Assignment – 6

Parmeet Singh Ahluwalia

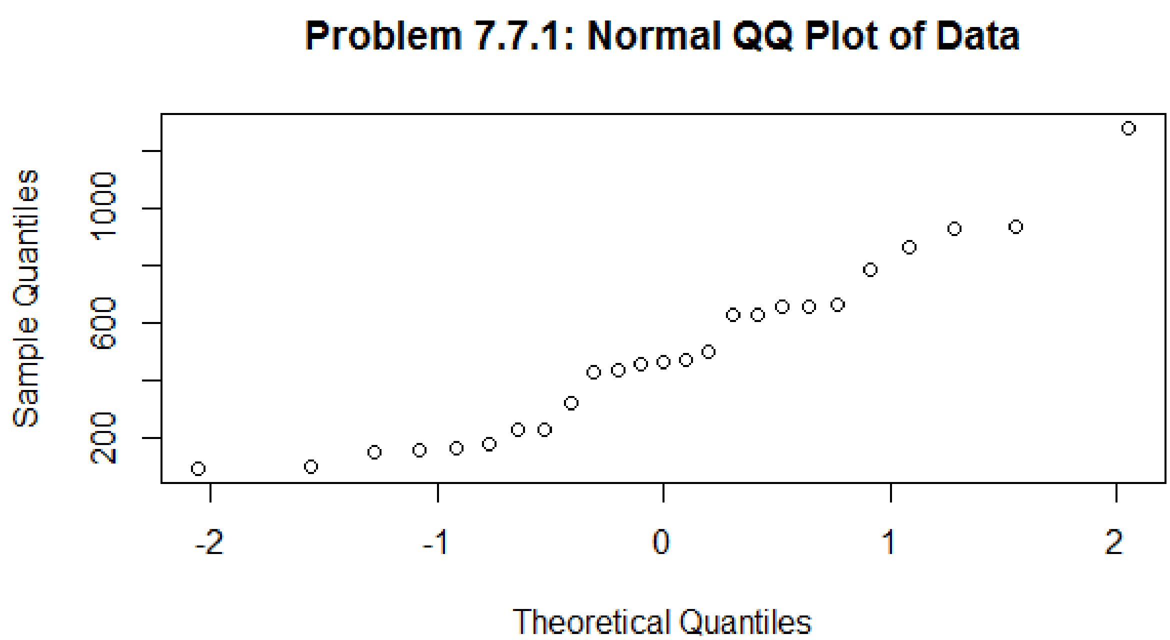
Stat S-520

<u>Answers</u>

A1)

<u>7.7.1 f:</u>

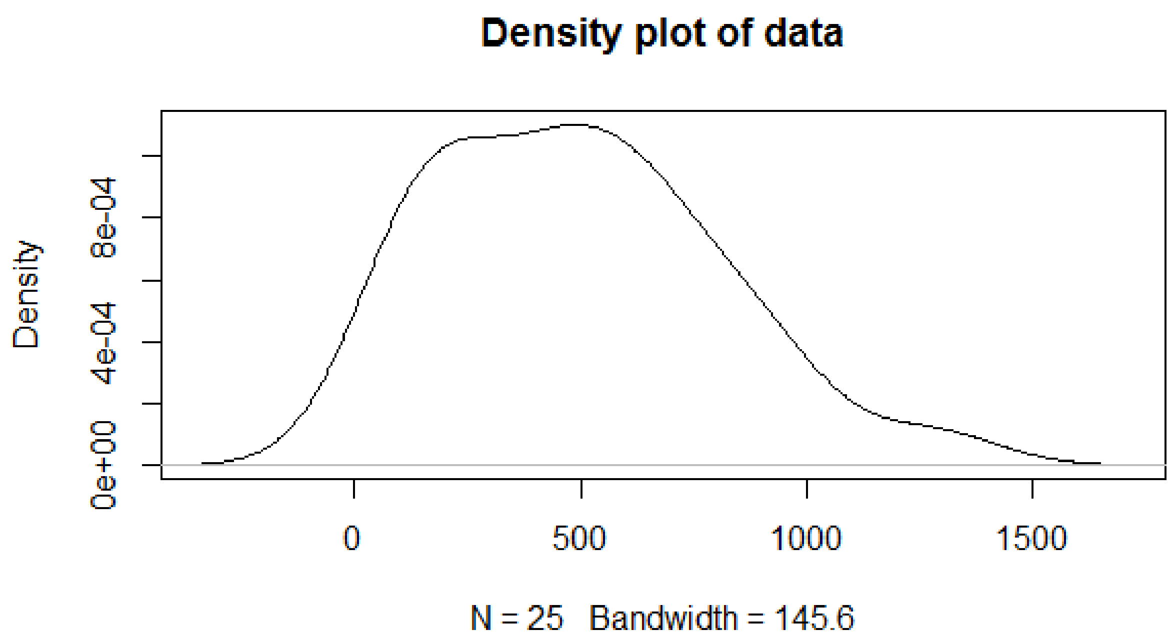
```
> data = c(462,425,164,784,625,472,658,658,663,928,92,230,
96,626,1277,225,150,320,496,157,458,933,861,174,431)

> qqnorm(data, main ="Problem 7.7.1: Normal QQ Plot of Data")
```



