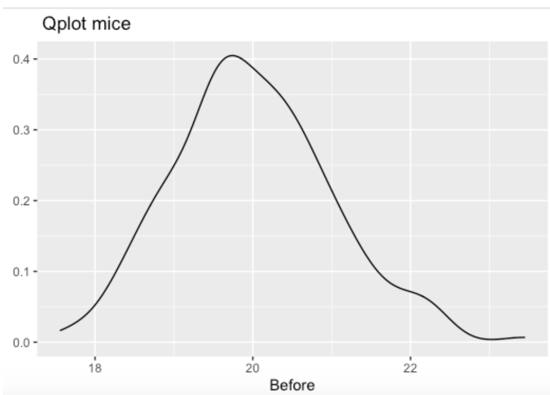
HYPOTHESIS TESTING IN R
Statistical Techniques In r
(A university of Greenwich london assignment)

TASK 1 :

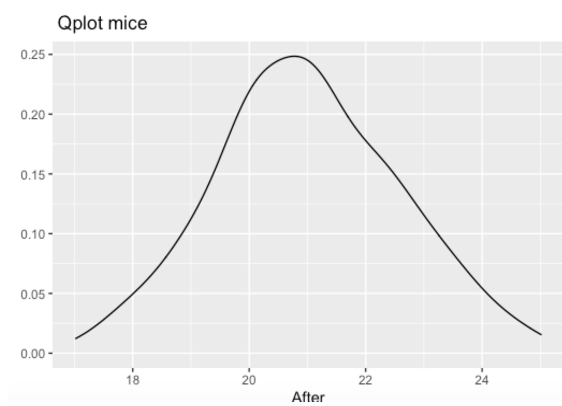
1. The mice dataset is a normal distribution data generated by the function **rnorm.**
   (weights Before - mean 20,vd=1)
   (weights After - mean 21,vd=2.5).
   The rats dataset is a weibull distribution data generated by the function **rweibull.**
   (Rats before - shape=10,scale=20)
   (Rats After- shape=9,scale=21.)

