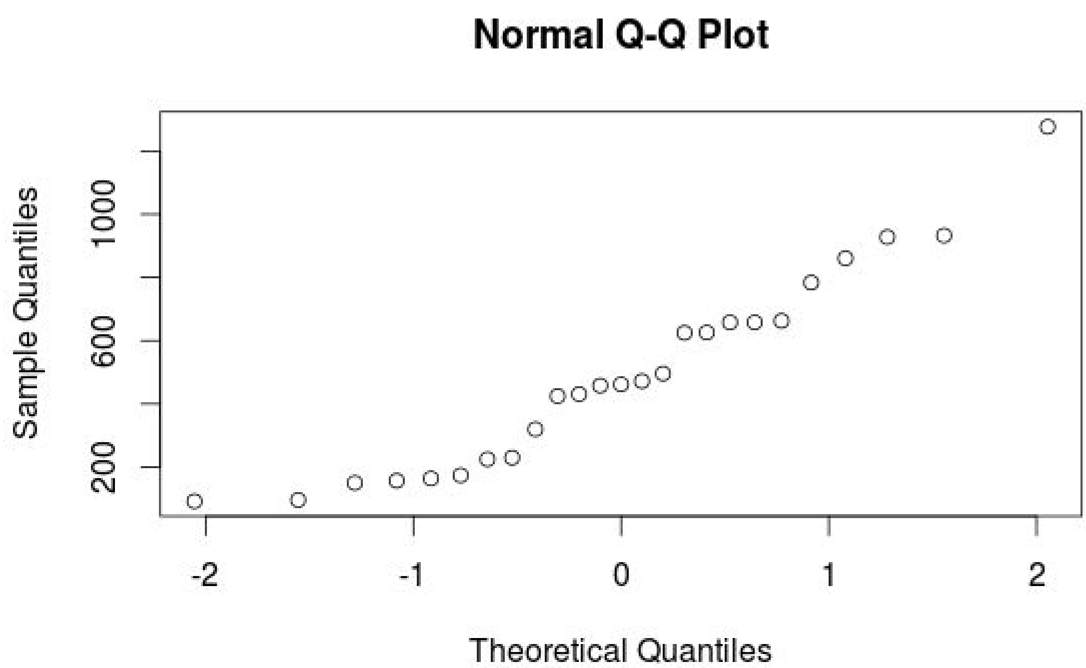
HOMEWORK 6
STAT-S 520
RAMPRASAD BOMMAGANTY

7.7.1:

(f) :
The code for normal probability plot is as follows:

```
x <-
c(462,425,164,784,625,472,658,658,663,928,92,230,96,626,1277,225,150,320,496,157,458,93
3,861,174,431)
qqnorm(x)
```

The plot is as follows:

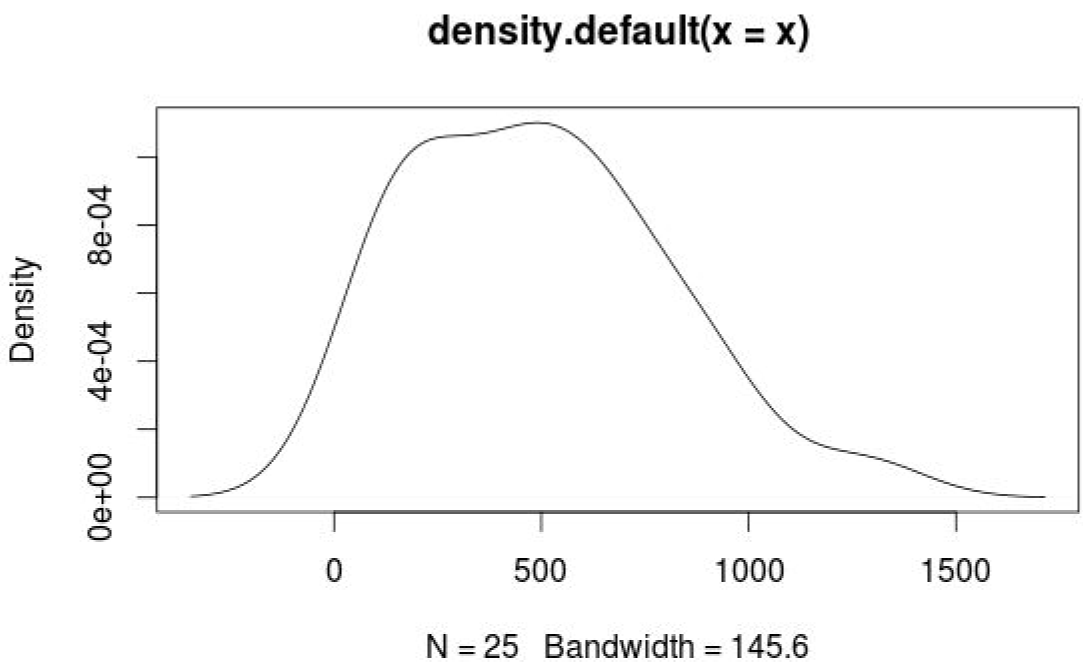


(g) :

The code for kernel density estimate is as follows:

```
x <-
c(462,425,164,784,625,472,658,658,663,928,92,230,96,626,1277,225,150,320,496,157,458,93
3,861,174,431)

plot(density(x))
```

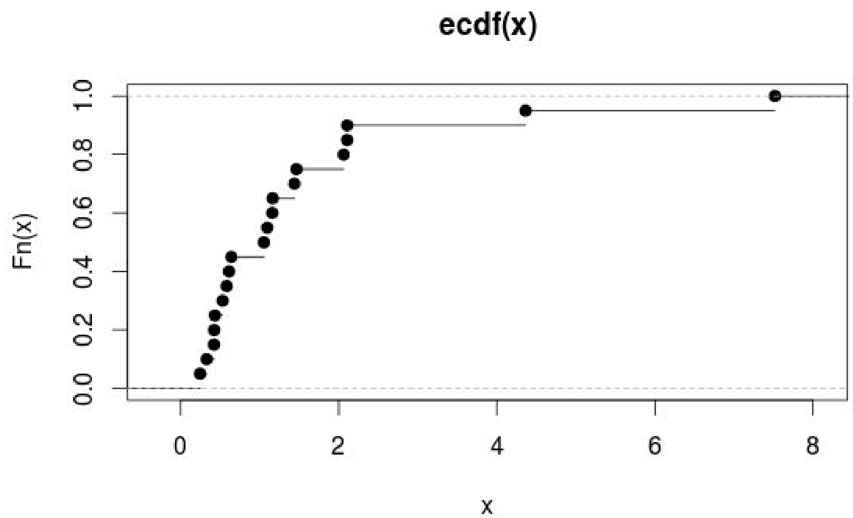


(h) : No, the sample was not drawn from a normal distribution. The reasons are as follows:

a. According to the normal probability plot, it is not linearly increasing as in the case of a sample taken from a normal distribution.
b. According to the kernel density estimate plot, it is a little left skewed and not symmetric.

7.7.4:

a. The plot for empirical cdf is as follows:
   R Code: plot(ecdf(x))



b. The R code for plug in estimate for mean and variance are:

```
x <-
c(0.246,0.327,0.423,0.425,0.434,0.530,0.583,0.613,0.641,1.054,1.098,1.158,1.163,1.439,1.464,
2.063,2.105,2.106,4.363,7.517)
n <- length(x)
plug.mean <- mean(x)
plug.var <- mean(x^2) - plug.mean^2
print(plug.mean)
print(plug.var)
```

Output of plug in mean : 1.4876

Output of plug in variance: 2.787554

R code for Median: median(x)

Output of plug in median: 1.076

R code for inter-quartile range: IQR(x)

Output of inter quartile range: 1.10775

c. The R code for calculating the square root of plug in variance is as follows:

```
sqrt(plug.var)
```

