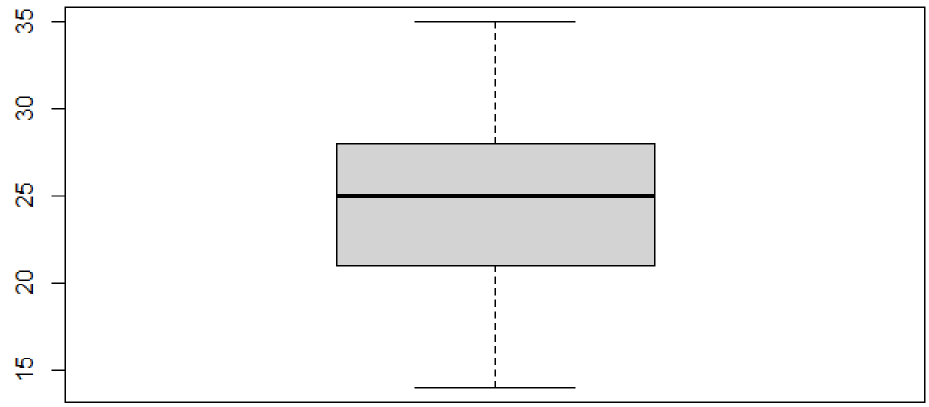
2)

3.3a

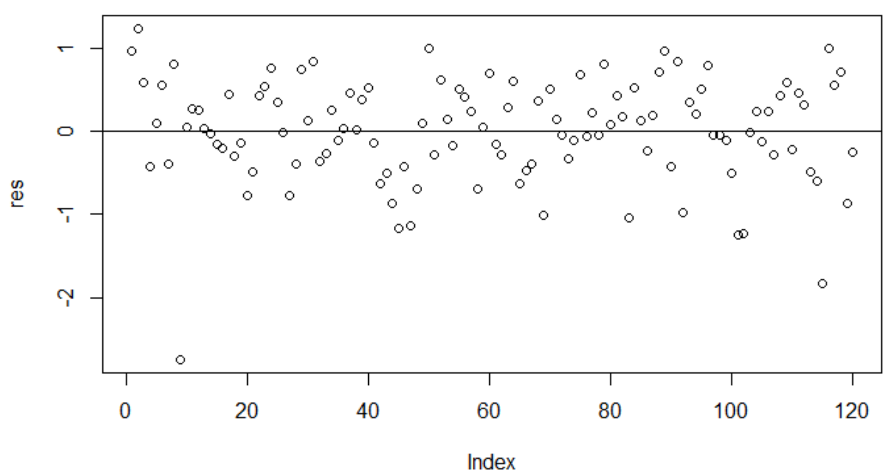
```
f=file.choose()
TestScores=read.table(f)
colnames(TestScores)= c("GPA","ACT")
boxplot(TestScores$ACT)
```



There aren’t any huge outliers.

3.3b

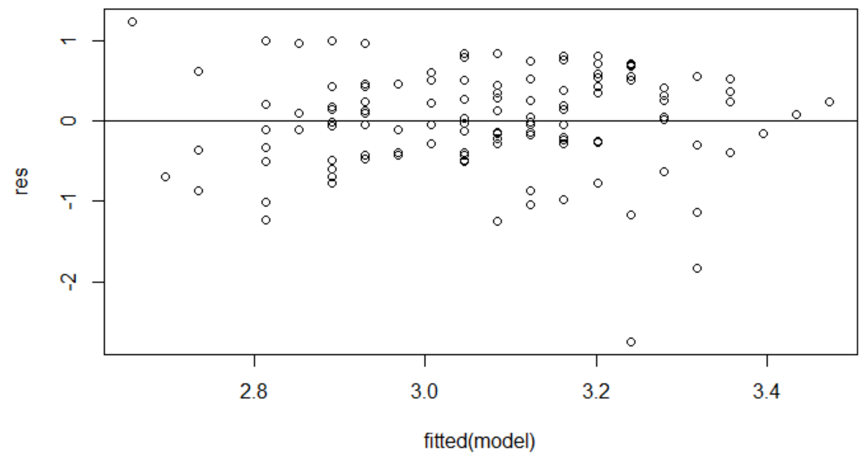
```
model=lm(GPA~ACT, data = TestScores)
res=resid(model)
plot(res)
abline(0,0)
```



Through the residual model, it is clear a linear model is appropriate.