2. Using Qplot with geom='density', the following distribution of the data is observed as follows:

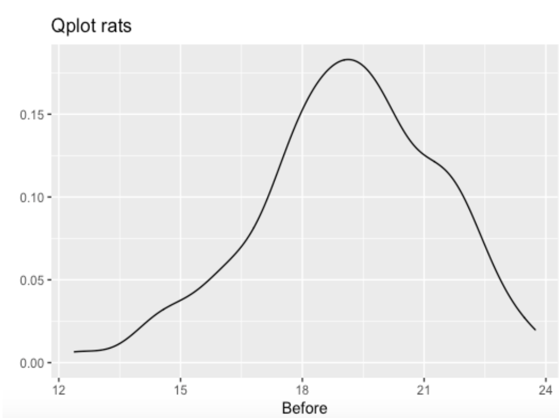
Mice Before



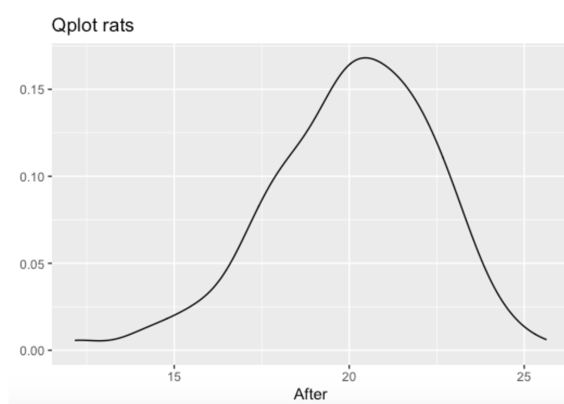
Mice After



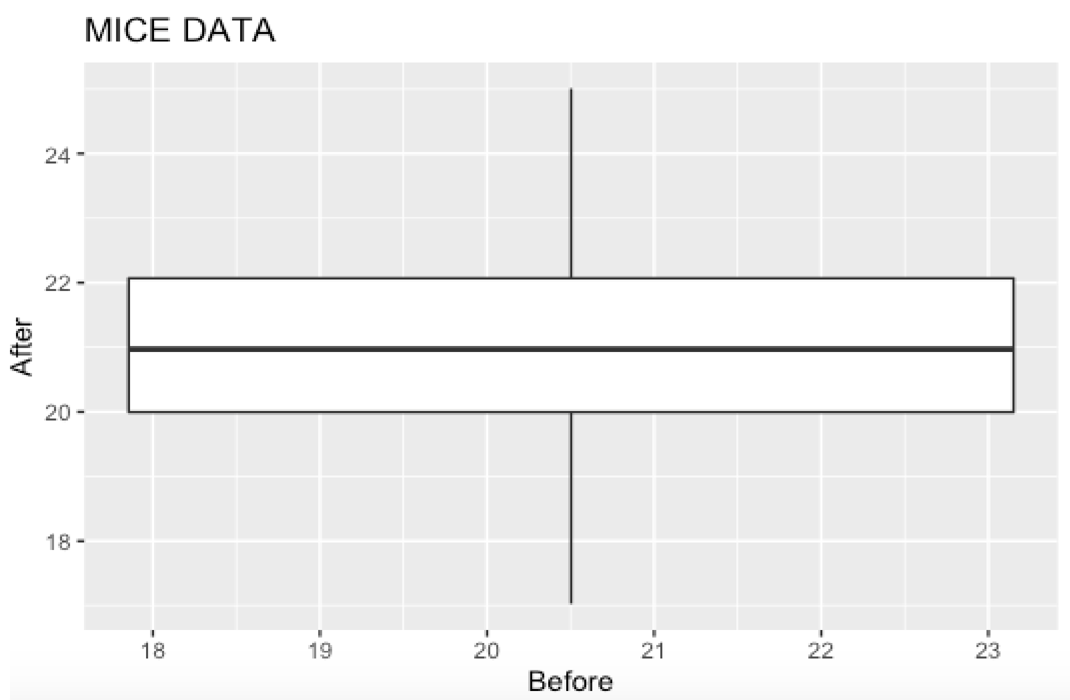
Rats Before

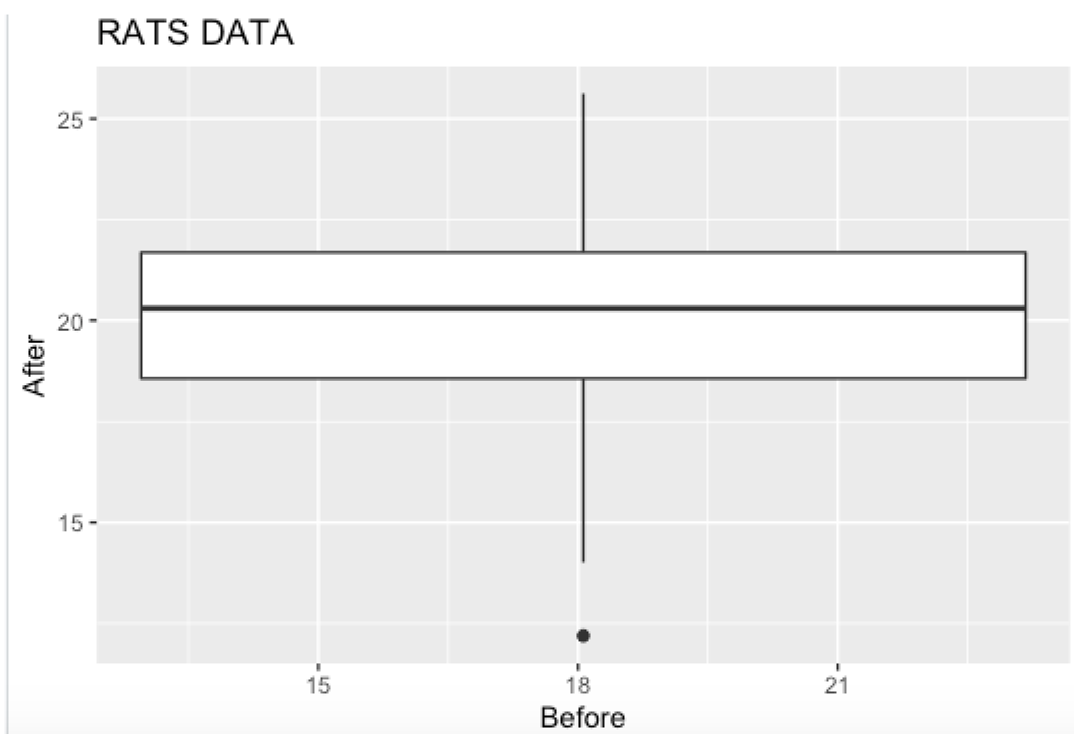