The output is as follows: 1.669597

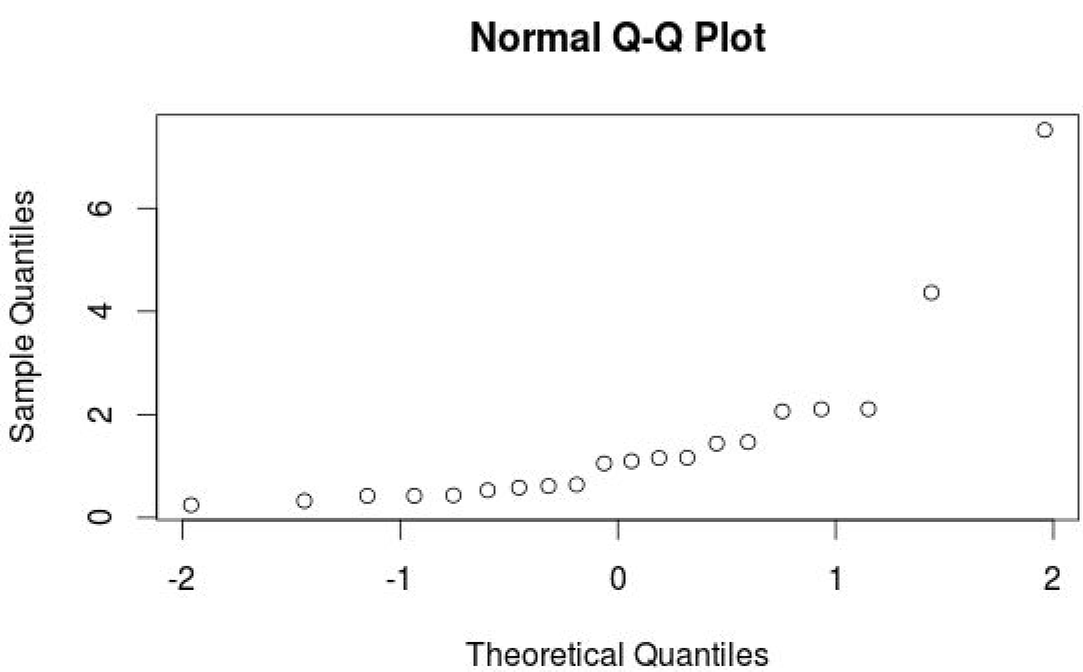
The square root of plug in variance is higher than the value of plug in estimate of inter-quartile range.

The ratio of IQR to Standard deviation is : 1.10775/ sqrt(2.7875) = 0.66

No, the sample was not taken from a normal distribution. The reason being: the ratio of IQR to SD is 0.66.

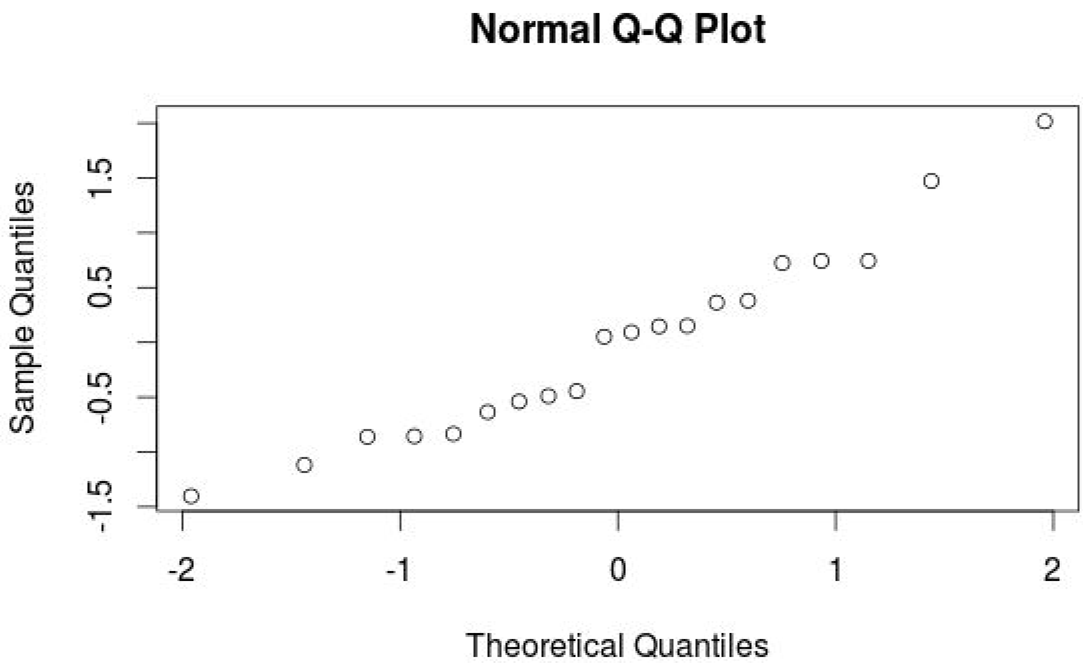
d. The R code for calculating the normal probability plot is as follows: qqnorm(x)

