T-Test

R Markdown

#Introduction#

```
library(tidyverse)
library(dplyr)

#load data
cancerPredictionDS <- read_csv("C:\\Users\\faqeerrehman\\MSU\\OneDrive - Montana State University\\Rese
head(cancerPredictionDS)
```

```
## # A tibble: 6 x 4
##    Source   MR0  MR12  MR21
##     <dbl> <dbl> <dbl> <dbl>
## 1  0.219 0.13  0.243 0.274
## 2  0.209 0.127 0.153 0.205
## 3  0.105 0.13  0.178 0.237
## 4  0.228 0.117 0.22  0.223
## 5  0.117 0.118 0.165 0.191
## 6  0.176 0     0.279 0.14
```

Table 1: Summary of Data

	n	Min	1Q	Median	Mean	Q3	Max	Standard Deviation
Source	100	0.071	0.1490	0.1795	0.18454	0.21700	0.296	0.0503662
MR0	100	0.000	0.1170	0.1245	0.11777	0.13325	0.150	0.0339800
MR12	100	0.093	0.1590	0.1855	0.18990	0.22000	0.294	0.0449160
MR21	100	0.068	0.1455	0.1810	0.18268	0.22000	0.281	0.0480040

#Data Statistics
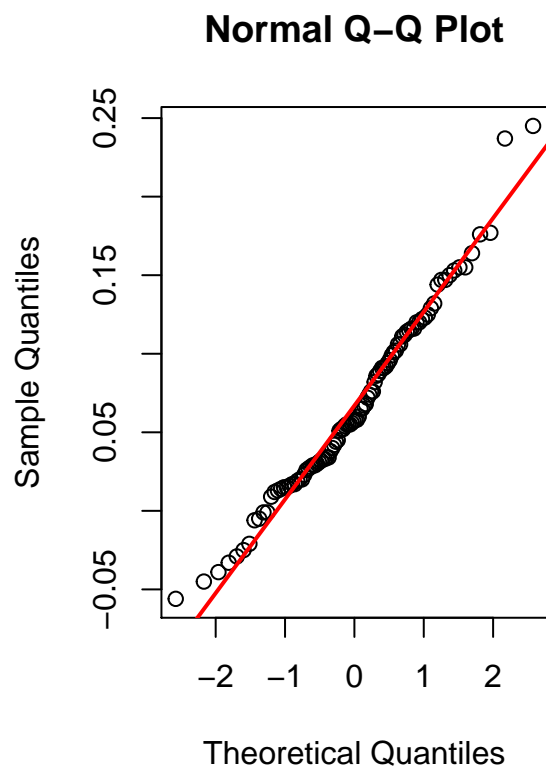
######################Mutant1

#Mutant1 - MR0

##MR0 Failed##

```
diffBwSrcFlw = cancerPredictionDS$Source- cancerPredictionDS$MR0
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
```

```
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot


```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw1)$out
```

In the boxplot, it can be seen that there are two potential outliers. Therefore, we have provided the results with and without the outliers. It can be seen that for both cases, we have strong evidence to suggest that there is difference in the source and follow-up executions (p-value<0.0001).

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
```

```
## data:  diffBwSrcFlw
## t = 11.404, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.05515279 0.07838721
## sample estimates:
## mean of x
##   0.06677
```

```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

