David O Neill - 0813001 - Assignment 2 - MA4413

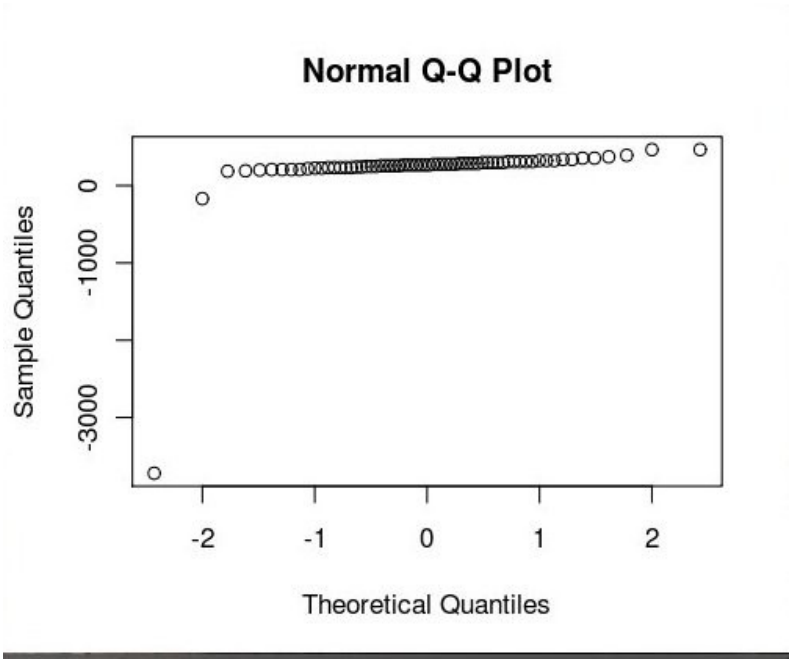
Q1.

The data was adjusted to display it in a readable and plotteable form.

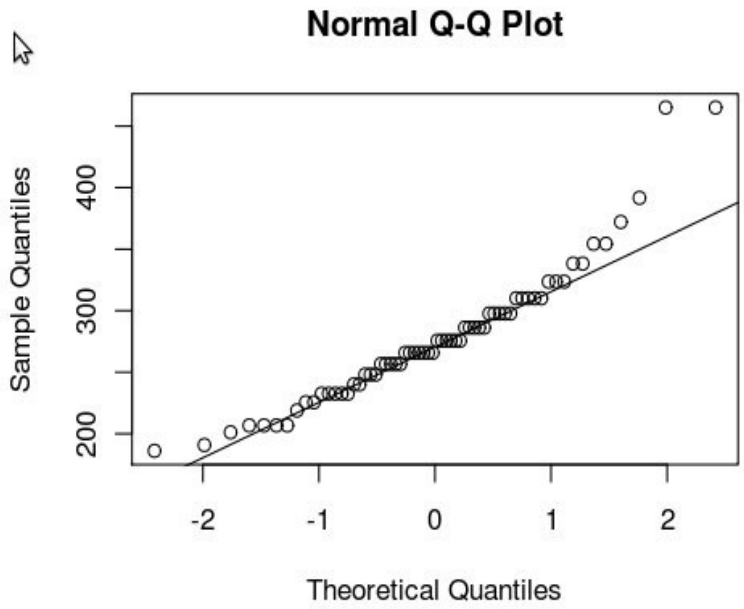
```
# stem show negatively skewed data
#   -4 | 4
#   -3 |
#   -2 |
#   -1 |
#   -0 | 2
#    0 |
#    1 | 669
#    2 | 011223334444455555666667777778888888899999
#    3 | 0001122222334666679
#    4 | 0
```

Q2

Becuase of the error in the data the standard deviation for this dataset indicates it is not normally distributed



our initial assumption of negatively skewed data set was wrong by removing the error we can now see it is positively skewed

