

HW10

Sumeet Mishra

4/8/2019

1. Problem Set A

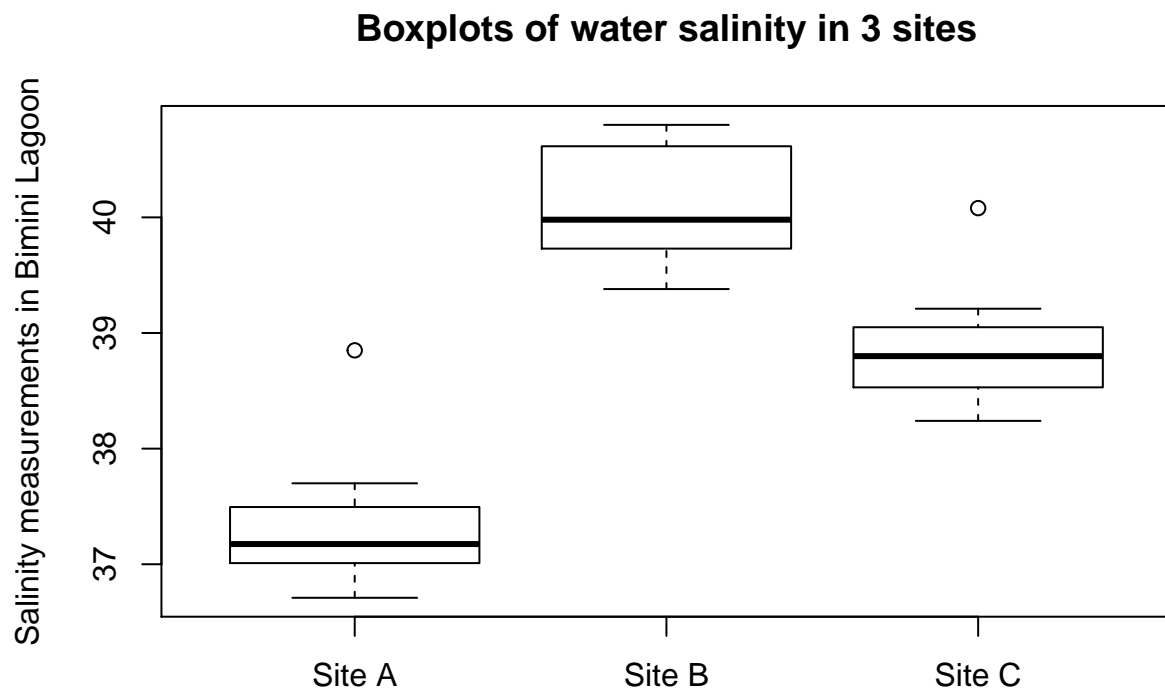
1.

```
A=c(37.54,37.01,36.71,37.03,37.32,37.01,37.03,37.70,37.36,36.75,37.45,38.85)
```

```
B=c(40.17,40.80,39.76,39.70,40.79,40.44,39.79,39.38)
```

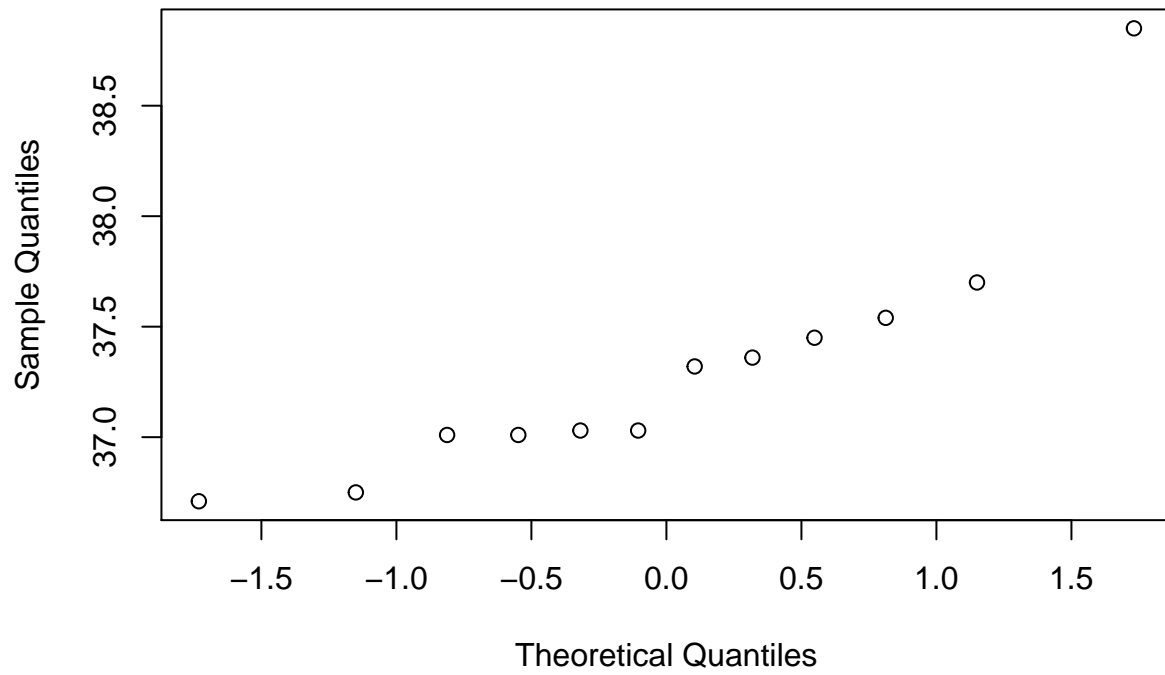
```
C=c(39.04,39.21,39.05,38.24,38.53,38.71,38.89,38.66,38.51,40.08)
```

```
boxplot(A,B,C,main = 'Boxplots of water salinity in 3 sites', ylab = 'Salinity measurements in Bimini Lagoon')
```