<u>7.7.1 g:</u>

```
>require("stats")

>density(data)

Call:

      density.default(x = data)

Data: data (25 obs.); Bandwidth 'bw' = 145.6
```

```
       x                  y            
 Min.   :-344.9   Min.   :1.233e-06  
 1st Qu.: 169.8   1st Qu.:8.725e-05  
 Median : 684.5   Median :3.641e-04  
 Mean   : 684.5   Mean   :4.852e-04  
 3rd Qu.:1199.2   3rd Qu.:9.565e-04  
 Max.   :1713.9   Max.   :1.102e-03  
>plot(density(data), main="Density plot of data")
```



h) The data is not a straight line, hence it is not a normal distribution.

A2) Trosset Exercise 7.7.4:

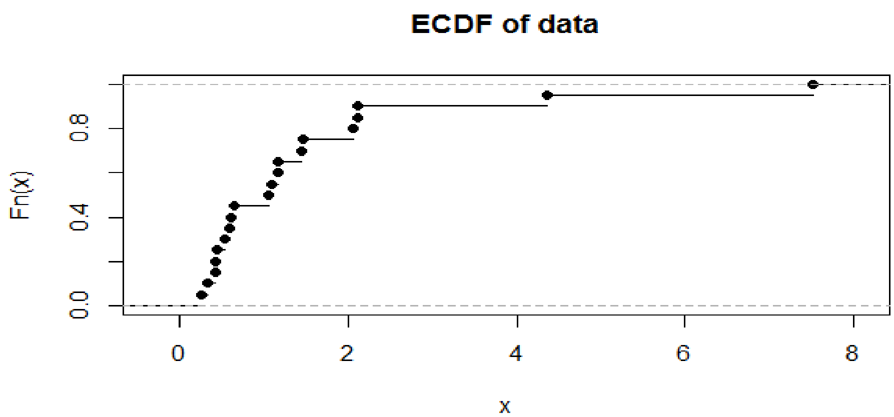
The **R** code is as follows:

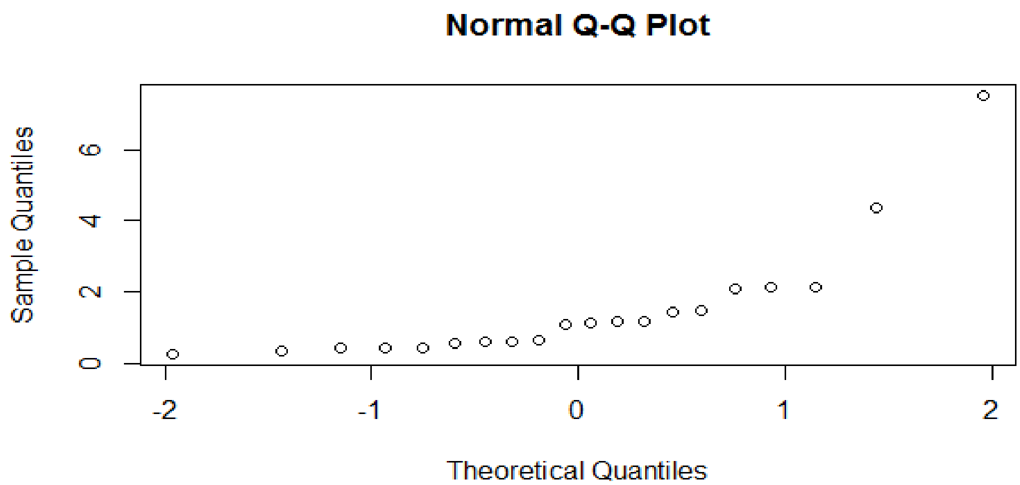
```
> data = scan("http://mypage.iu.edu/~mtrosset/StatInfeR/Data/sample774.dat")
Read 20 items
> plot(ecdf(data), main="ECDF of data")
> summary(data)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  0.246   0.506   1.076   1.488   1.614   7.517 

> mean(data^2) - mean(data)^2 # Plug-in variance
[1] 2.787554
```