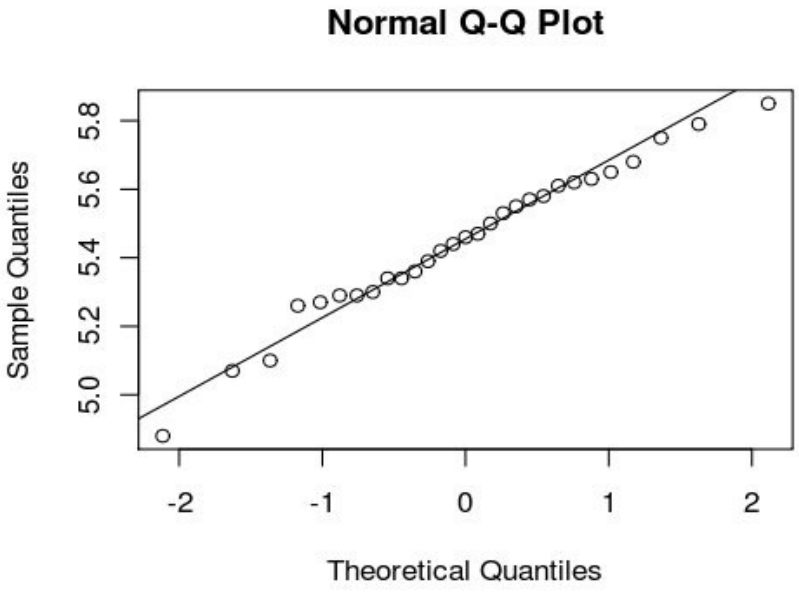


By plotting the values the data minus the negative values we get a more normal linear data indicated by the diagram.

Q3

---

yes normal distribution



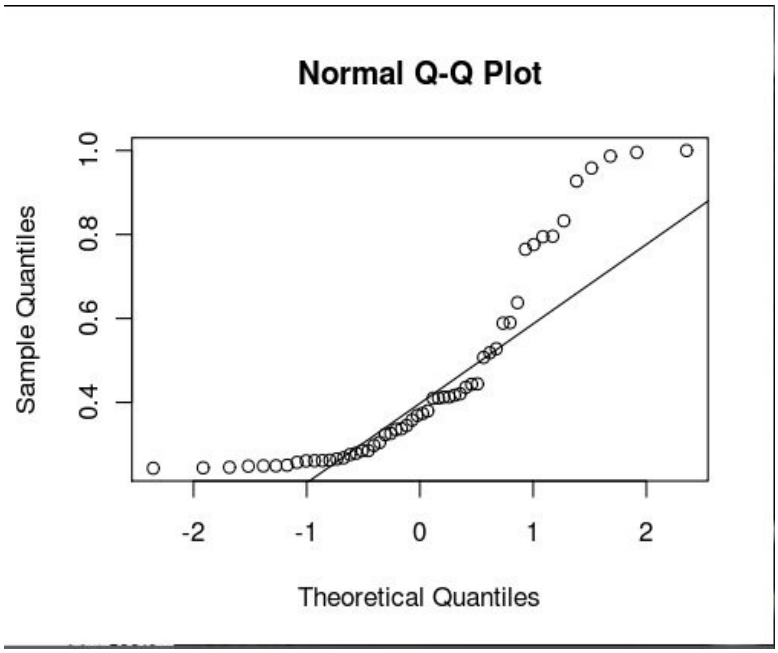
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
4.880	**5.300**	**5.460**	**5.448**	**5.610**	**5.850**

Using a box plot we can see the **50%** of the data is between **5.3 - 5.6**
getting the mean of these would be **5.45**
which is the close to the mathematical mean of the data.

Q4

Gambia had the largest increase of **56.58%**
Nigeria's population of **1987 at 106700** people is greatly above the mean at **11420**

Our initial view of the data does not indicate normal distribution.



On further analysis by analyzing the interquartile
Omitting the the upper and lower (tails) we can
formalize our hypothesis that the distribution is normal .

Using the **shapiro test** we can obtain a **p-value** greater than **0.05**

(0.08373) indicating
that **h0** is of normal distribution
the tails or lower and higher quartiles of the distibution

~ **50% p-value = 0.08373**
~ **55% p-value = 0.0688**
~ **60% p-value = 0.05783**

Q5

The natural log on the negatively skewed dstribution can adjust the values to more accurately describe a normal distrubtion (indicated by the HIST and qqnorm) .
The sample quantiles of the natural log is twice that of the base 10 log . In a trace analysis, using a logarithm as a transformation would adjust the values
into a noramaly distrubuted fashion.

Q6

Using the Normal Distribution Calulations
mean **3400,** SD = **550**

0.0509 (5.09%) of the babies are below **2500g.**
0.5721 (57.21%) of the babies are below **3500g.**
0.1377 (13.77%) of babies are greater than **4000g.**

Q7

Using the Normal Distribution Calulations
mean = **3400,** SD = **550**

The babies whose weight above **90% (10%) is 4105g.**
The babies below **1% is 2121g.**
The babies above **95 (5%) is 4305g.**
First quartile is **3029g.**

The second quartile is **3400g (mean)** as expected in a gaussian normal distribution.
The third quartile is **3771g.**

Q8

---

mean = **266,** SD = **16**
Using the Normal Distribution Calulations

Babies percentage born before **290 (93.32%)**
subtracting babies born before **250 days 0.1587 (15.87%)**

babies between **250 and 290 days= 77.45%**
babies after **205 days 0.0074%**

The middle **95%** of pregnancies last between **234.6 and 297.4 days**
At least **296.1 days** require special care