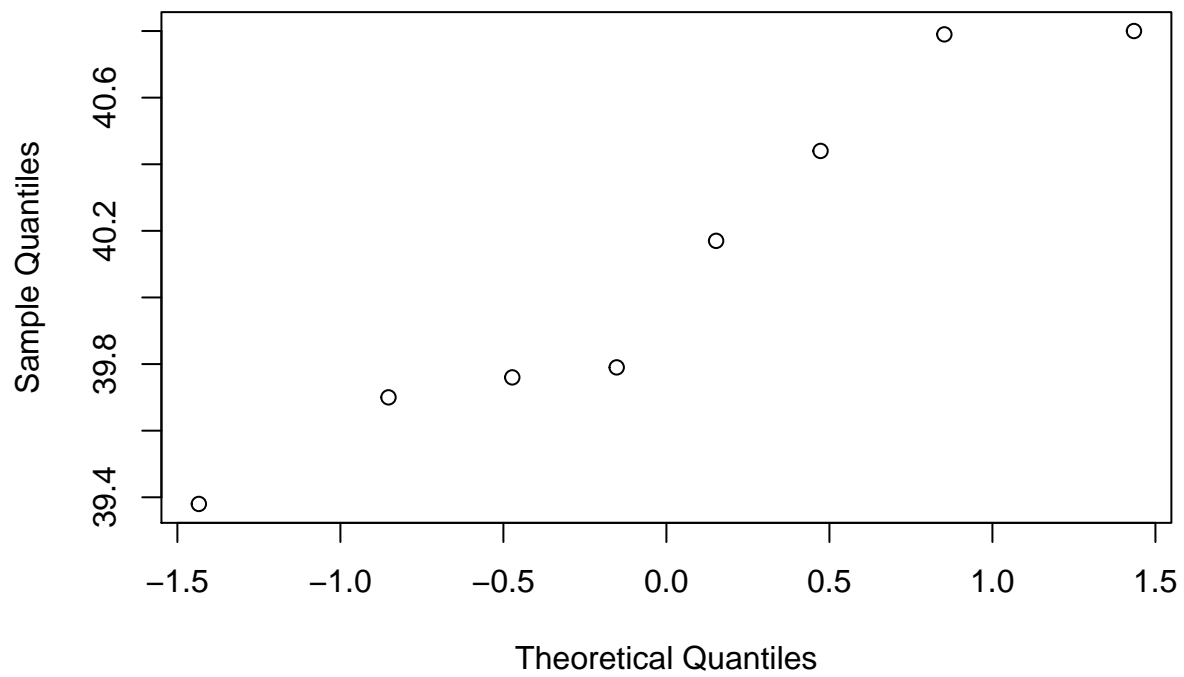
```
qqnorm(A)
```

Normal Q-Q Plot



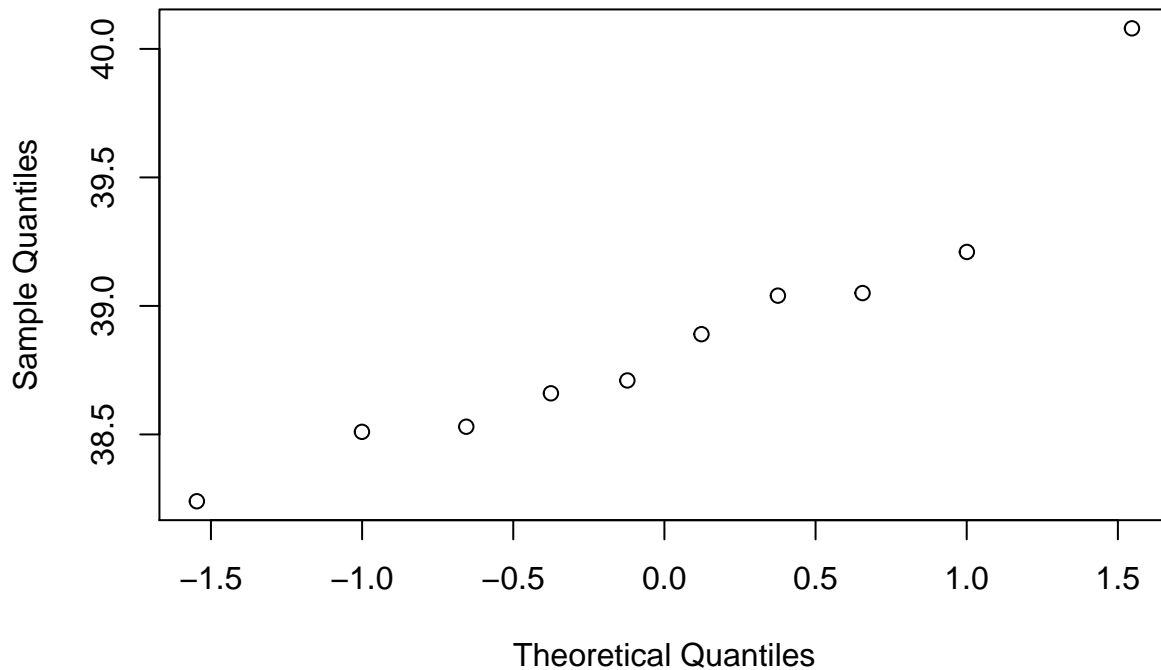
`qqnorm(B)`

Normal Q-Q Plot



`qqnorm(C)`

Normal Q-Q Plot



From the above qqnorm plots group B seems somewhat closer to normality when compared to the graphs of group A and group C, but cannot jump to the conclusion that the data of all 3 groups are normal because of the outliers in group A and C. But the data is sufficient to perform ANOVA test.

```
sd(A)
```

```
## [1] 0.5727975
```

```
sd(B)
```

```
## [1] 0.5313846
```

```
sd(C)
```

```
## [1] 0.510812
```

The sample standard deviation of A,B,C are all close(though a little different in variance). So we can conclude that homoscedasticity assumption is plausible for the ANOVA test.

2.

Null Hypothesis is that the three sites have the same salinity.

$$H_0 : \mu_A = \mu_B = \mu_C$$

Alternate Hypothesis is that atleast one of the three sites does not have the same salinity.

H_1 :Atleast 1 of the μ_i is different.

```
A.mean=mean(A)
```

```
B.mean=mean(B)
```

```
C.mean=mean(C)
```

```
S=c(A,B,C)
```

```
grand.mean=mean(S)
```

```
n1=length(A)
```

```
n2=length(B)
```

```
n3=length(C)
```

```

N=n1+n2+n3
k=3
SSB=n1*(A.mean-grand.mean)^2 + n2*(B.mean-grand.mean)^2 + n3*(C.mean- grand.mean)^2
between.df=k-1
between.meansquare=SSB/between.df
print(paste("The between degrees of freedom",between.df))

## [1] "The between degrees of freedom 2"

print(paste("The between sum of squares is",SSB))

## [1] "The between sum of squares is 38.8008825"

print(paste("The between mean square is",between.meansquare))

## [1] "The between mean square is 19.40044125"

SSW=(n1-1)*var(A) + (n2-1)*var(B) + (n3-1)*var(C)
within.df=N-k
within.meansquare=SSW/within.df
print(paste("The within degrees of freedom",within.df))

## [1] "The within degrees of freedom 27"

print(paste("The within sum of squares is",SSW))

## [1] "The within sum of squares is 7.934014166666666"

print(paste("The within mean square is",within.meansquare))

## [1] "The within mean square is 0.29385237654321"

F=between.meansquare/within.meansquare
print(paste("The F test-statistic is",F))

## [1] "The F test-statistic is 66.0210459354486"

pval=1-pf(F, df1=between.df, df2=within.df)
print(paste('The p-value is',pval))

```

```
## [1] "The p-value is 4.00864896832331e-11"
```

p value is $4.0086e-11 < \alpha = 0.05$.Hence we reject the null hypothesis. Since the p value is very very less, we do not have evidence to conclude that the three sites have the same salinity.