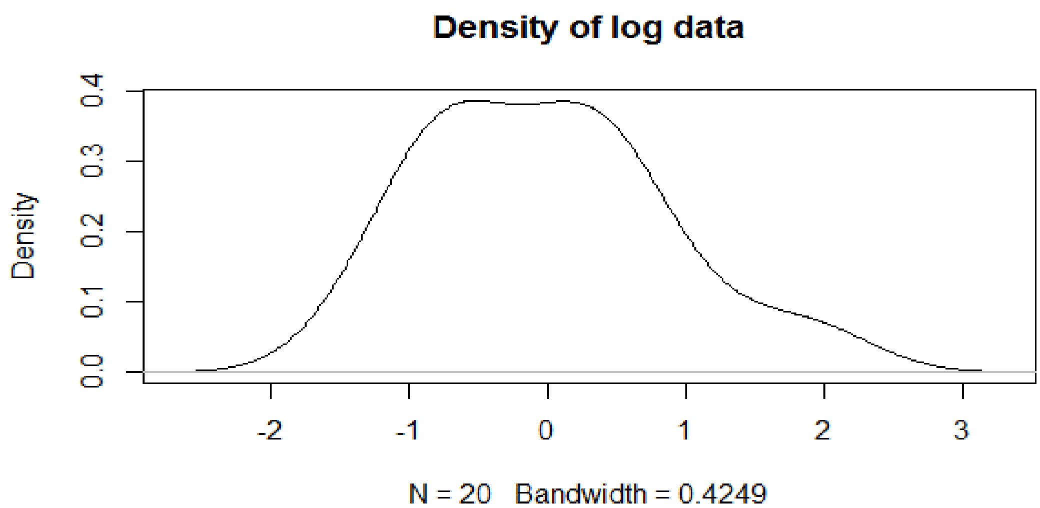
```
> sort(data)
 [1] 0.246 0.327 0.423 0.425 0.434 0.530 0.583 0.613 0.641 1.054 1.098 1.158
1.163
[14] 1.439 1.464 2.063 2.105 2.106 4.363 7.517
> (1.464+2.063)/2 - (.434+.530)/2 # Plug-in IQR
[1] 1.2815
> 1.2815 / sqrt(mean(data^2) - mean(data)^2) # IQR/SD
[1] 0.7675505
> qqnorm(data)
> plot(density(log(data)), main="Density of log data")
> qqnorm(log(data))
```