The output is as follows:

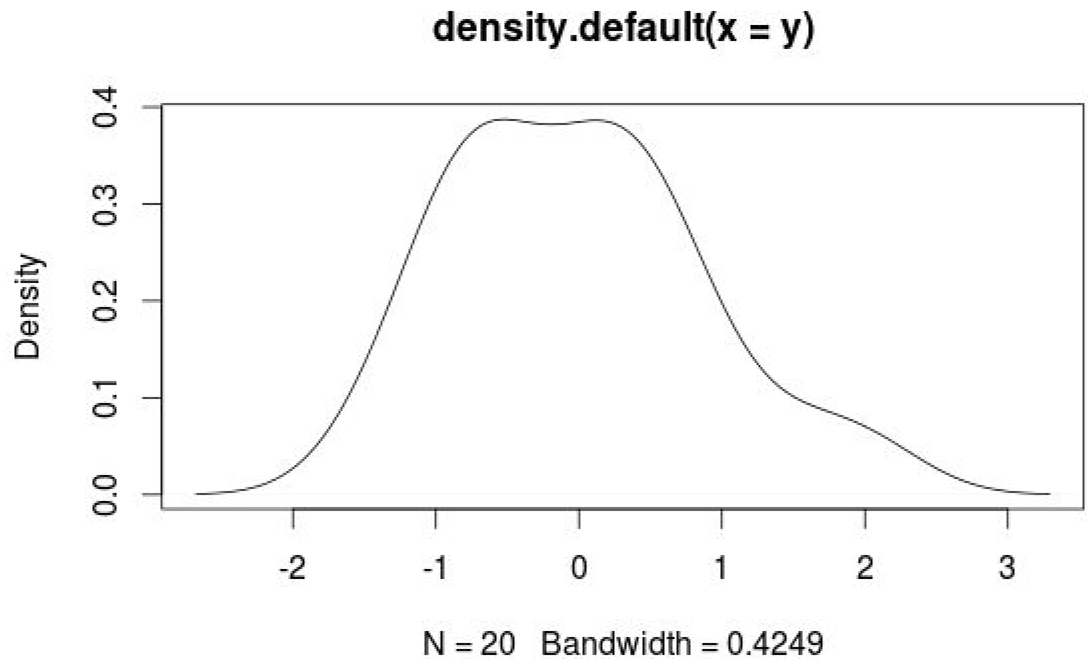


No, the sample was not taken from a normal distribution. The reason being : The plot is not as straight as a that of a normal distribution.

e. The normal probability plot of y = log(x) is as follows:
R code: qqnorm(y)



The kernel density estimate plot is as follows:
R code: plot(density(x))