Anova table-

```

group=factor(c(rep("A",n1),rep("B",n2),rep("C",n3)))
anova(lm(S~group))

```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: S
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## group      2  38.801  19.4004   66.021 4.009e-11 ***
```

```
## Residuals 27   7.934   0.2939
```

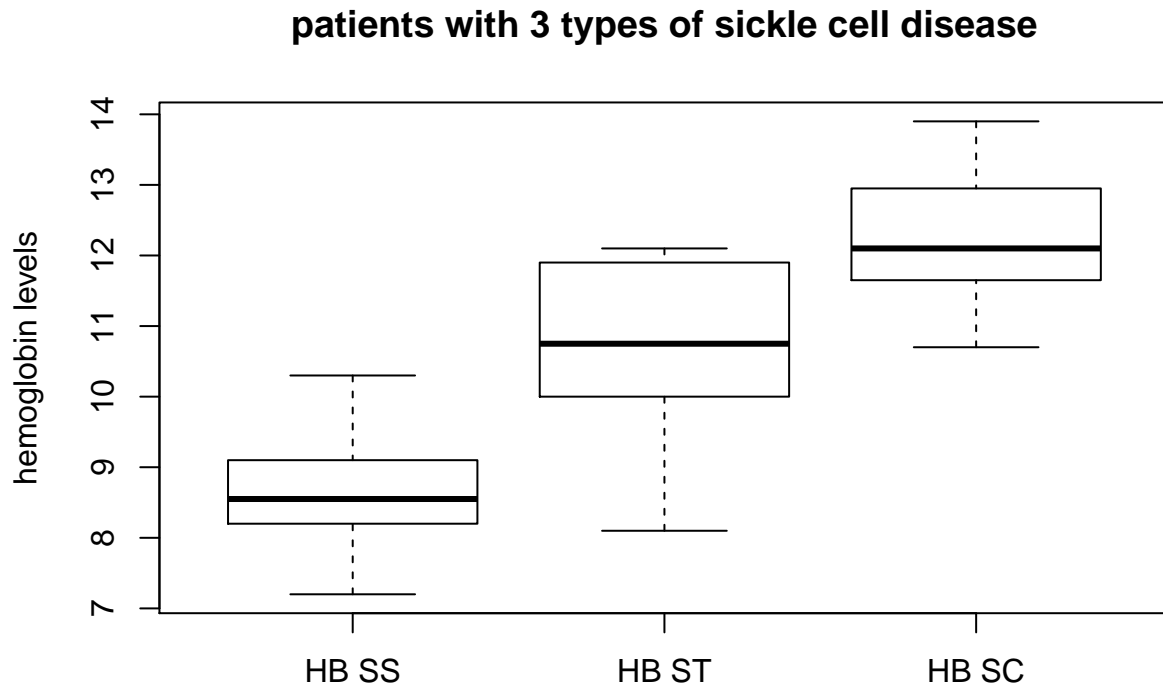
```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2.Problem Set B

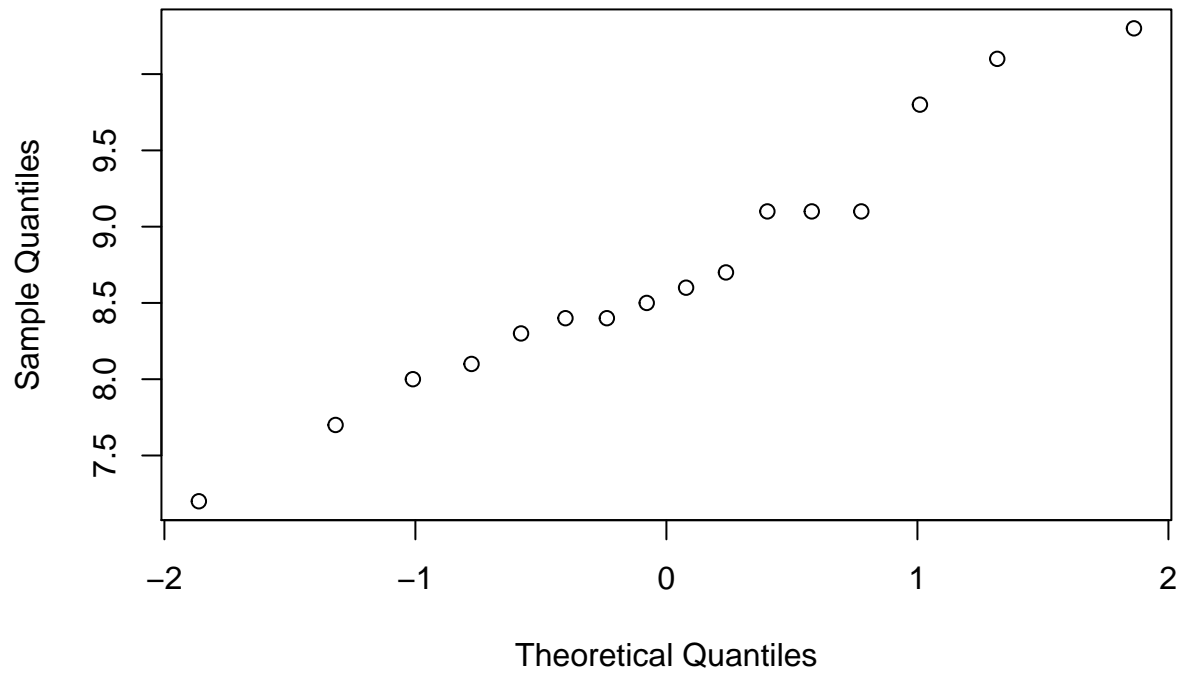
1.

```
SS=c(7.2,7.7,8.0,8.1,8.3,8.4,8.4,8.5,8.6,8.7,9.1,9.1,9.1,9.8,10.1,10.3)
ST=c(8.1,9.2,10.0,10.4,10.6,10.9,11.1,11.9,12.0,12.1)
SC=c(10.7,11.3,11.5,11.6,11.7,11.8,12.0,12.1,12.3,12.6,12.6,13.3,13.3,13.8,13.9)
boxplot(SS,ST,SC,main = 'patients with 3 types of sickle cell disease', ylab = 'hemoglobin levels', nam
```