3.3c

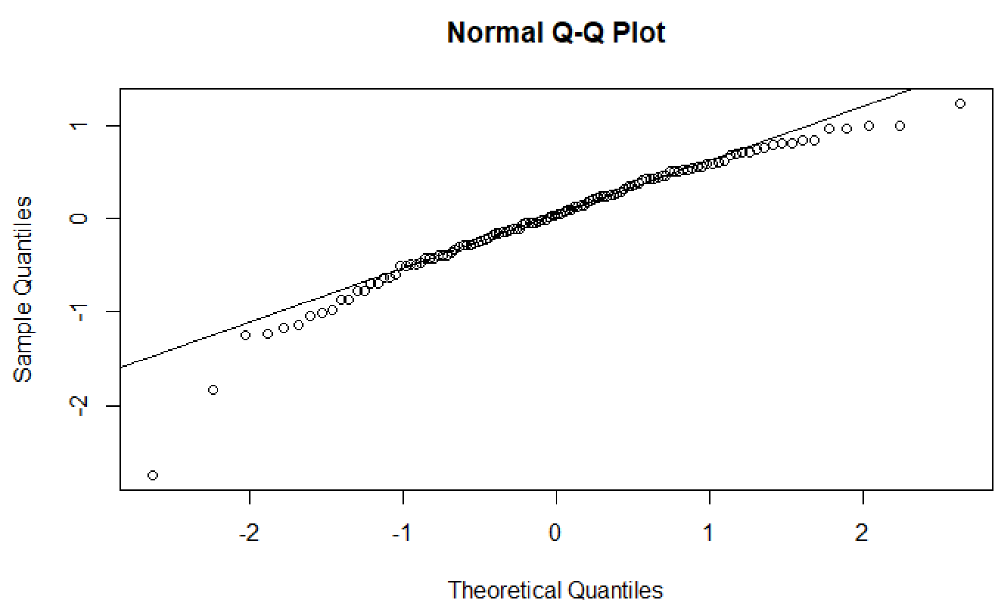
```
plot(fitted(model),res)
abline(0,0)
```



Few outliers, low variance, linear

3.3d

```
qqnorm(res)
qqline(res)
```

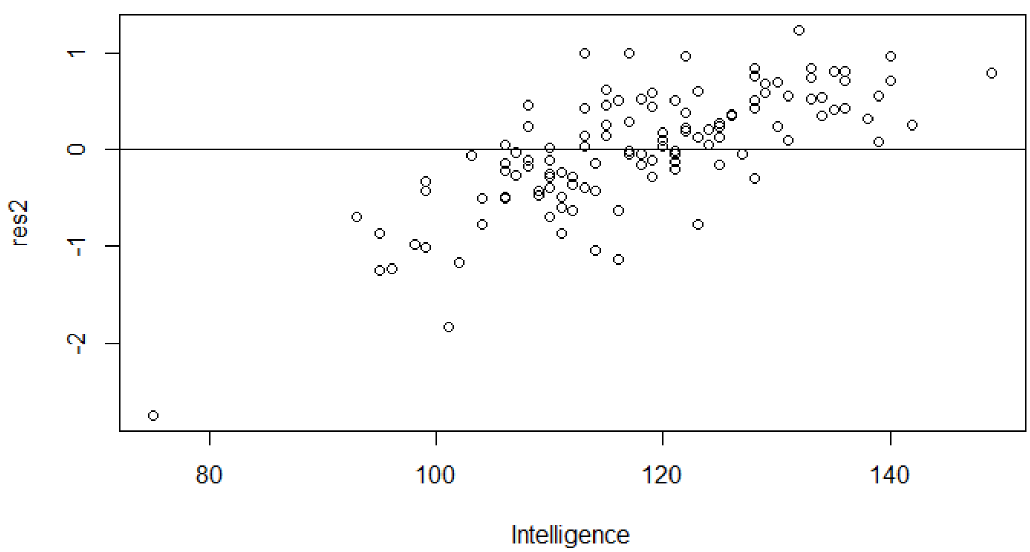



```
r=cor(GPA,ACT)
r
[1] 0.2694818
```

