6.18

```
f=file.choose()
Properties=read.table(f)
colnames(Properties)=c('Y','X1','X2','X3','X4')
```

a)

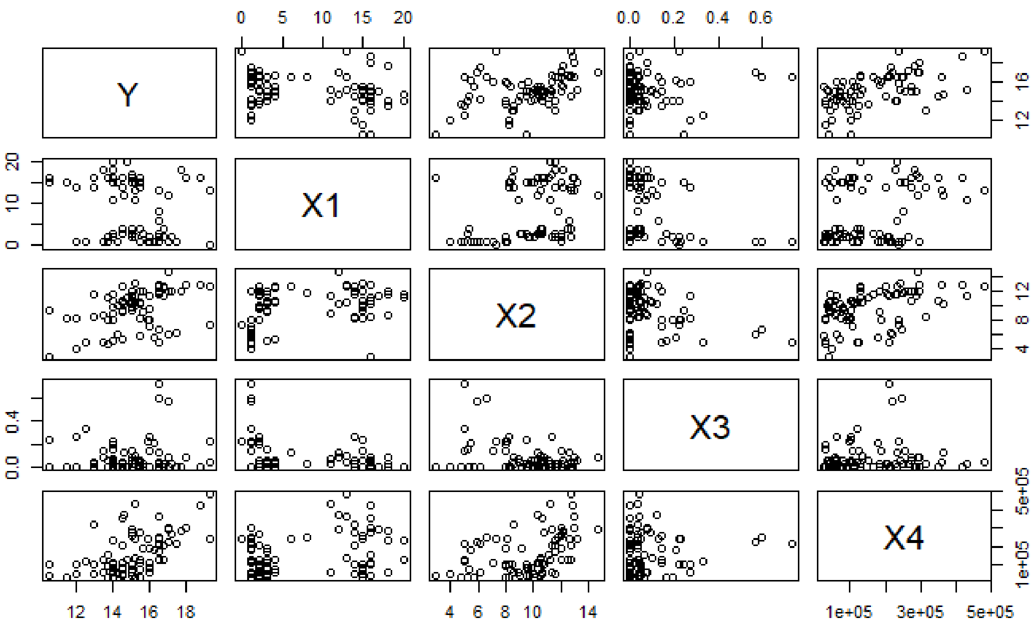
```
stem(Properties$X1)

  The decimal point is at the |

   0 | 0000000000000000
   2 | 0000000000000000000000
   4 | 00000
   6 | 0
   8 | 0
  10 | 00
  12 | 00000
  14 | 0000000000000
  16 | 0000000000
  18 | 000
  20 | 00

> stem(Properties$X2)

  The decimal point is at the |

   2 | 0
   4 | 080003358
   6 | 012613
   8 | 00001223456001555689
  10 | 013344566677778123344666668
  12 | 00011115777889002
  14 | 6

> stem(Properties$X3)

  The decimal point is 1 digit(s) to the left of the |

  0 | 00000000000000000000000000002333333333334444445555556678889
  1 | 023444469
```

```
  2 | 1223477
  3 | 3
  4 |
  5 | 7
  6 | 0
  7 | 3

> stem(Properties$X4)

  The decimal point is 5 digit(s) to the right of the |

  0 | 333333444444
  0 | 5556666677778899
  1 | 000001111222333334
  1 | 578889
  2 | 011122334444
  2 | 555788899
  3 | 002
  3 | 567
  4 | 23
  4 | 8
```

How spread out the data is

b)

```
plot(Properties)
```



> cor(Properties)

	Y	X1	X2	X3	X4
Y	1.00000000	-0.2502846	0.4137872	0.06652647	0.53526237
X1	-0.25028456	1.0000000	0.3888264	-0.25266347	0.28858350
X2	0.41378716	0.3888264	1.0000000	-0.37976174	0.44069713
X3	0.06652647	-0.2526635	-0.3797617	1.00000000	0.08061073
X4	0.53526237	0.2885835	0.4406971	0.08061073	1.00000000

Y and X4 have the highest correlation while Y and X3 have the smallest.

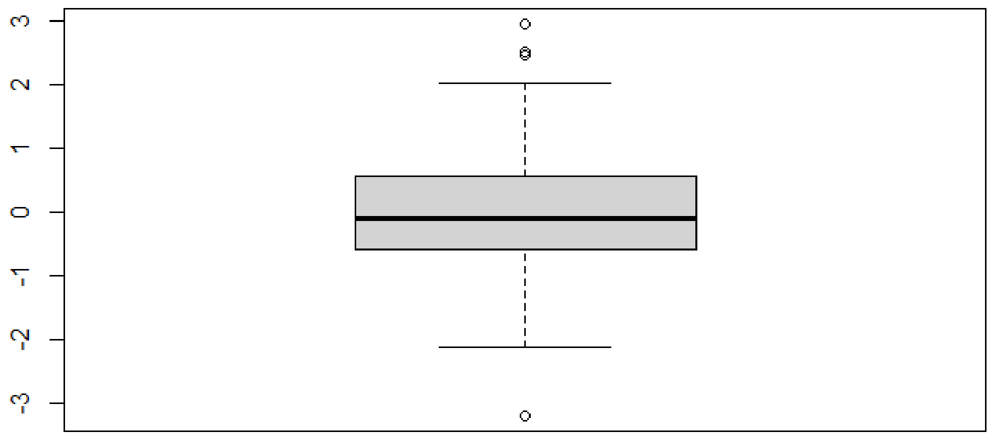
c)

```
> model=lm(Y~X1+X2+X3+X4,data = Properties)
> model

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4, data = Properties)

Coefficients:
(Intercept)         X1          X2          X3          X4
  1.220e+01   -1.420e-01   2.820e-01   6.193e-01   7.924e-06
```