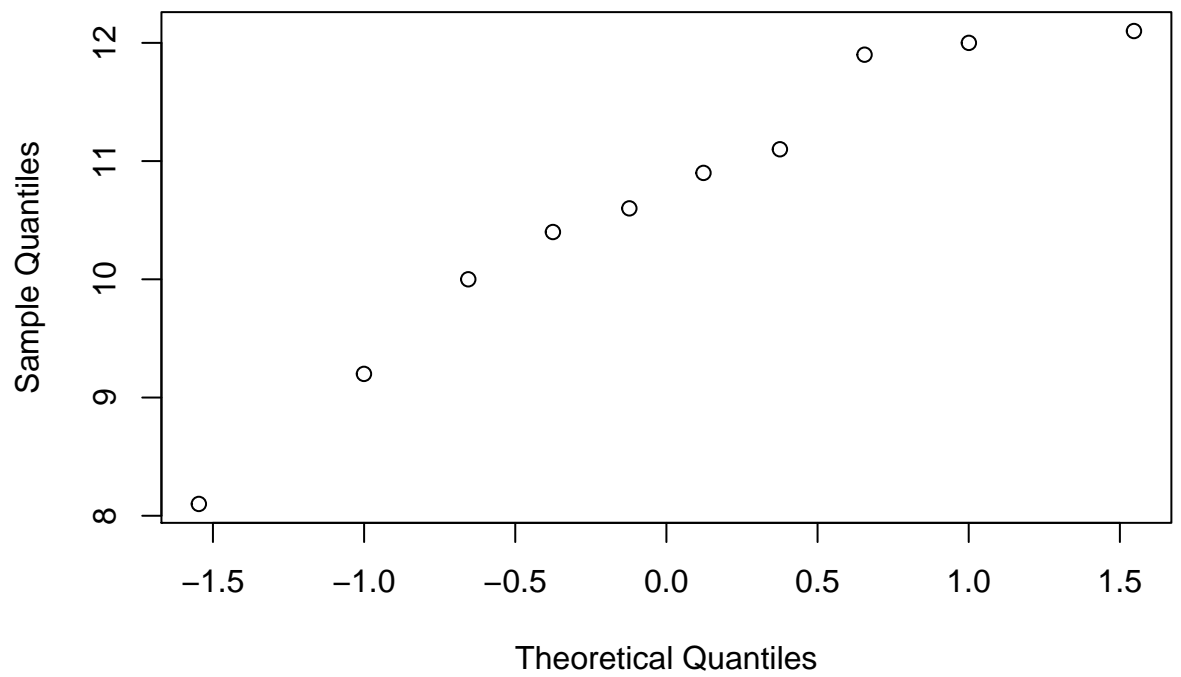
```
qqnorm(SS)
```

Normal Q-Q Plot

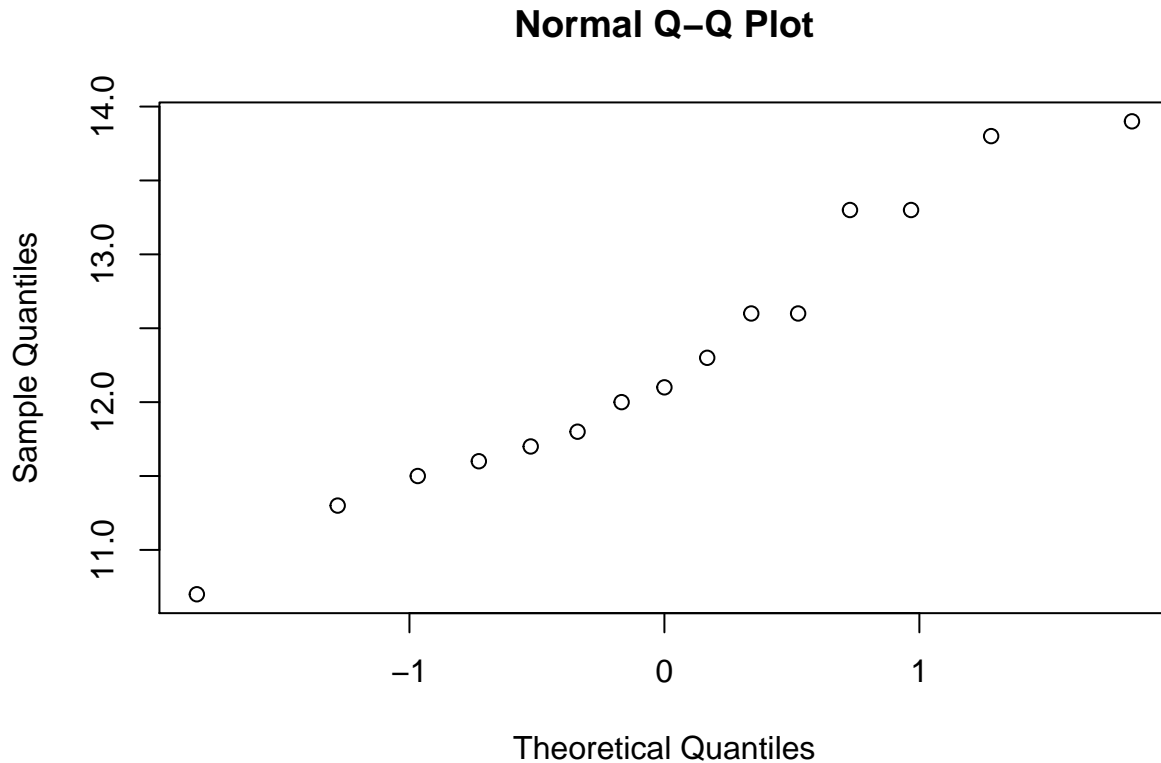


```
qqnorm(ST)
```

Normal Q-Q Plot



```
qqnorm(SC)
```



From the above qqnorm plots, dataset SS and SC seems to be normal while the graph of dataset ST is skewed to the left. We can certainly conclude that for the groups SS and SC normality assumption hold good but cannot conclude with certainty that the data of ST is normal.

```
sd(SS)
```

```
## [1] 0.844492
```

```
sd(ST)
```

```
## [1] 1.284134
```

```
sd(SC)
```

```
## [1] 0.9418826
```

The sample standard deviation of SS, ST and SC are not equal, but close enough to carry on with the ANOVA test.

2.

Null Hypothesis is three types of sickle cell disease have the same mean hemoglobin levels.

$H_0 : \mu_{SS} = \mu_{ST} = \mu_{SC}$

Alternate Hypothesis is that atleast one of the three sickle cell disease does not have the same mean hemoglobin levels.

H_1 :Atleast 1 of the μ_i is different.

```
SS.mean=mean(SS)
```

```
ST.mean=mean(ST)
```

```
SC.mean=mean(SC)
```

```
S=c(SS,ST,SC)
```

```
grand.mean=mean(S)
```

```
n1=length(SS)
```

```
n2=length(ST)
```

```

n3=length(SC)
N=n1+n2+n3
k=3
SSB=n1*(SS.mean-grand.mean)^2 + n2*(ST.mean-grand.mean)^2 + n3*(SC.mean- grand.mean)^2
between.df=k-1
between.meansquare=SSB/between.df
print(paste("The between degrees of freedom",between.df))

```

```
## [1] "The between degrees of freedom 2"
```

```
print(paste("The between sum of squares is",SSB))
```

```
## [1] "The between sum of squares is 99.8893048780488"
```

```
print(paste("The between mean square is",between.meansquare))
```

```
## [1] "The between mean square is 49.9446524390244"
```

```
SS.within=(n1-1)*var(SS) + (n2-1)*var(ST) + (n3-1)*var(SC)
```

```
within.df=N-k
```

```
meansquare.within=SS.within/within.df
```

```
print(paste("The within of freedom is",within.df))
```

```
## [1] "The within of freedom is 38"
```

```
print(paste("The within sum of squares is",SS.within))
```

```
## [1] "The within sum of squares is 37.9585"
```

```
print(paste("The within mean squares is",meansquare.within))
```

```
## [1] "The within mean squares is 0.998907894736842"
```

```
F.ts=between.meansquare/meansquare.within
```

```
print(paste("The F test-statistic is",F.ts))
```

```
## [1] "The F test-statistic is 49.9992568906286"
```

```
pvalue=1-pf(F.ts, df1=between.df, df2=within.df)
```

```
print(paste('The p-value is',pvalue))
```

```
## [1] "The p-value is 2.28178587136085e-11"
```

p value is $2.281e-11 < \alpha(0.05)$. Hence we reject the null hypothesis. Since the p value is very very less, we do not have evidence to conclude that the three types of sickle cell disease have the same mean hemoglobin levels.