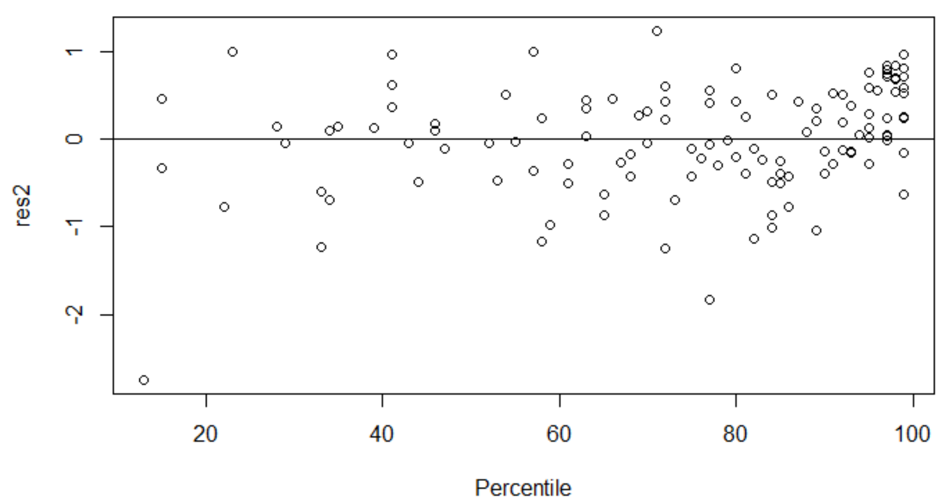
H0: Normal, Ha: not normal. Correlation critical value is .987. Our correlation coefficient is less than the critical so reject the null.

3.3f

```
f=file.choose()
Scores=read.table(f)
mod=lm(GPA~ACT, data = Scores)
res2=resid(mod)
plot(res2)
abline(0,0)
attach(Scores)
plot(Intelligence,res2)
abline(0,0)
```




```
plot(Percentile,res2)
abline(0,0)
```



Based off the graphs, it seems as if the model could be improved with an inclusion of the intelligence test of a student.

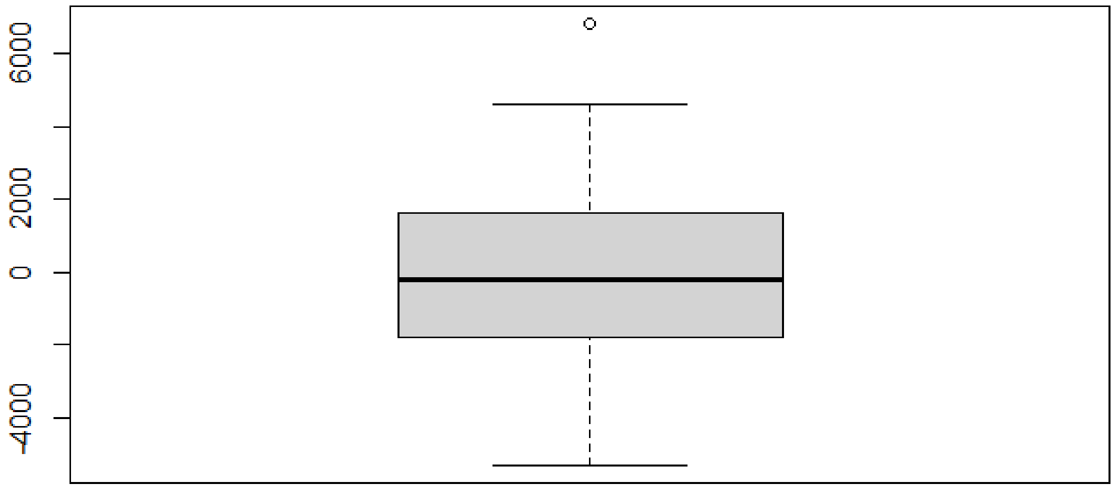
3.8a

```
f=file.choose()
CrimeRate=read.table(f)
colnames(CrimeRate)=c("Crime Rate","Diploma%")
stem(CrimeRate$`Diploma%`)
  The decimal point is 1 digit(s) to the right of the |

  6 | 1444
  6 | 5678
  7 | 00334444
  7 | 55556666777777788888888999999
  8 | 000011111112222222233333344444
  8 | 55578889
  9 | 11
```