```
t.test(diffBwSrcFlw[-c(14, 97)],mu=0,conf.level=0.95,alternative="two.sided")
```
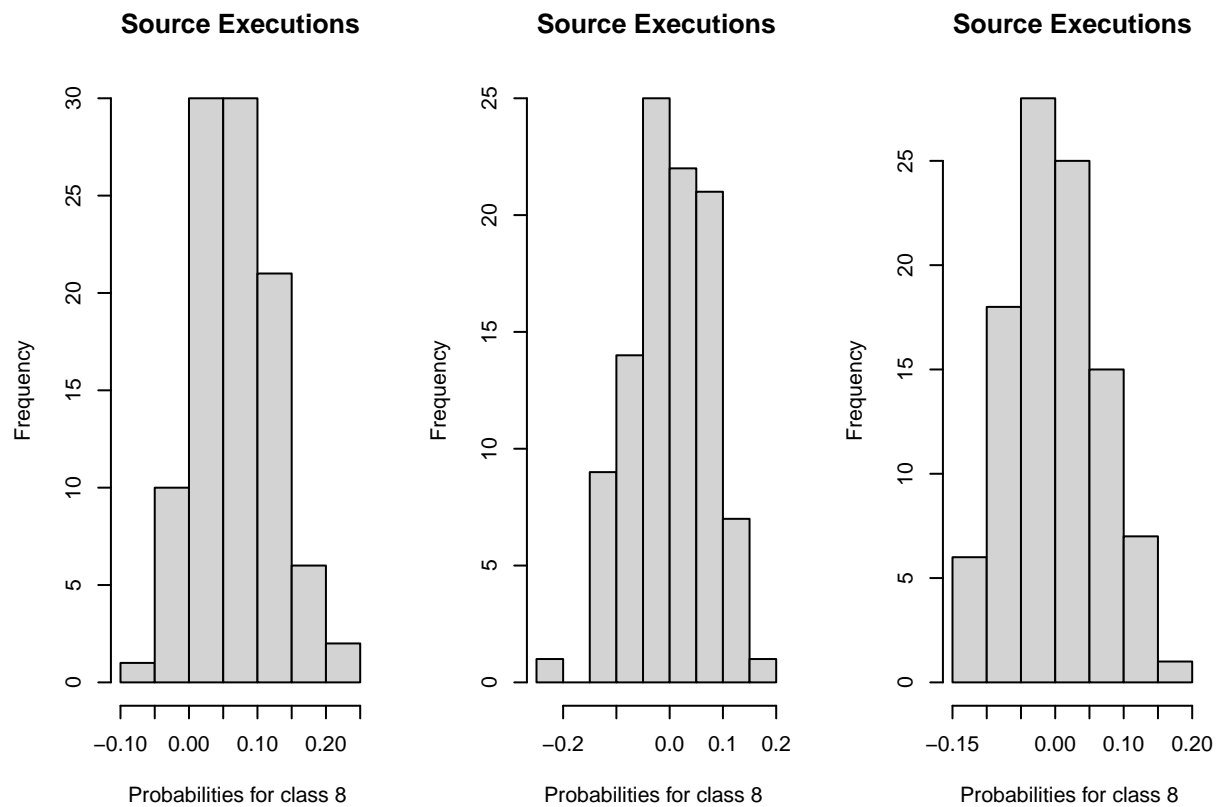
```
##
##  One Sample t-test
##
## data:  diffBwSrcFlw[-c(14, 97)]
## t = 11.703, df = 97, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.05249327 0.07393530
## sample estimates:
##  mean of x
## 0.06321429
```

```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

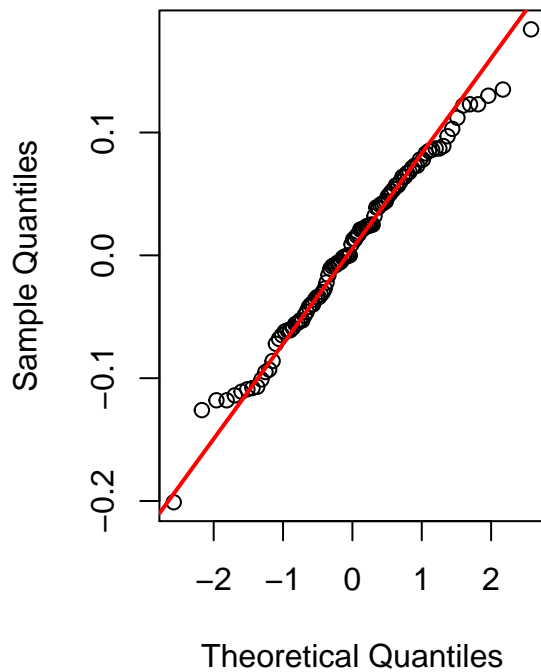#Mutant1 - MR12 ##MR12 Passed##

```
diffBwSrcFlw = cancerPredictionDS$MR12 - cancerPredictionDS$Source
```

```
hist(cancerPredictionDS$MR12 - cancerPredictionDS$Source,
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

Source Executions	Source Executions	Source Executions

Probabilities for class 8