Anova table-

```

group=factor(c(rep("SS",n1),rep("ST",n2),rep("SC",n3)))
anova(lm(S~group))

```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: S
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## group      2  99.889   49.945  49.999 2.282e-11 ***
```

```
## Residuals 38  37.959    0.999
```

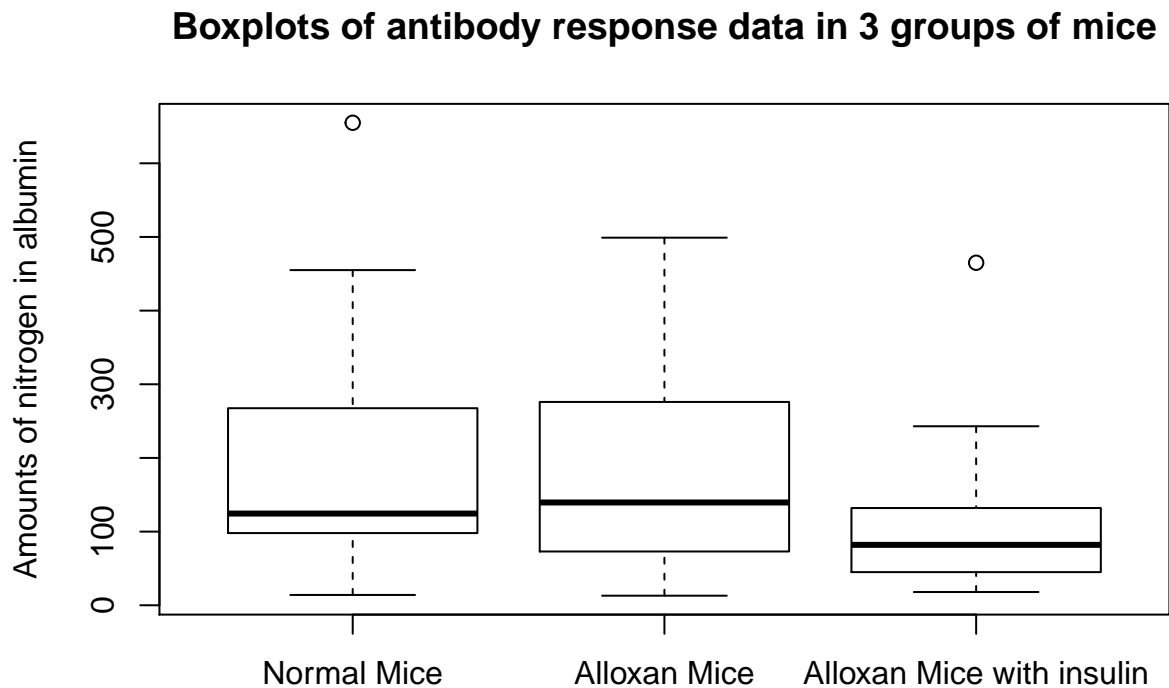
```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


3.Problem Set G

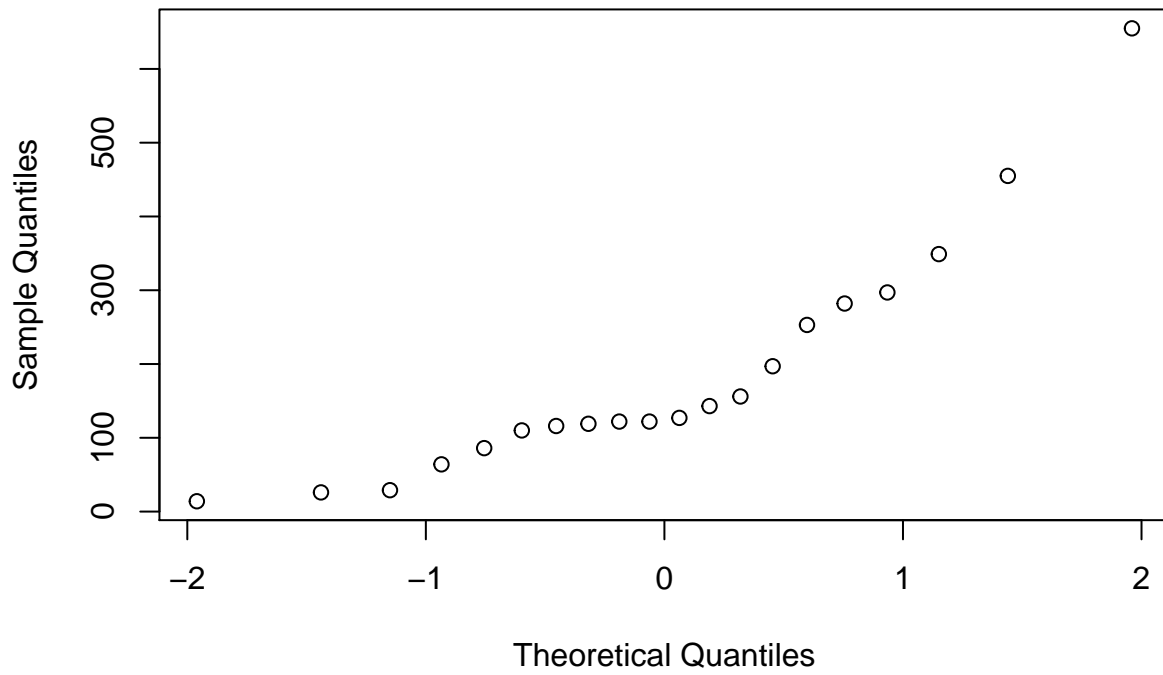
1.

```
normal_mice=c(156,282,197,297,116,127,119,29,253,122,349,110,143,64,26,86,122,455, 655,14)
n_nm=length(normal_mice)
alloxan_mice=c(391,46,469,86,174,133,13,499,168,62,127,276,176,146,108,276,50,73)
n_am=length(alloxan_mice)
alloxan_insulin_mice=c(82,100,98,150,243,68,228,131,73,18,20,100,72,133,465,40,46, 34,44)
n_aim=length(alloxan_insulin_mice)
N_total=n_nm+n_am+n_aim
boxplot(normal_mice, alloxan_mice, alloxan_insulin_mice,
main = 'Boxplots of antibody response data in 3 groups of mice', ylab = 'Amounts of nitrogen in albumin
```