Many in the 70%

3.8b

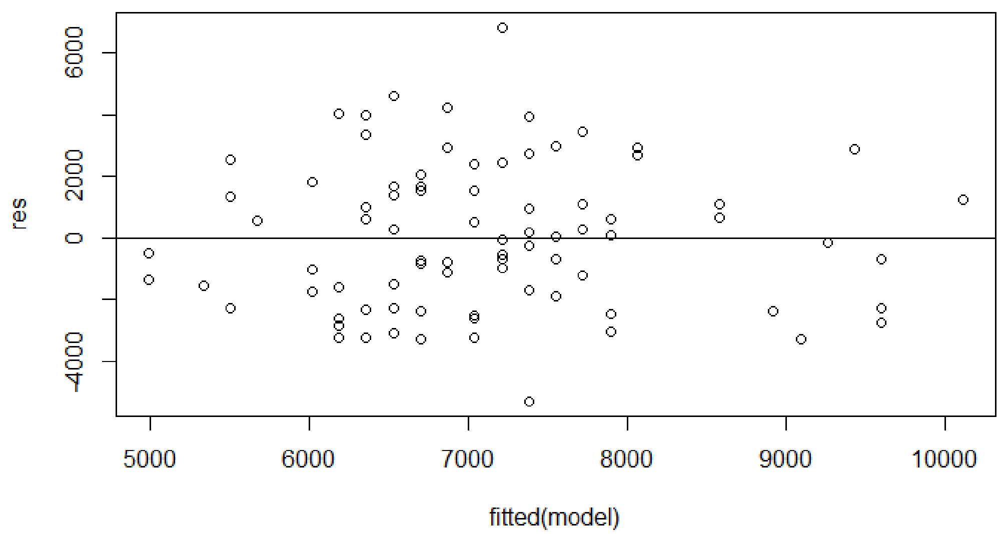
```
attach(CrimeRate)
model=lm(`Crime Rate`~ `Diploma%`)
res=resid(model)
boxplot(res)
```



No.

3.8c

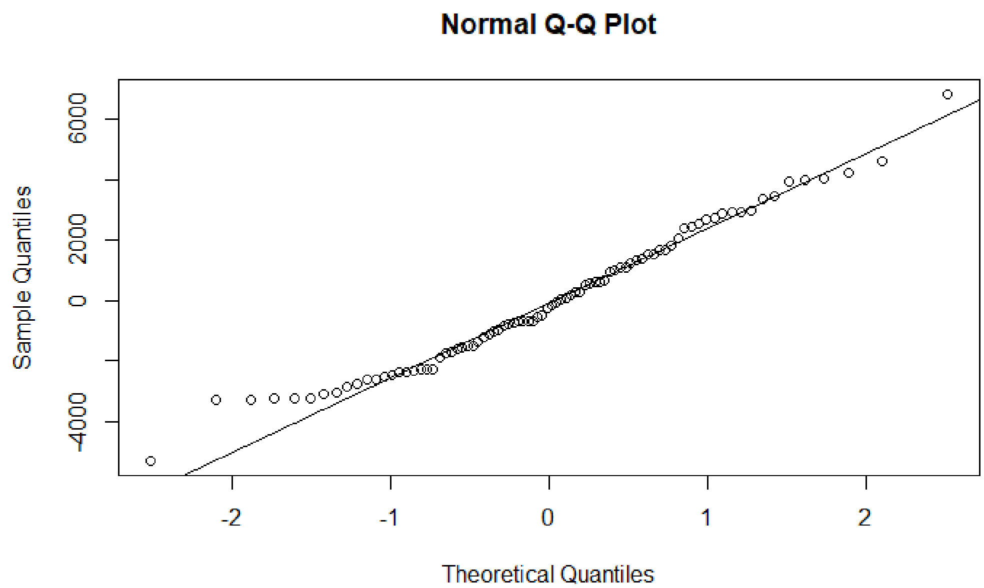
```
plot(fitted(model),res)
abline(0,0)
```



Through this plot,it is clear a linear model is appropriate.

3.8d

```
qqnorm(res)
qqline(res)
```



H0: Normal Ha: not normal. r = 0.98876. If r ≥ r* then fail to reject the null. r* = 0.9854 so fail to reject the null.

3.15a

f=file.choose()
SolutionConc=read.table(f)
colnames(SolutionConc)=c("Concentration","Hours")
lm(Concentration~Hours, data = SolutionConc)
Call:
lm(formula = Concentration ~ Hours, data = SolutionConc)

Coefficients:
(Intercept)   Hours
      2.575   -0.324

Y=2.575-.324x

3.15b

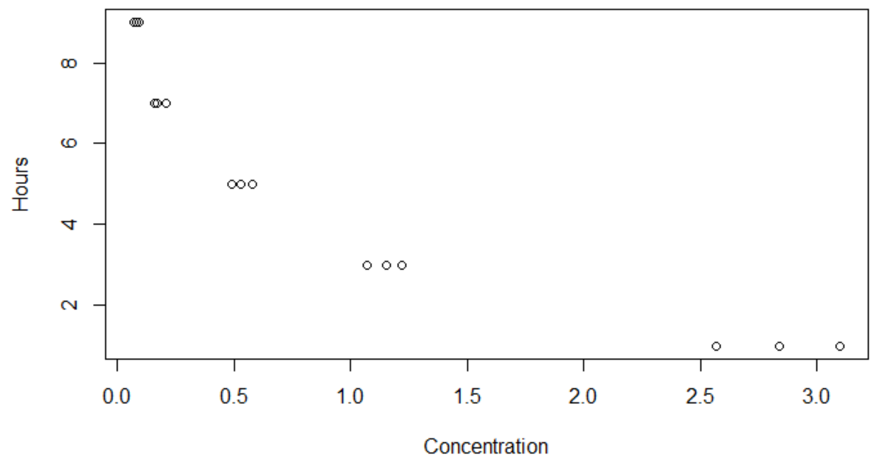
H0: E{Y } = β0 + β1X, Ha: E{Y } 6= β0 + β1X.
F-statistic: 55.99 on 1 and 13 DF, p-value: 4.611e-06.P value is very low so reject the null.