```r
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot


```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  diffBwSrcFlw
## t = 0.74838, df = 99, p-value = 0.456
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.008851242  0.019571242
## sample estimates:
## mean of x
##    0.00536
```
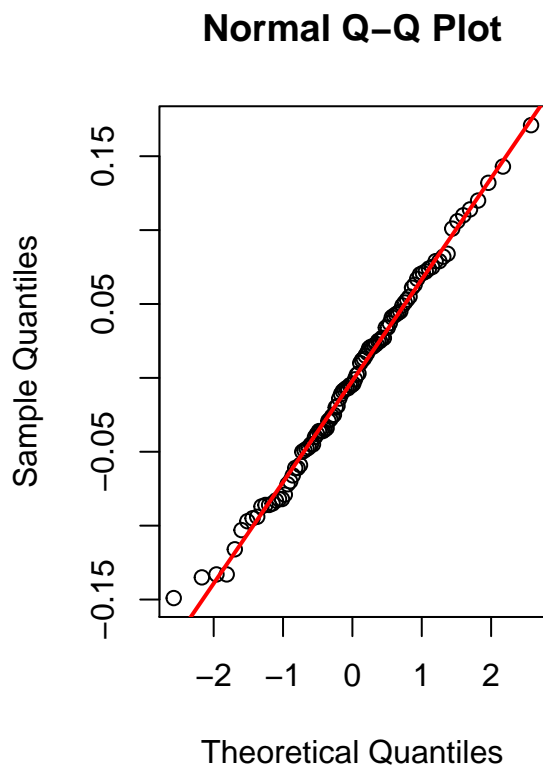
```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

#Mutant1 - MR21 ##MR21 Passed##

```
diffBwSrcFlw = cancerPredictionDS$MR21 - cancerPredictionDS$Source
```

```r
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

```r
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot


```r
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```r
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
```

```
## data:  diffBwSrcFlw
## t = -0.27351, df = 99, p-value = 0.785
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.01535362  0.01163362
## sample estimates:
## mean of x
##  -0.00186
```

```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

######################Mutant5

#Mutant5 - MR0

##MR0 Failed##

```
library(tidyverse)
library(dplyr)

#load data
cancerPredictionDS <- read_csv("C:\\Users\\faqeerrehman\\MSU\\OneDrive - Montana State University\\Rese
head(cancerPredictionDS)
```

```
## # A tibble: 6 x 4
##    Source   MR0  MR12  MR21
##     <dbl> <dbl> <dbl> <dbl>
## 1  0.718 0.117 0.638 0.552
## 2  0.942 0.109 0.641 0.002
## 3  0.737 0.103 0.629 0.866
## 4  0.108 0.129 0.834 0.001
## 5  0.585 0.119 0.033 0.984
## 6  0.355 0.121 0.273 0.818
```

```
diffBwSrcFlw = cancerPredictionDS$Source- cancerPredictionDS$MR0
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
     main="Source Executions",
     xlab= "Probabilities for class 8 ",breaks=10)
```

Applied log transformation but normality assumption is violated more by transformed data than original data, so we will use the original data.