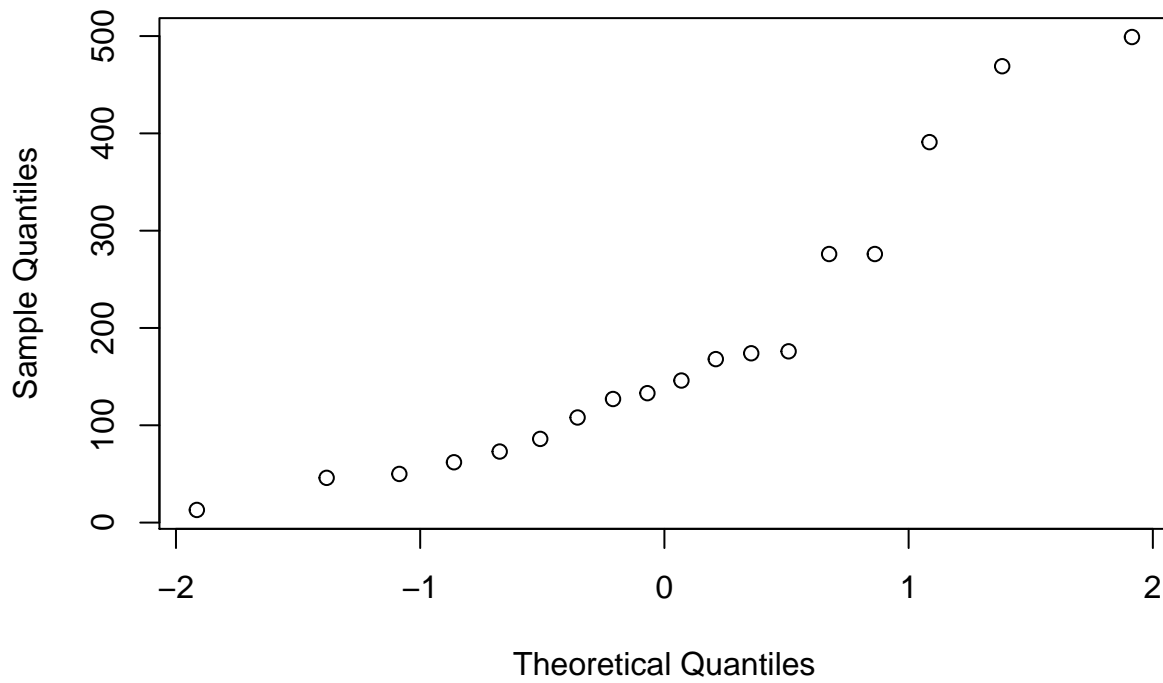
```
qqnorm(normal_mice)
```

Normal Q-Q Plot



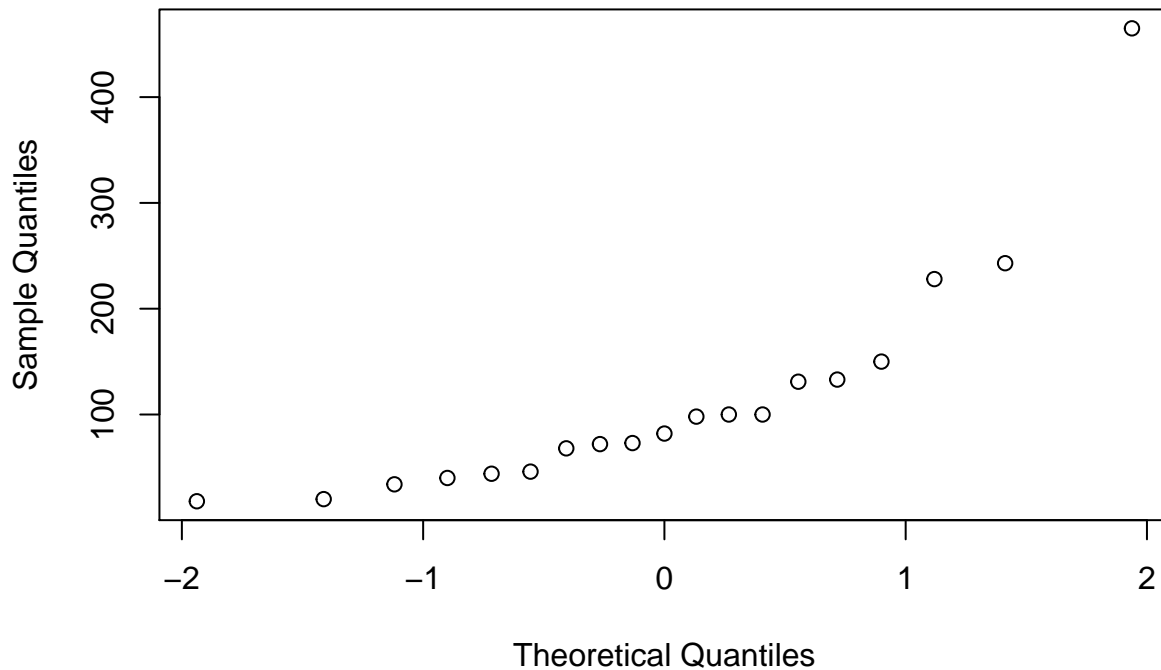
```
qqnorm(alloxan_mice)
```

Normal Q-Q Plot



```
qqnorm(alloxan_insulin_mice)
```

Normal Q-Q Plot



From the above qqplots, the graphs are not linear. Hence, we can conclude that the normality assumptions do not hold true.

```
sd(normal_mice)
```

```
## [1] 158.8349
```

```
sd(alloxan_mice)
```

```
## [1] 144.8493
```

```
sd(alloxan_insulin_mice)
```

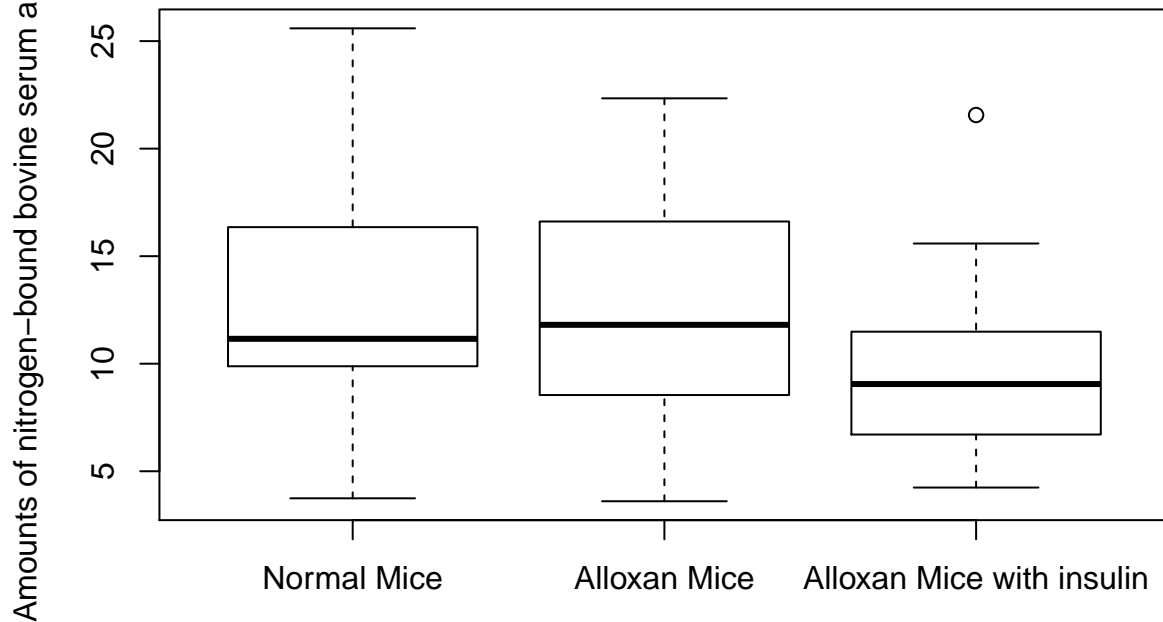
```
## [1] 105.7896
```

Although, the standard deviation of the 3 groups are not equal, the difference between the 3 SD is not huge enough (not twice the other), hence we can assume that the homoscedasticity is plausible.

2.