3.15c

Since there is a lack of fit, a linear model is not appropriate for this data

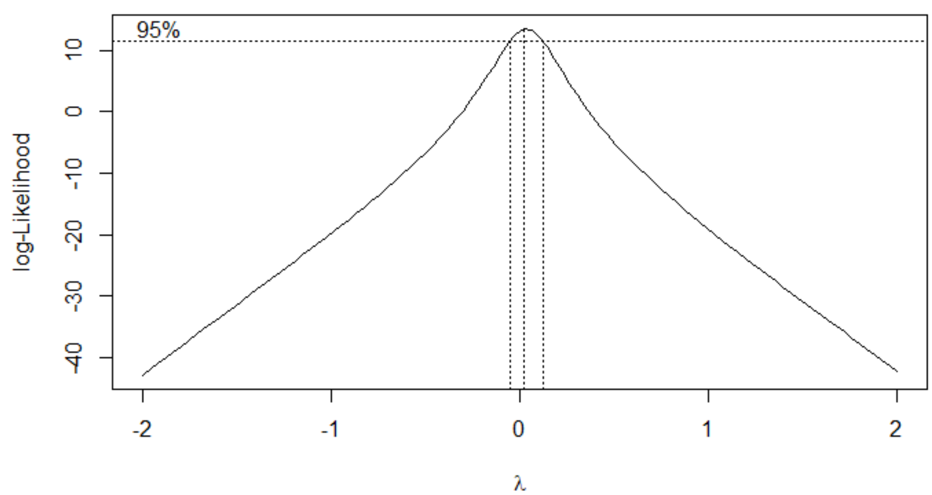
3.16a

plot(SolutionConc)



The Y'=log10(y) transformation most fits the pattern of the scatter plot

3.16b
bc=boxcox(Concentration~Hours)



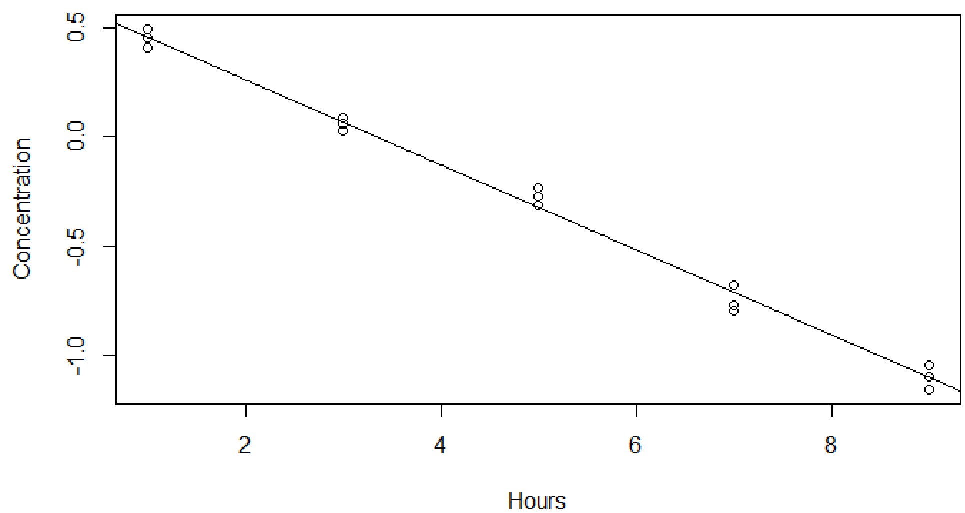
lamdba=bc$x[which.max(bc$y)]
lamdba
[1] 0.02020202

3.16c
Concentration=log10(Concentration)
newmod=lm(Concentration~Hours)
newmod
Call:

lm(formula = Concentration ~ Hours)