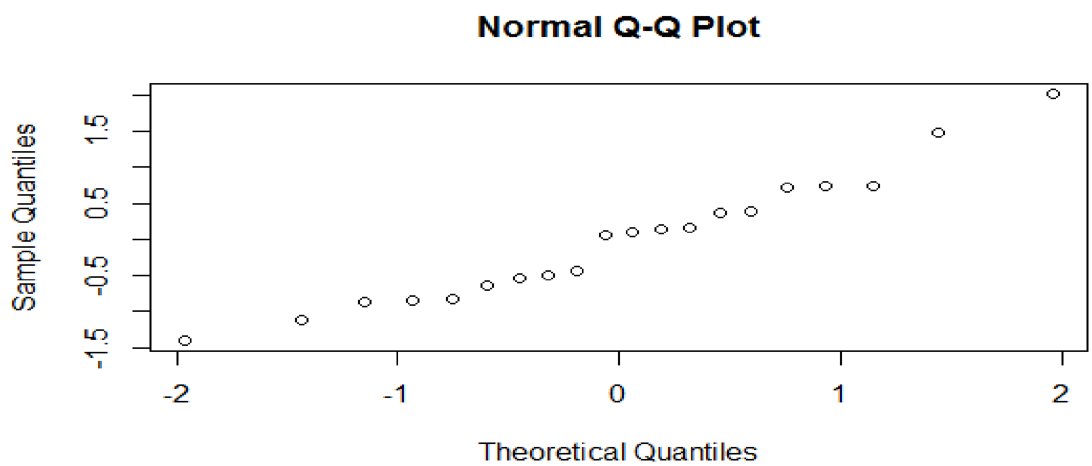
From execution of the above code we can see the following: The plug-in estimates are mean = 1. 49, variance = 2.79, median= 1.076, IQR= 1.28. The IQR/SD value is 0.77.

The plots are given below:







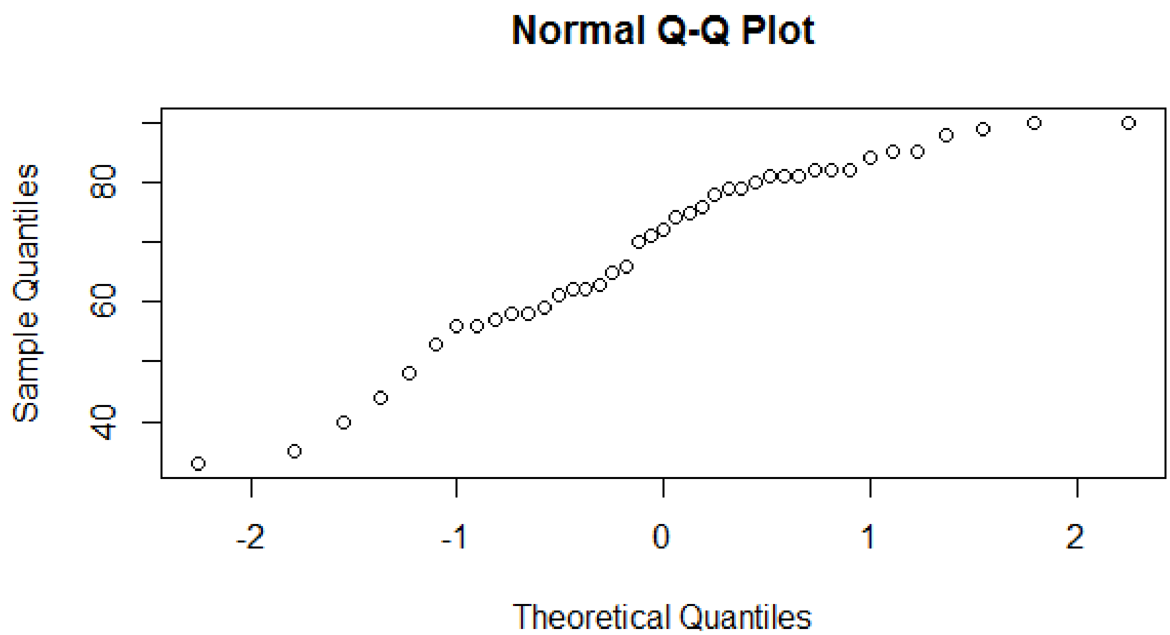


From the above plots, we can tell that the distribution is not normal as the QQ-plot is curved upwards. But from the QQ-plot of the log data, the log data is normally distributed as it is almost a straight line.