Rats After



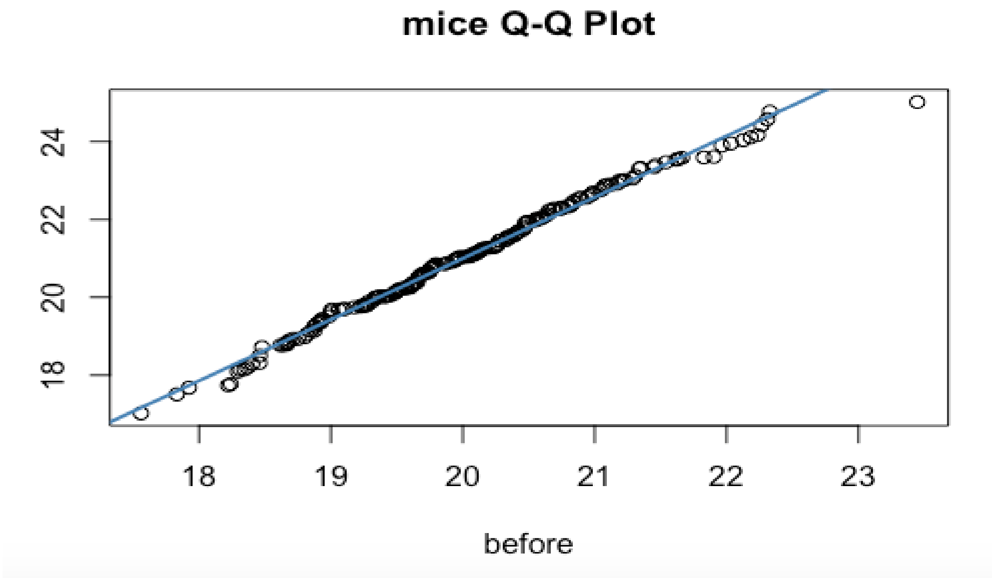
3. The Boxplot of the both the datasets is as follows:





TASK 2:

1. **For the mice data:**
   Normality qualitatively test:



The graph is almost a straight line for the mice data, Hence It passes the normal qualittaively test.

Shapiro wilk test:

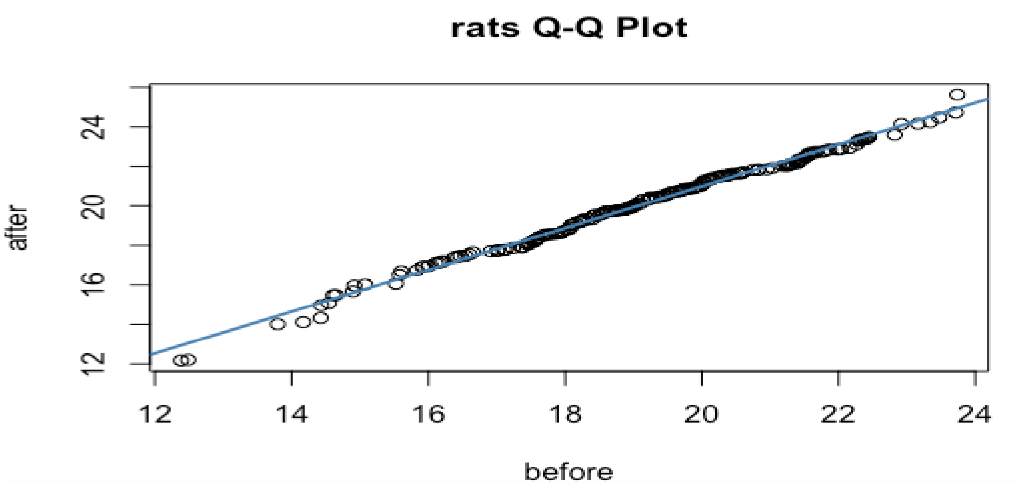
A random sample of 10 observations have been picked from the mice data and the result is observed as follows:

For both the columns(Before and After)The p values observed **are 0.2298 and 0.7776 which are greater than 0.05.**

Hence mice data passes the Shapiro wilk test.

```
> shapiro.test(mice$Before)

        Shapiro-Wilk normality test

data:  mice$Before
W = 0.99076, p-value = 0.2298

> shapiro.test(mice$After)

        Shapiro-Wilk normality test

data:  mice$After
W = 0.99519, p-value = 0.7776
```

2. **For the rats data:**
   Normality qualitatively test:



The graph in this case seems to be a straight line, hence it **passes normality quantitatively test:**

Shapiro wilk test:
A random sample of 10 observations have been picked and the result is observed as follows:
The p value for both the columns(Before and After) **is 0.04 and 0.006 which is less than 0.05.**

Hence Rats data fail Shapiro wilk test suggesting they do not come from a normal distribution.

```
> shapiro.test(rats$Before)

        Shapiro-Wilk normality test

data:  rats$Before
W = 0.98589, p-value = 0.04332

> shapiro.test(rats$After)

        Shapiro-Wilk normality test

data:  rats$After
W = 0.98003, p-value = 0.006012
```

Mice data passes both quantitative and qualitative tests whereas rats pass only quantitative test.

TASK 3 :

Paired T- test for mice data:

Paired t-test was performed on the mice data and the following observations were made:

```
> res$statistic
        t
-7.470234
> res$p.value
[1] 2.469652e-12
> res$conf.int
[1] -1.2639752 -0.7360248
attr(,"conf.level")
[1] 0.95
> res$estimate
mean difference
             -1
> res$parameter
 df
199
```

Absolute value of T (7.47) is greater than critical value of 0.05 level significance and 199 df, **we reject the null hypothesis.** Also p value<0.000001 strongly corroborates our claim to reject null hypothesis

Therefore we conclude that the mean difference between the data is -1 i.e there is significant difference between the means of the two populations, at given level of confidence.

Wilcox test for Rats data: Wilcox test was performed on rats data and following observations were made:

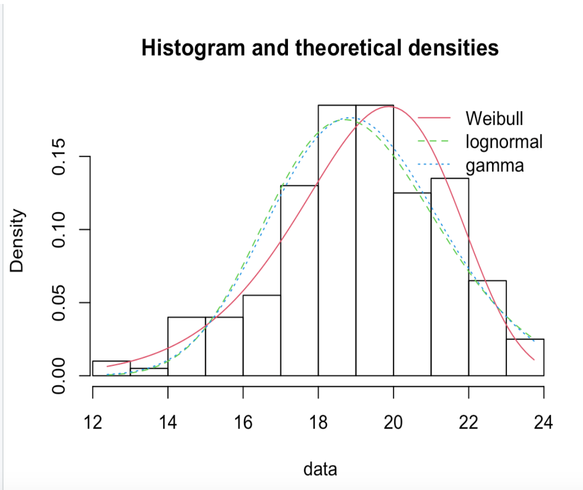
```
> res1 <- wilcox.test(rats$Before, rats$After, paired = TRUE)
> res1

        Wilcoxon signed rank test with continuity correction

data:  rats$Before and rats$After
V = 6503, p-value = 1.509e-05
alternative hypothesis: true location shift is not equal to 0

> res1$statistic
   V
6503
> res1$p.value
[1] 1.509253e-05
>
```