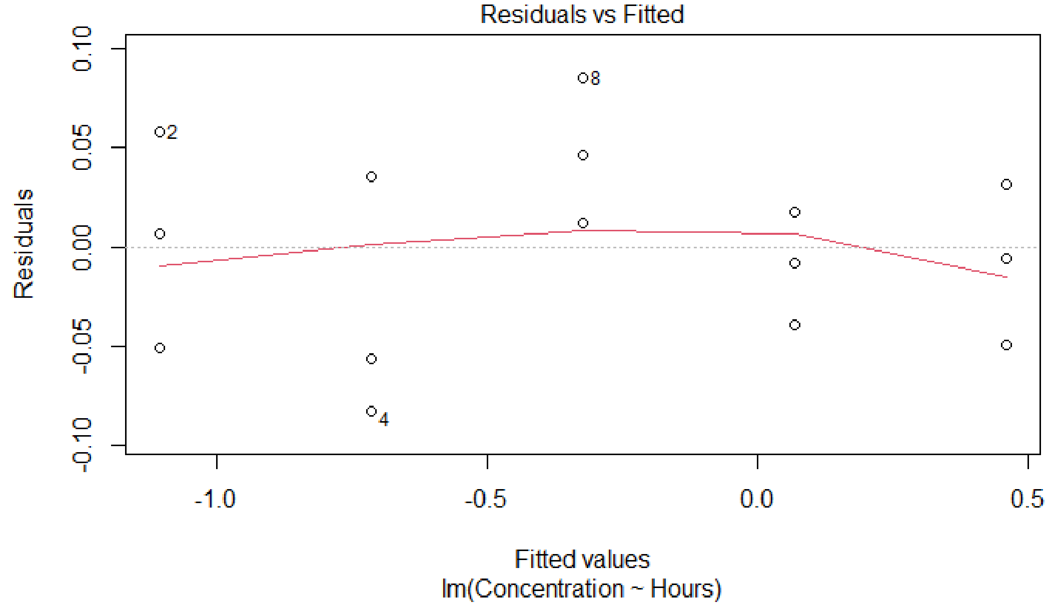
Coefficients:
(Intercept)   Hours
     0.6549   -0.1954
Y'=.6549-.1954x

3.16d
plot(Hours,Concentration)
abline(newmod)

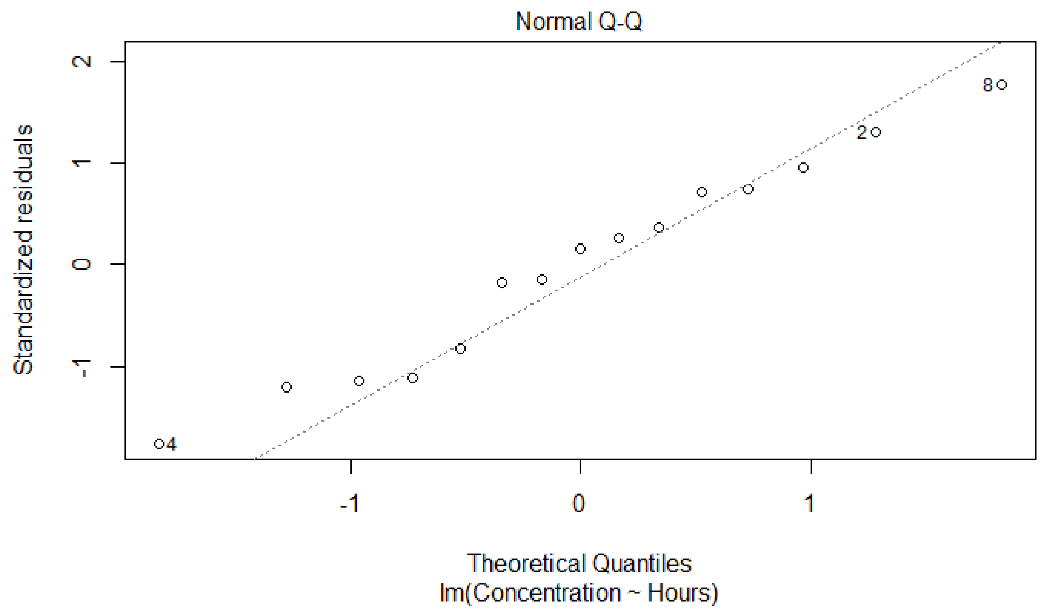


Yes

3.16e
plot(newmod)