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```
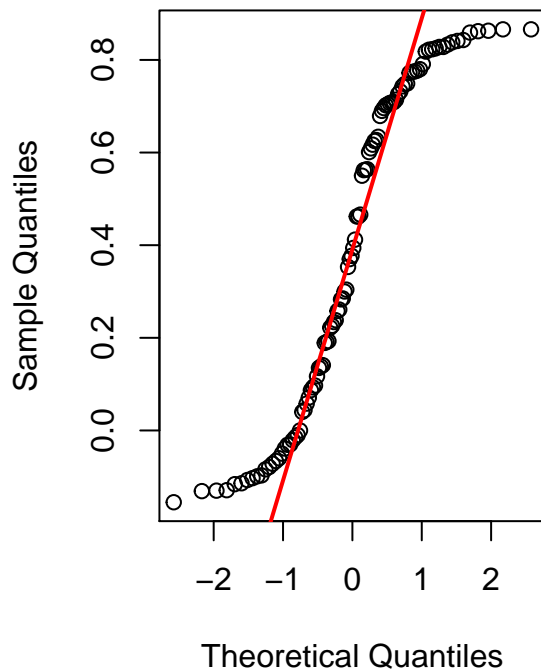
Normal Q–Q Plot


```r
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```r
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  diffBwSrcFlw
## t = 11.112, df = 99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.3204064 0.4597136
## sample estimates:
## mean of x
##   0.39006
```

```r
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```
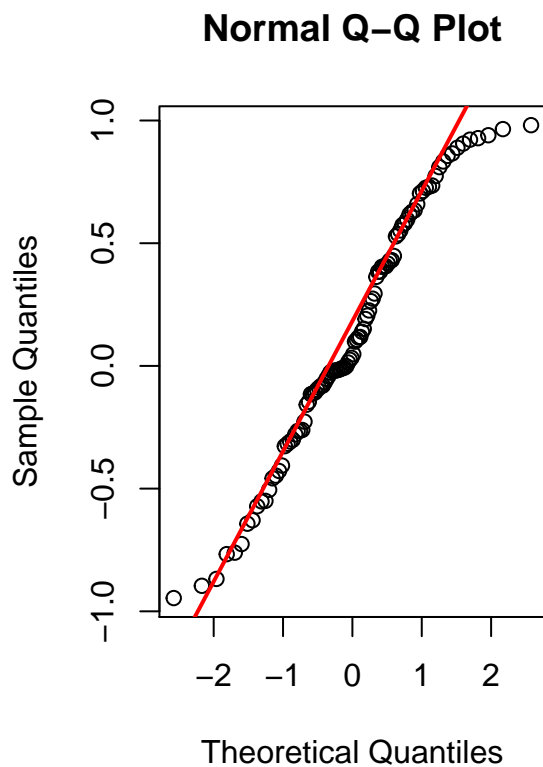
#Mutant5 - MR12 ##MR12 Failed##

```
diffBwSrcFlw = cancerPredictionDS$MR12 - cancerPredictionDS$Source
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot


```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
```

```
## data:  diffBwSrcFlw
## t = 2.5545, df = 99, p-value = 0.01216
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.02804581 0.22319419
## sample estimates:
## mean of x
##   0.12562
```
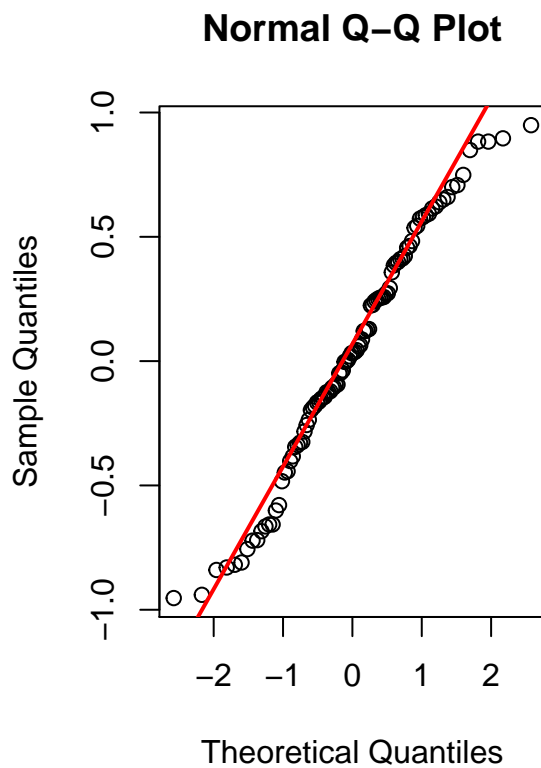
```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

#Mutant5 - MR21 ##MR21 Passed##