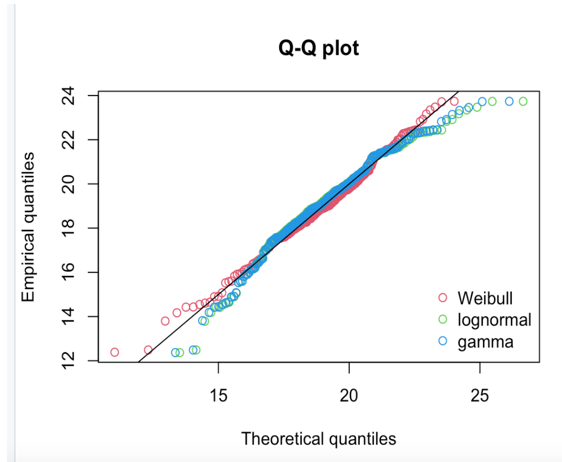
From the given p value which is conspicuously less than 0.05 , we reject the null hypothesis **the median of the population of differences between the paired data is zero.**
the distribution of one population is shifted to the left or right of the other

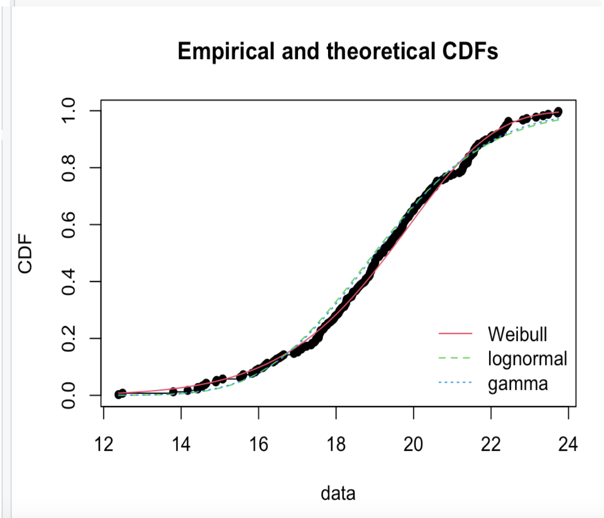
Task 4 :

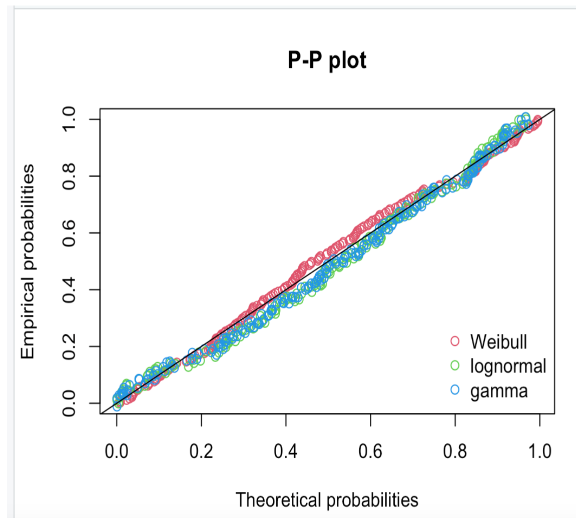
Fitting lognormal,weibull and gamma distribution on rats data.

Before Treatment : Various distributions have been fit on data and their comparisons can be studied from the below graphs.









After Treatment :