Values are spread out with no pattern



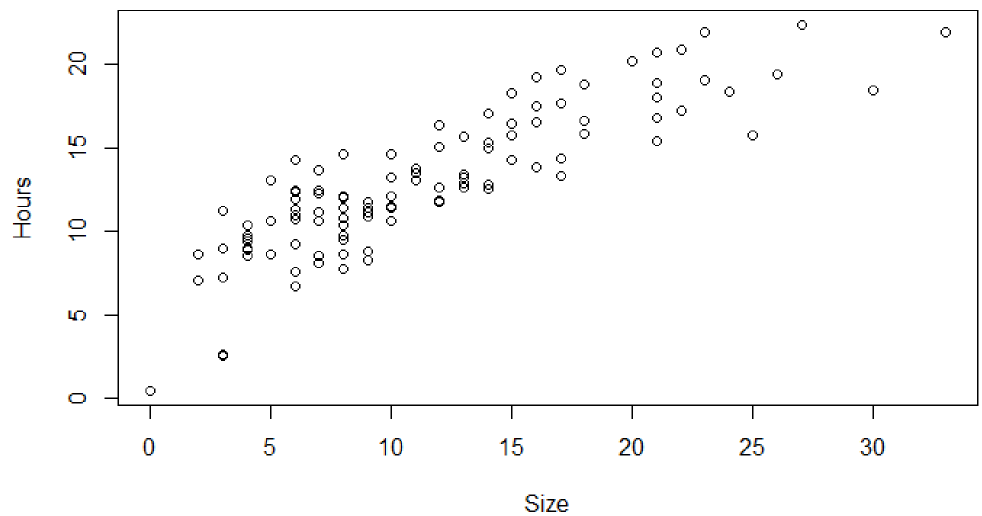
Many points don't actually fall on the line(not normal), but still linear.

3.16f

4.51731(.63768)^x

3.18a

```
f=file.choose()
ProdTime=read.table(f)
colnames(ProdTime)=c("Hours","Size")
attach(ProdTime)
plot(Size,Hours)
```



No. Transformation on X specifically to sqrt(X). Normality is the only issue.

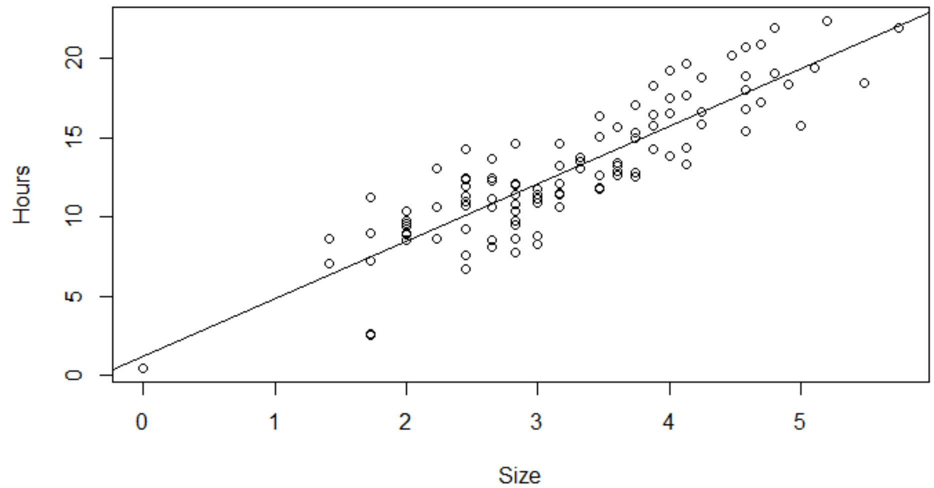
3.18b

```
Size=sqrt(Size)
NewProdMod=lm(Hours~Size)
NewProdMod

Call:
lm(formula = Hours ~ Size)

Coefficients:
(Intercept)         Size
      1.255        3.624
```

Y=1.255 + 3.624x’

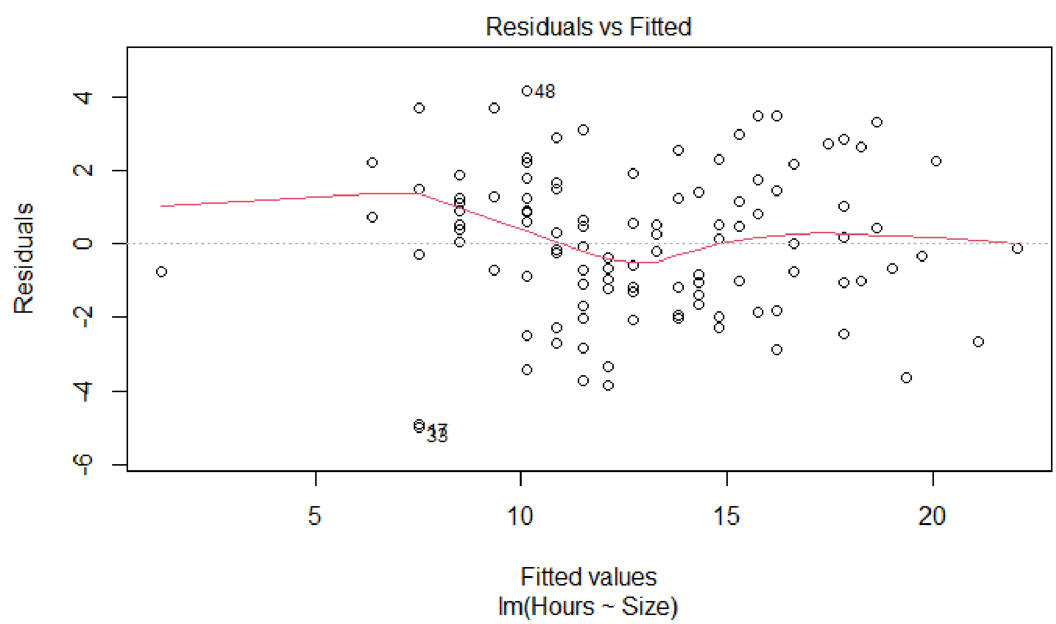
3.18c

```
plot(Size,Hours)
abline(NewProdMod)
```