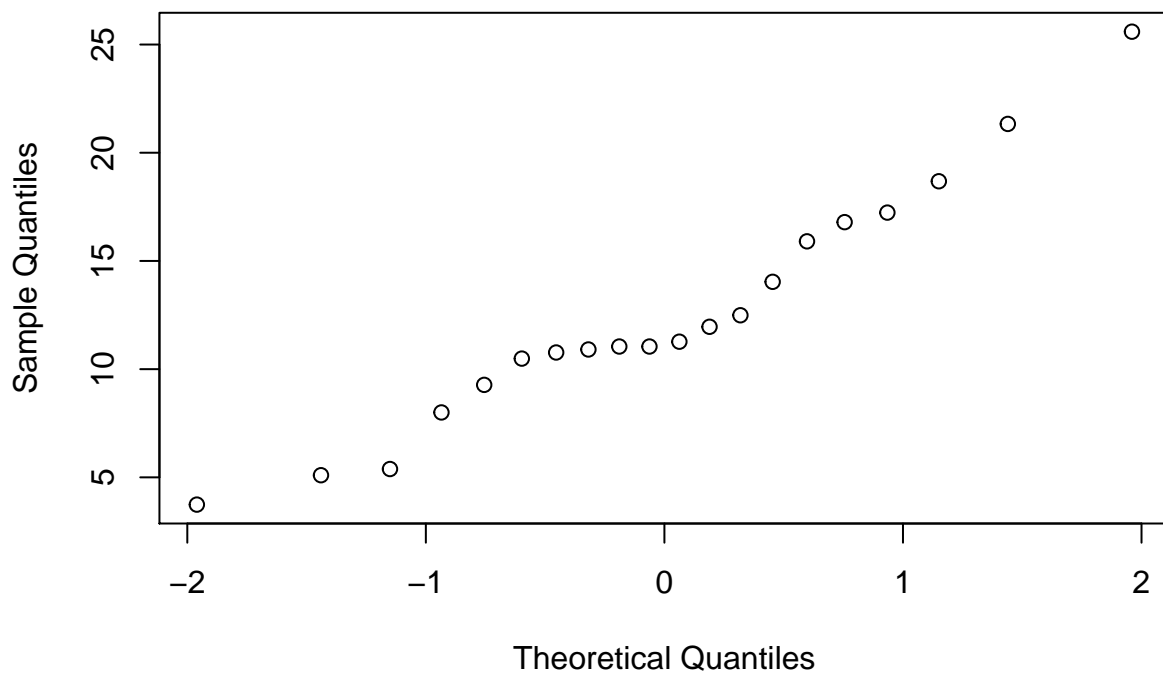
```
trans.normal_mice=sqrt(normal_mice)
trans.alloxan_mice=sqrt(alloxan_mice)
trans.alloxan_insulin_mice=sqrt(alloxan_insulin_mice)
trans.all=c(trans.normal_mice,trans.alloxan_mice,trans.alloxan_insulin_mice)
boxplot(trans.normal_mice, trans.alloxan_mice, trans.alloxan_insulin_mice,
main = 'Boxplots of transformed antibody response data in 3 groups of mice',
ylab = 'Amounts of nitrogen-bound bovine serum albumin', names=c('Normal Mice','Alloxan Mice','Alloxan Insulin Mice'))
```

Boxplots of transformed antibody response data in 3 groups of mic



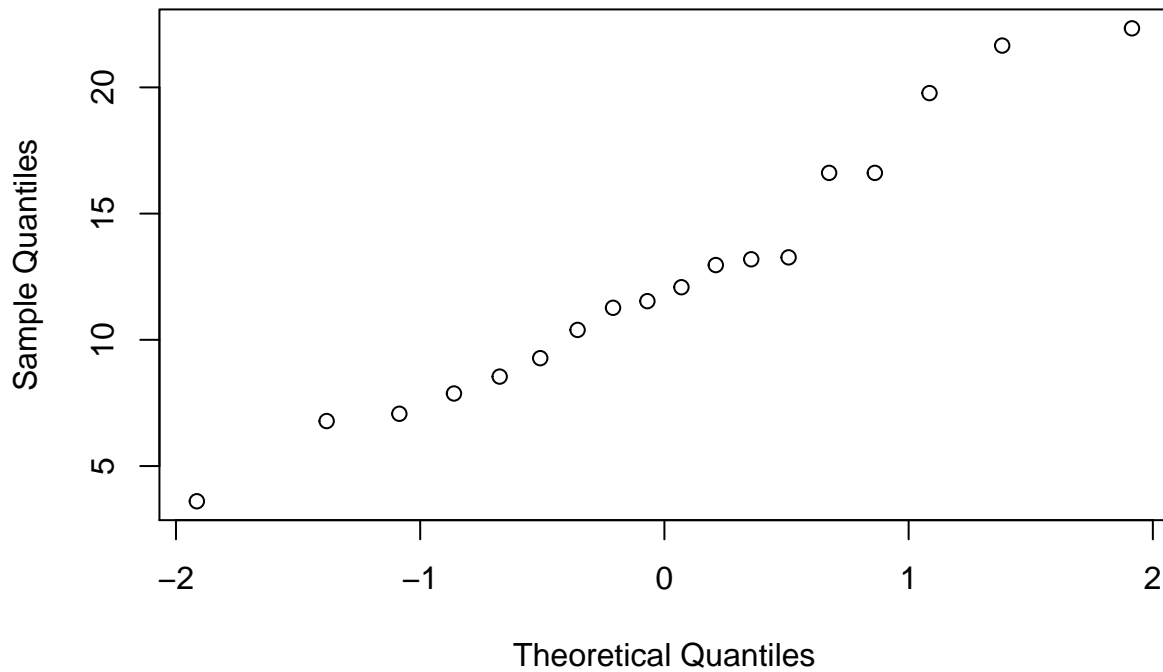
```
qqnorm(trans.normal_mice)
```

Normal Q-Q Plot



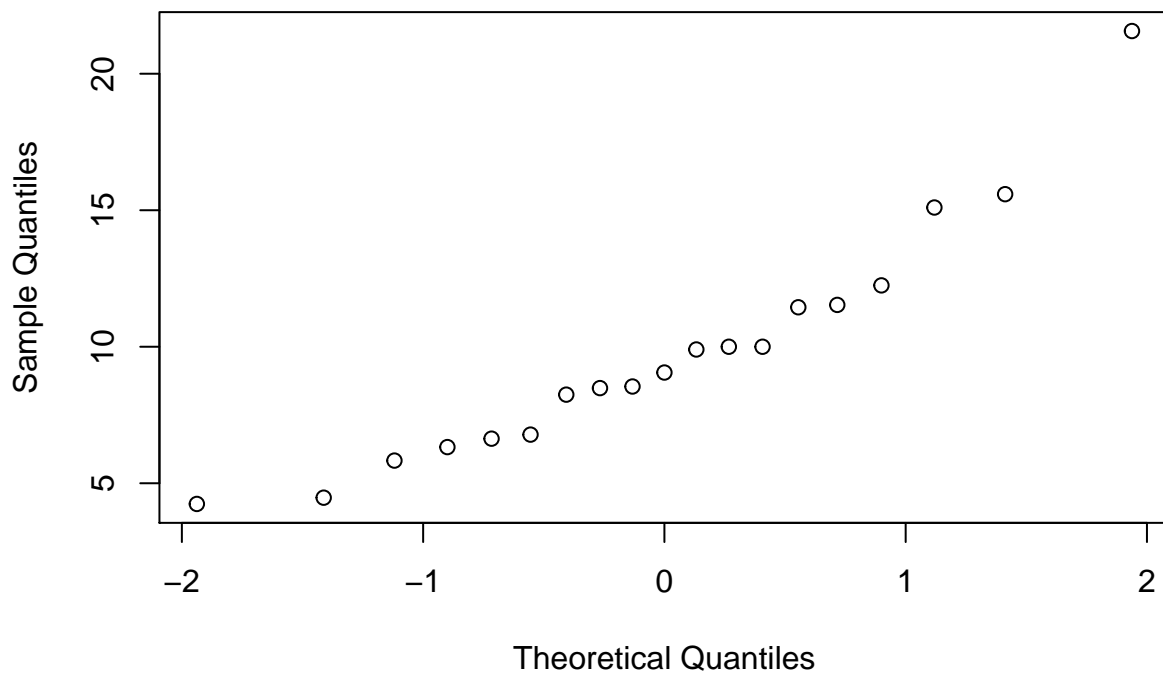
```
qqnorm(trans.alloxan_mice)
```

Normal Q-Q Plot



```
qqnorm(trans.alloxan_insulin_mice)
```

Normal Q-Q Plot



From the normal and density plots, we see that all the 3 groups are close to normality, while alloxan mice with insulin group is skewed to the right. We can conclude that normality assumption holds plausible.

```
sd(trans.normal_mice)
```

```
## [1] 5.480753
```

```
sd(trans.alloxan_mice)
```

```
## [1] 5.226939
```

```
sd(trans.alloxan_insulin_mice)
```

```
## [1] 4.244474
```

