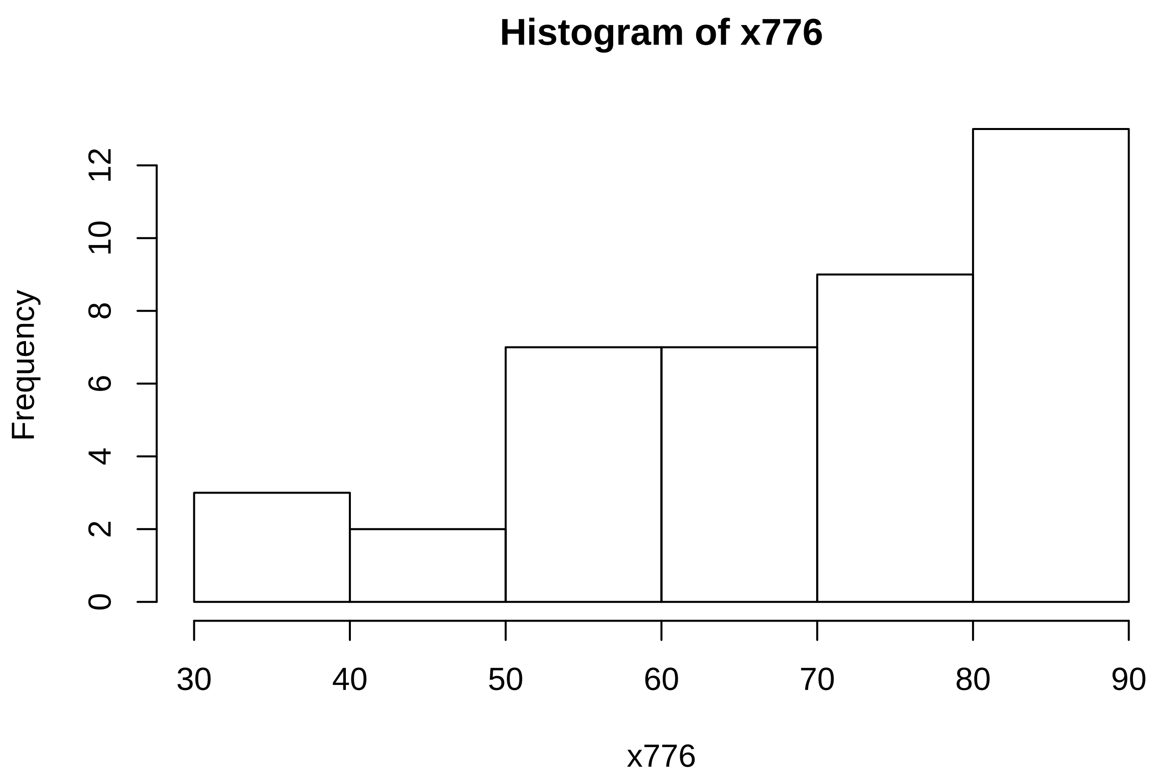
It's not a normal distribution as the points on the graph together does not make a line from bottom left to top right(linear)

(b) Does this list of numbers have any interesting anomalies?

```
plot(density(x776), main ='Kernel density plot')
```


Kernel density plot
Density
0.000
0.010
0.020
20
40
60
80
100
N = 41   Bandwidth = 6.602

```
hist(x776)
```


Histogram of x776
Frequency
0
2
4
6
8
10
12
30
40
50
60
70
80
90
x776

From the above 2 plots and the normal density plot we observed that the distribution is taken from a bimodal distribution skewed to the left.

3.

n=20,mean=20x5=100,sd=(30/60)x20=10

```
1-pnorm(105,100,10/sqrt(20))
```

```
## [1] 0.01267366
```

4.(a)

the expected value of X is-

EX=(-2x0.3)+(-1x0.6)+(12x0.1)=0

(b)Var(X) is-

```
(((-2)^2*0.3)+((-1)^2*0.6)+((12)^2*0.1))-((-2*0.3)+(-1*0.6)+(12*0.1))^2
```

```
## [1] 16.2
```

(c)

The expected value is-

0

(d)

The variance is 16.2/n.

(e)

n = 100

```
1-pnorm(0.5,0,sqrt(16.2/100))
```

```
## [1] 0.1070703
```

5.

```
data=read.csv('Book2.csv', header = TRUE)
#View(data)
```

(a) Lacking any other information, our best estimate for the population mean household size is the sample mean. What is the sample mean of our data?

```
n=sum(data$Number_of_households)
mean=sum(data$Number_of_households*data$Household_size)/n
mean
```

```
## [1] 2.5
```

(b) What is our estimate for the standard deviation of household sizes?

```
EX2=sum((data$Household_size)^2*data$Number_of_households)/n
var=EX2-(mean)^2
sd=sqrt(var)
sd
```

```
## [1] 1.403567
```

(c) What is the estimated standard error of the sample mean? (That is, based on our answer to (b), what is our estimate for the standard deviation of the distribution of the sample mean?)

```
stderror=sd/sqrt(n)
stderror
```

```
## [1] 0.1403567
```

(d) Our error is the difference between the sample mean and the population mean. Using the normal distribution, find the approximate probability that the absolute value of the error in a survey of this form and size is less than 0.5.

$P(|X|<0.5) => P(-0.5<X<0.5)$

$=> F(0.5)-F(-0.5)$

```
pnorm(0.5,0,stderror)-pnorm(-0.5,0,stderror)
```

```
## [1] 0.9996325
```