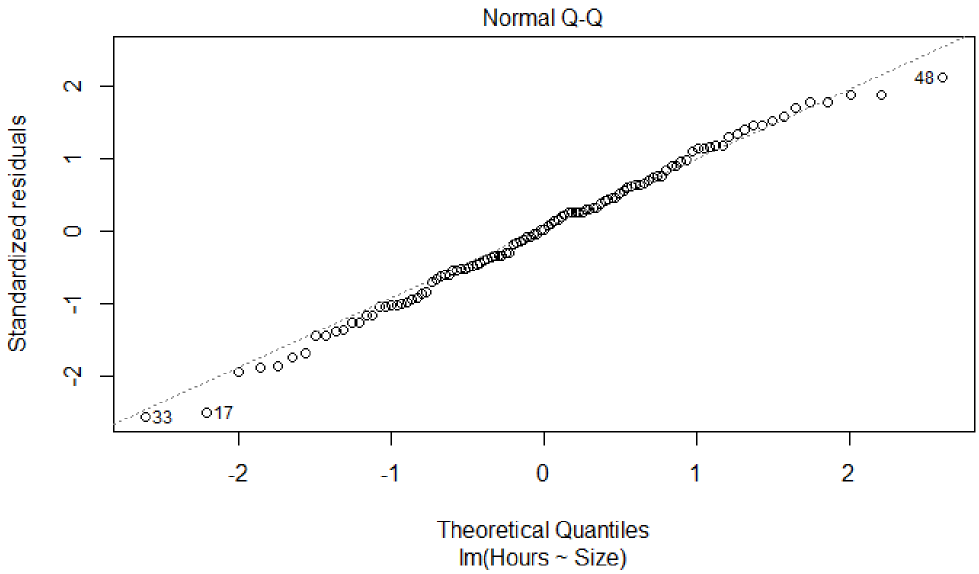


Yes

3.18d
plot(NewProdMod)



Even spread with few outliers



Normal with few outliers

3.18e

Y=1.255 + 3.624sqrt(X)

3.20

If the trasnformation is on x, it will remain normal, but if the transformation is on y, then it will not remain normal.

3)

a)

> sum(model$residuals)
[1] -2.942091e-15
Which is approximately 0

b)

Old

summary(model)

Call:
lm(formula = GPA ~ ACT, data = TestScores)

Residuals:

Min	1Q	Median	3Q	Max
-2.74004	-0.33827	0.04062	0.44064	1.22737

Coefficients:

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.11405    0.32089   6.588  1.3e-09 ***
ACT          0.03883    0.01277   3.040  0.00292 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6231 on 118 degrees of freedom
Multiple R-squared:  0.07262,	Adjusted R-squared:  0.06476
F-statistic:  9.24 on 1 and 118 DF,  p-value: 0.002917
```

New

```
TestScores[1,1]=.897
model=lm(GPA~ACT, data = TestScores)
summary(model)
Call:
lm(formula = GPA ~ ACT, data = TestScores)

Residuals:
     Min       1Q   Median       3Q      Max
-2.73511 -0.31973  0.04944  0.47245  1.30273

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.97295    0.33146   5.952 2.78e-08 ***
ACT          0.04352    0.01319   3.299  0.00128 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6436 on 118 degrees of freedom
Multiple R-squared:  0.08443,	Adjusted R-squared:  0.07667
F-statistic: 10.88 on 1 and 118 DF,  p-value: 0.001284
```

The residuals on the new got bumped up a little bit while the intercepts and p value went down.