```
diffBwSrcFlw = cancerPredictionDS$MR21 - cancerPredictionDS$Source
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot

```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  diffBwSrcFlw
## t = 0.69811, df = 99, p-value = 0.4867
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.06125517  0.12775517
## sample estimates:
## mean of x
##    0.03325
```

```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

######################Mutant5

#Mutant4 - MR0

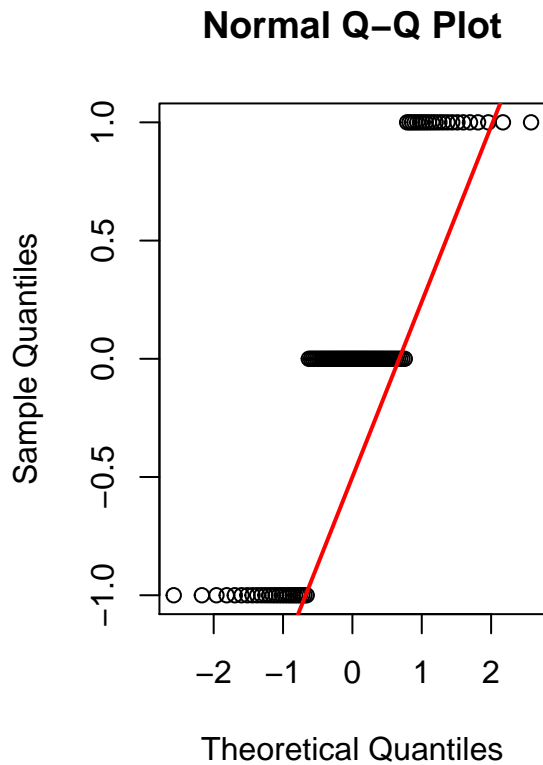##MR0 Passed (Normality assumptions violated)##

```
library(tidyverse)
library(dplyr)

#load data
cancerPredictionDS <- read_csv("C:\\Users\\faqeerrehman\\MSU\\OneDrive - Montana State University\\Rese
#view(cancerPredictionDS)
```

```
diffBwSrcFlw = cancerPredictionDS$Source - cancerPredictionDS$MR0
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
     main="Source Executions",
     xlab= "Probabilities for class 8 ",breaks=10)
```

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot


```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  diffBwSrcFlw
## t = -0.57542, df = 99, p-value = 0.5663
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.17793267  0.09793267
## sample estimates:
## mean of x
##     -0.04
```
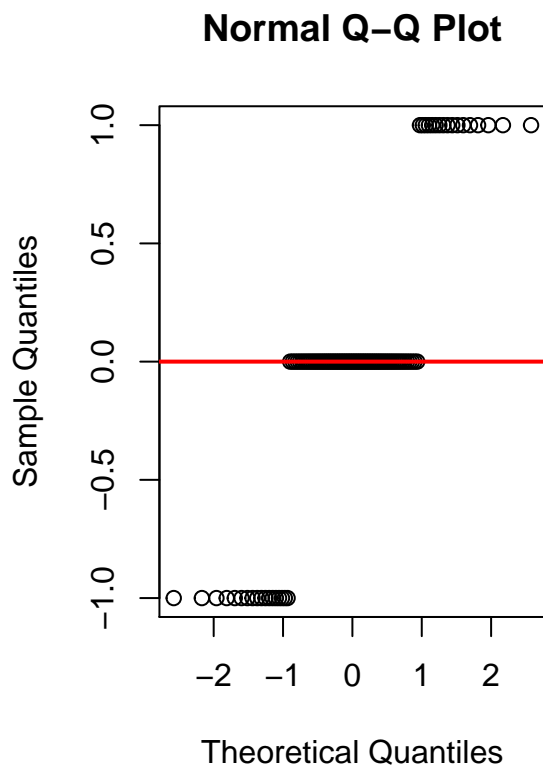
```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