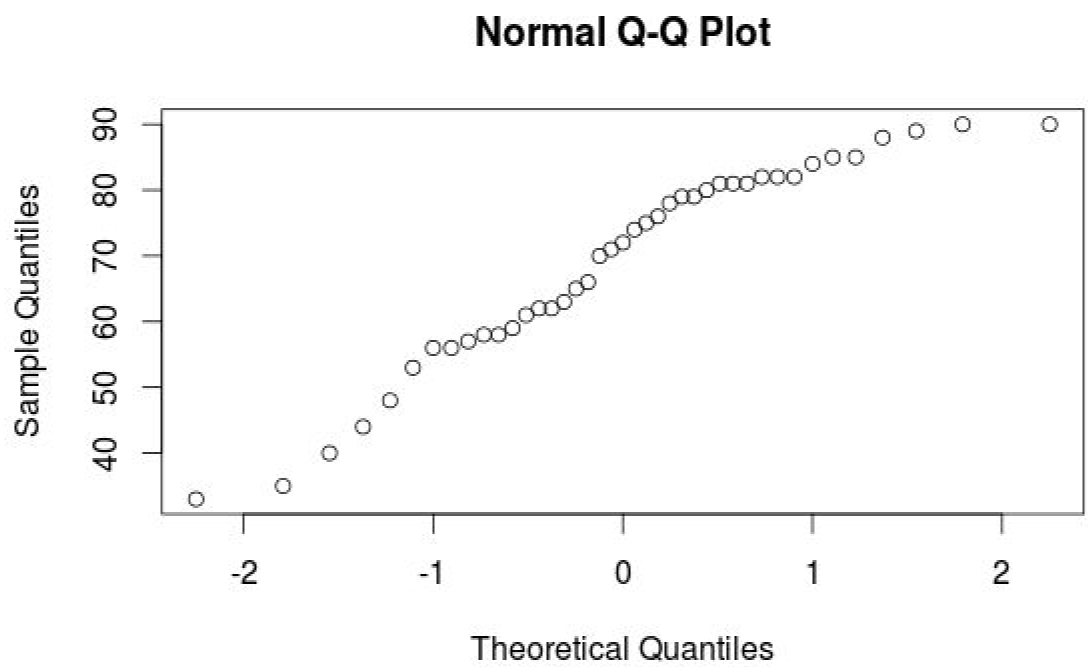


Yes, y might have been drawn from a normal distribution. The normal probability plot and the kernel density plots are more or less similar to that of a normal distribution.

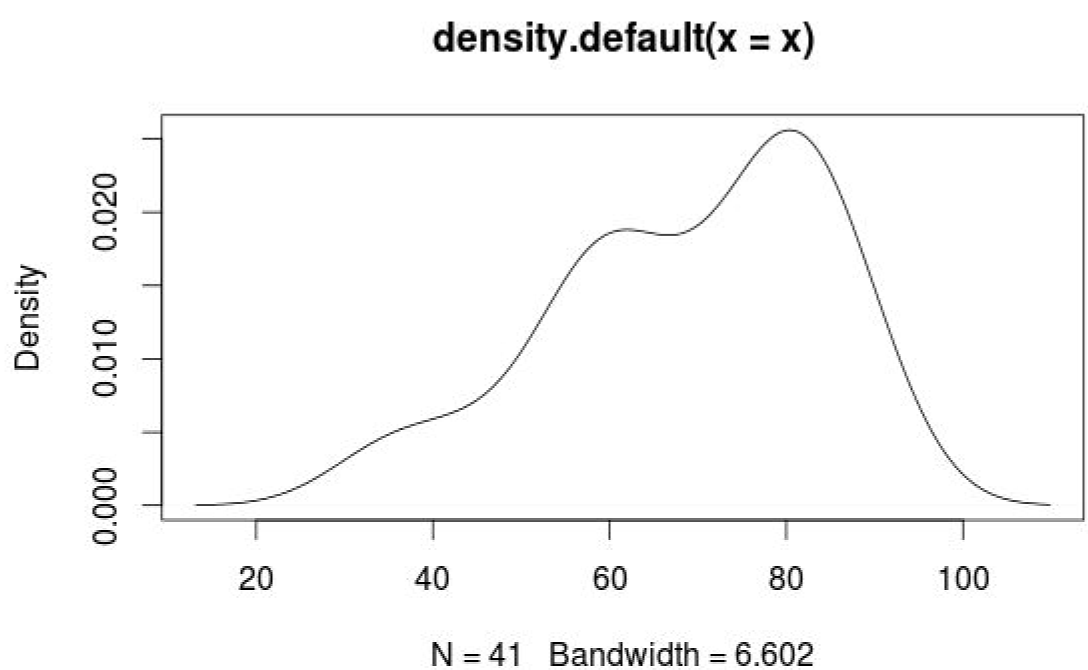
7.7.6::

a. The normal probability plot for the sample of number is as follows:
   R code:

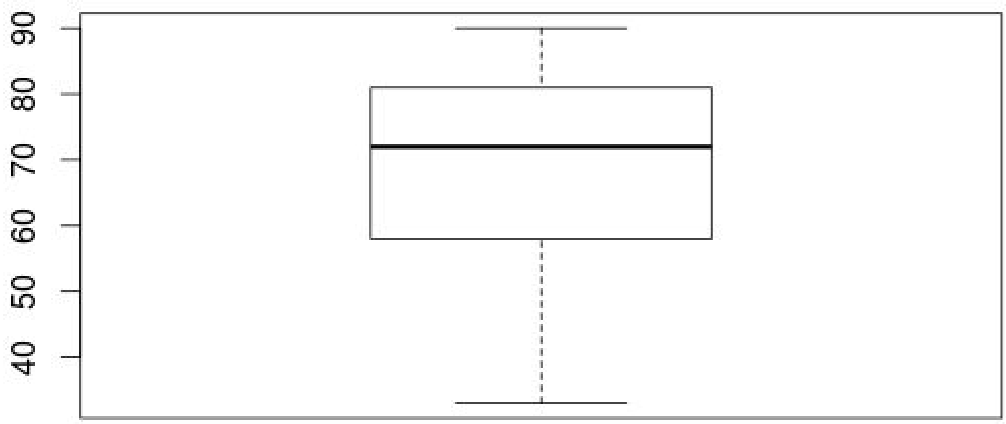
```
x <-
c(90,90,89,88,85,85,84,82,82,82,81,81,81,80,79,79,78,76,75,74,72,71,70,66,65,63,62,62,61,59,
58,58,57,56,56,53,48,44,40,35,33)
qqnorm(x)
plot(density(x))
```



The kernel density estimate is as follows:



The box plot of the number is as follows:



No, these numbers do not appear to be a random sample from a normal distribution.

b. The interesting anomalies with respect to the sample of numbers are: The density plot is heavily left skewed.

8.4.4::

Given mean =5, sd = 0.5.

Therefore, variance = 0.25

We need to find out the probability that the batteries last at least 105 hours.

Since Chris buys twenty 2 packs, the new values are as follows:

Mean = 20 * 5 = 100
Variance = 20 * 0.25 = 5
Standard Deviation = sqrt(5)

The R code is as follows:

```
1 - pnorm(105,100,sqrt(5))
```

The output is as follows: 0.01267366

5:
Answer:

a . EX = (0.3)* (-2) + 0.6 * (-1) + 0.1 * 12 + 0 = 0
b. Var (X)

EX^2 = (0.3)* (4) + 0.6 * (1) + 0.1 * 144 + 0 = 16.2

Var(X) = 16.2 - 0 = 16.2

c. Expected value of the sample mean distribution is also 0.

d. Variance of sample mean distribution = 16.2/n.

e. The R code is as follows:

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

The output is : 0.1070703

6::
Answer:

a. R code is as follows:

```
sample <- c(rep(1,27),rep(2,34),rep(3,16),rep(4,13),rep(5,6),rep(6,3),rep(7,1))
mean(sample)
```

Output : 2.5

b. R code is as follows:

```
plug.var <- mean(sample^2) - mean(sample)^2
sqrt(plug.var)
```

Output : 1.403567

c. Standard error of the sample mean = 1.41 / sqrt(27+34+16+13+6+3+1) = 0.141

d. R code for probability that absolute error of the survey is less than 0.5 :

```
pnorm(0.5,0,sqrt(plug.var)/sqrt(100)) - pnorm(-0.5,0,sqrt(plug.var)/sqrt(100))
```

Output : 0.9996325

e. Yes, we can be sure that the average household size is between 2 and 3 because there is a probability of 0.99 that the error in the survey is less than 0.5.