A3) Trosset Exercise 7.7.6:

The **R** code is as follows:

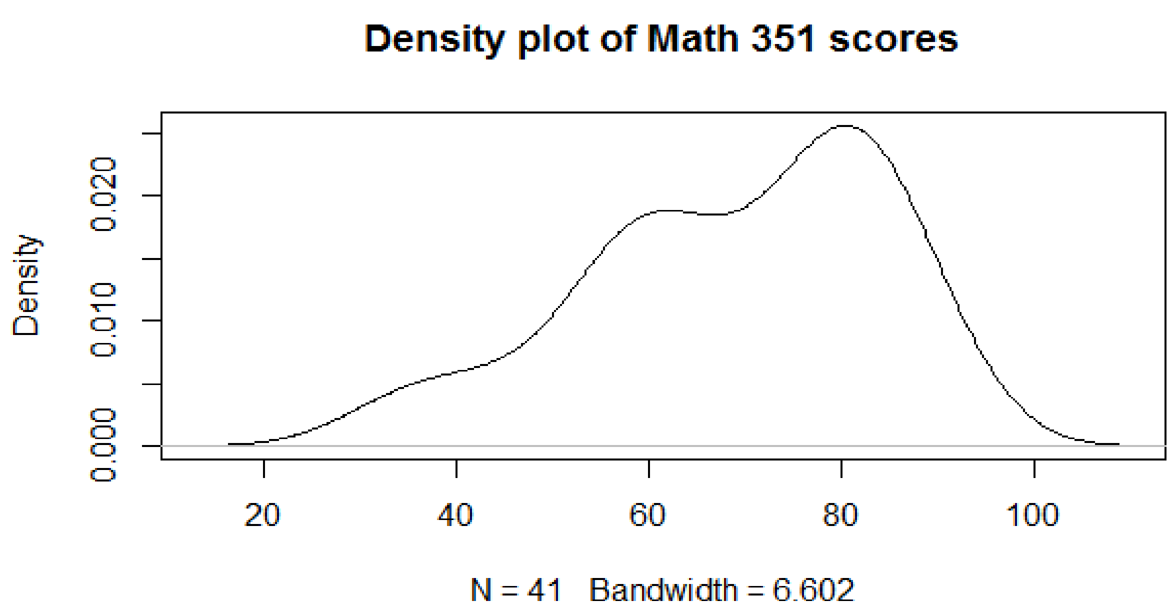
```
>scores=scan("http://mypage.iu.edu/~mtrosset/StatInfeR/Data/test351.da
t")
Read 41 items
> qqnorm(scores)
```



From the above plot we can see that it is not a normal distribution as the plot is curved.
The Density plot is given below:

```
> plot(density(scores), main="Density plot of Math 351 scores")
```

From the Density plot, we can easily see that the data is left-skewed. This is generally not the case with real data.



A4)

The distribution of the life of one battery (in hours) has mean 5, standard deviation 0.5, and variance 0.25.

We can use the Central Limit Theorem here.

$$\sum_{i=1}^{n} X_i = Normal(n\mu, n\sigma^2)$$

Here , mean 20×5 = 100, variance 20×0.25 = 5, and standard deviation √5.

Hence the probability that the battery will last 105 hours is as follows:

```
> 1-pnorm(105,100,sqrt(5))
[1] 0.01267366
```

A5)
a)Find EX.

$EX = (-2\times 0.3) + (-1\times 0.6) + (12\times 0.1) = 0.$

b)Find Var(X).

$EX^2 = (4\times 0.3) + (1\times 0.6) + (144\times 0.1) = 16.2$
$VarX = EX^2 - (EX)^2 = 16.2 - 0 = 16.2$

(c) 0.

(d) As it depends on *n*, the answer will be in terms of *n*.
$\sigma^2/n = 16.2/n.$

(e) Suppose n = 100.
We can Use **R** to find the approximate probability that X’ is greater than 0.5.

```
> 1 - pnorm(0.5, mean=0, sd=sqrt(16.2/100))
[1] 0.1070703
```

A6)

a)We can use **R** for calculating the mean.

```
> data = c(rep(1,27), rep(2,34), rep(3,16), rep(4, 13), rep(5, 6),
rep(6,3), 7)

> mean(data)
[1] 2.5
```

b)
```
> sd(data)
[1] 1.410638
```

c)
```
> sd(data)/(sqrt(100))
[1] 0.1410638
```

d)
```
> prob=pnorm(0.5, 0, sd(data)/sqrt(100)) - pnorm(-0.5, 0, sd(data)/sqr
t(100))

> prob
[1] 0.9996066
```

The probability is 99.96%.

e) Yes, we can be sure as we have seen above the error is less than 0.5 in more than 99% of the households. Hence here too it will be the same and we can easily say than the average household size is between 2 and 3.