#Mutant4 - MR12 ##MR12 Passed (Normality assumptions violated)##

```
diffBwSrcFlw = cancerPredictionDS$MR12 - cancerPredictionDS$Source
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot


```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
```

```
## data:  diffBwSrcFlw
## t = -0.16821, df = 99, p-value = 0.8668
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1279624  0.1079624
## sample estimates:
## mean of x
##      -0.01
```
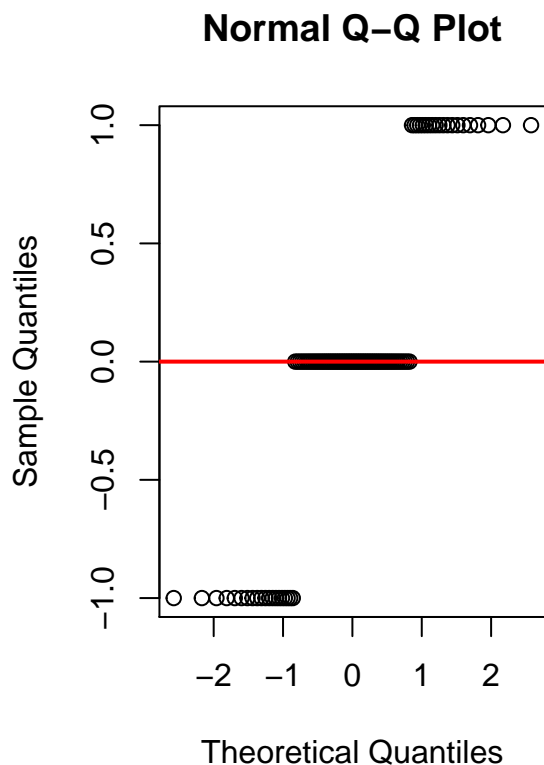
```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MR0,mu=0,conf.level=0.95,alternative="two.sided",p
```

#Mutant4 - MR21 ##MR21 Passed (Normality assumptions violated)##

```
diffBwSrcFlw = cancerPredictionDS$MR21 - cancerPredictionDS$Source
```

```
par(mfrow=c(1,2))
hist(diffBwSrcFlw,
    main="Source Executions",
    xlab= "Probabilities for class 8 ",breaks=10)
```

```
par(mfrow=c(1,2))
qqnorm(diffBwSrcFlw)
qqline(diffBwSrcFlw,col="red",lwd=2)
```

Normal Q–Q Plot

```
#BoxPlot
boxplot(diffBwSrcFlw,
        main = "Scores produced by source and followup executions",
        names = c("Difference"),
        ylab= "Scores",
        col=c(3,4))
#boxplot.stats(diffBwSrcFlw)$out
```

```
t.test(diffBwSrcFlw,mu=0,conf.level=0.95,alternative="two.sided")
```

```
##
##  One Sample t-test
##
## data:  diffBwSrcFlw
## t = 0, df = 99, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1261251  0.1261251
## sample estimates:
## mean of x
##        0
```

```
#t.test(cancerPredictionDS$Source,cancerPredictionDS$MRO,mu=0,conf.level=0.95,alternative="two.sided",p
```