Although, the standard deviation of the 3 groups are not exactly equal, they are close enough to proceed with the ANOVA test with the transformed data. Even though there is a variation in the spread, this variation is not as much as in the original data, hence we can safely assume that homoscedasticity is plausible. ANOVA assumptions is plausible with the transformed data.

3. Null Hypothesis is three groups of mice have the same mean antibody response.

$$H_0 : \mu_{NormalMice} = \mu_{AlloxanMice} = \mu_{AlloxanInsulineMice}$$

Alternate Hypothesis is atleast one of the three groups of mice does not have the same mean antibody response.

H_1 :Atleast 1 of the μ_i is different.

```
trans.normal_mice.mean=mean(trans.normal_mice)
```

```
trans.alloxan_mice.mean=mean(trans.alloxan_mice)
```

```
trans.alloxan_insulin_mice.mean=mean(trans.alloxan_insulin_mice)
```

```
trans.all.mean=mean(trans.all)
```

```
SoSB=n_nm*(trans.normal_mice.mean-trans.all.mean)^2 + n_am*(trans.alloxan_mice.mean-trans.all.mean)^2+n_aim*(trans.alloxan_insulin_mice.mean-trans.all.mean)^2
```

```
dofb=k-1
```

```
msbetween=SoSB/dofb
```

```
print(paste("The between degrees of freedom is",dofb))
```

```
## [1] "The between degrees of freedom is 2"
```

```
print(paste("The between sum of squares is",SoSB))
```

```
## [1] "The between sum of squares is 94.735137611465"
```

```
print(paste("The between mean squares is",msbetween))
```

```
## [1] "The between mean squares is 47.3675688057325"
```

```
SoSW=(n_nm-1)*var(trans.normal_mice) + (n_am-1)*var(trans.alloxan_mice) + (n_aim-1)*var(trans.alloxan_insulin_mice)
```

```
dofw=N_total-k
```

```
mswithin=SoSW/dofw
```

```
print(paste("The within of freedom is",dofw))
```

```
## [1] "The within of freedom is 54"
```

```
print(paste("The within sum of squares is",SoSW))
```

```
## [1] "The within sum of squares is 1359.46974752583"
```

```
print(paste("The within mean squares is",mswithin))
```

```
## [1] "The within mean squares is 25.1753656949227"
```

```
Fstat=msbetween/mswithin
```

```
print(paste("The F test-statistic is",Fstat))
```

```
## [1] "The F test-statistic is 1.88150469708114"
```

```
p_value=1-pf(Fstat, df1=dofb, df2=dofw)
print(paste('The p-value is',p_value))
```

```
## [1] "The p-value is 0.162213354721056"
```

p value is 0.1622 > $\alpha(0.05)$. Hence we fail to reject the null hypothesis. We can conclude that the three groups of mice have the same mean antibody response.

The anova table -

```
class=factor(c(rep("normal_mice",n_nm),rep("alloxan_mice",n_am),rep("alloxan_insulin_mice",n_aim)))
anova(lm(trans.all~class))
```

```
## Analysis of Variance Table
##
## Response: trans.all
##          Df Sum Sq Mean Sq F value Pr(>F)
## class      2   94.74   47.368   1.8815 0.1622
## Residuals 54 1359.47   25.175
```

4.

significance level $=\alpha=0.05$ Since we have 3 groups, the results of Bonferroni t-test is compared with the new significance level, $\alpha/3$

```
alpha=0.05/3
alpha
```

```
## [1] 0.01666667
```

- a) Null hypothesis is the antibody response of alloxan diabetic mice does not differ from the antibody response of normal mice. Alternate hypothesis is the antibody response of alloxan diabetic mice differ from the antibody response of normal mice.

$$H_0 : \mu_{NormalMice} - \mu_{AlloxanMice} = 0$$

$$H_0 : \mu_{NormalMice} - \mu_{AlloxanMice} \neq 0$$

```
t.test(trans.alloxan_mice, trans.normal_mice)
```

```
##
## Welch Two Sample t-test
##
## data: trans.alloxan_mice and trans.normal_mice
## t = -0.035226, df = 35.867, p-value = 0.9721
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.585979 3.463552
## sample estimates:
## mean of x mean of y
## 12.49121 12.55242
```

The p value = 0.9721 is greater than $\alpha(0.0167)$. Hence we fail to reject null hypothesis and conclude that the true means of the transformed normal and transformed alloxan mice are same.

- b) Null hypothesis is the antibody response of alloxan diabetic mice with insulin does not differ from the the antibody response of alloxan diabetic mice. Alternate hypothesis is the antibody response of alloxan diabetic mice with insulin differ from the the antibody response of alloxan diabetic mice.

$$H_0 : \mu_{AlloxanMice} - \mu_{AlloxanInsulinMice} = 0$$

$$H_0 : \mu_{AlloxanMice} - \mu_{AlloxanInsulinMice} \neq 0$$

```
t.test(trans.alloxan_mice, trans.alloxan_insulin_mice)
```

```
##
## Welch Two Sample t-test
##
## data: trans.alloxan_mice and trans.alloxan_insulin_mice
## t = 1.7207, df = 32.789, p-value = 0.09474
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.4936299 5.8977548
## sample estimates:
## mean of x mean of y
## 12.491208 9.789145
```

The p value = 0.09474 is greater than $\alpha(0.0167)$. Hence we fail to reject null hypothesis and conclude that the true means of the transformed alloxan and transformed alloxan with insulin mice are same.

c)

Null hypothesis is the antibody response of alloxan diabetic mice treated with insulin does not differ from the antibody response of normal mice. Alternate hypothesis is the antibody response of alloxan diabetic mice treated with insulin differ from the antibody response of normal mice.

$$H_0 : \mu_{NormalMice} - \mu_{AlloxanInsulinMice} = 0$$

$$H_0 : \mu_{NormalMice} - \mu_{AlloxanInsulinMice} \neq 0$$

```
t.test(trans.normal_mice, trans.alloxan_insulin_mice)
```

```
##
## Welch Two Sample t-test
##
## data: trans.normal_mice and trans.alloxan_insulin_mice
## t = 1.7653, df = 35.59, p-value = 0.08608
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.4125421 5.9390938
## sample estimates:
## mean of x mean of y
## 12.552421 9.789145
```

The p value = 0.08608 is greater than $\alpha(0.0167)$. Hence we fail to reject null hypothesis and conclude that the true means of the transformed normal and transformed alloxan with insulin mice are same. None of the Null hypothesis is rejected.