$Y = 12.200 - .1420X1 + .2820X2 + 0.6193X3 + 0.0000079X4$

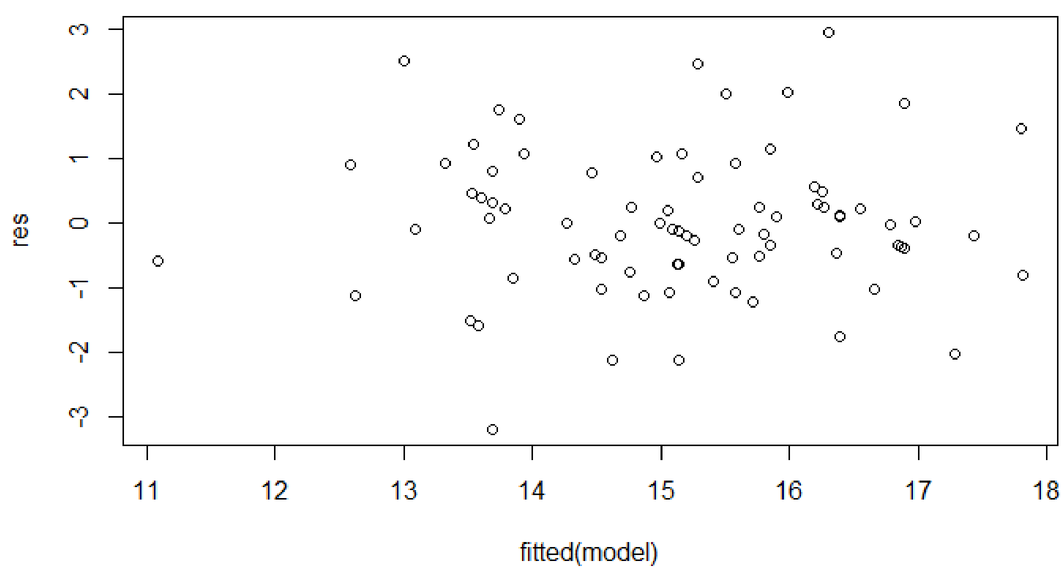
d)

```
res=resid(model)
boxplot(res)
```

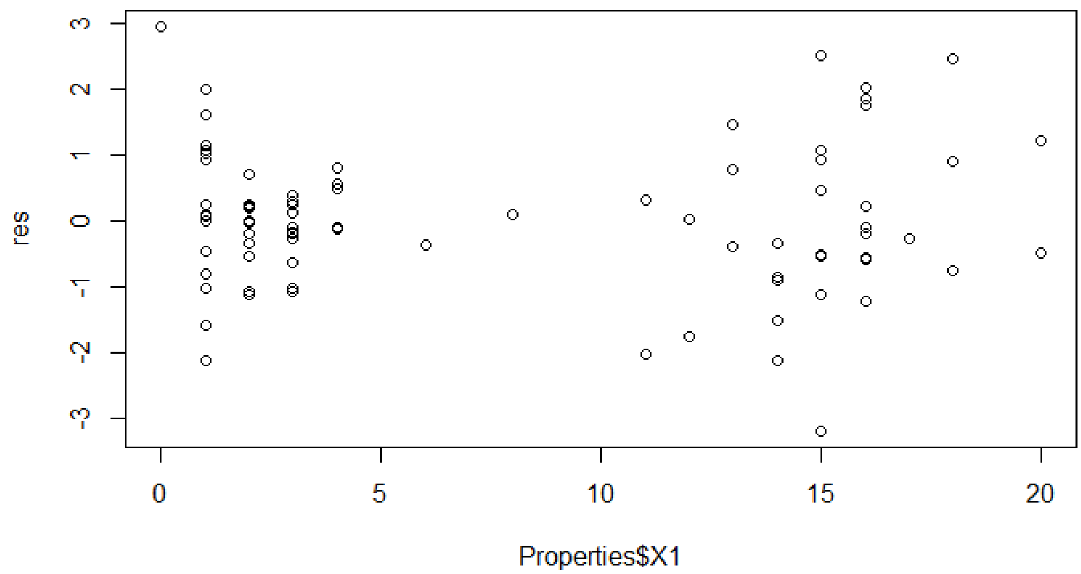


e)

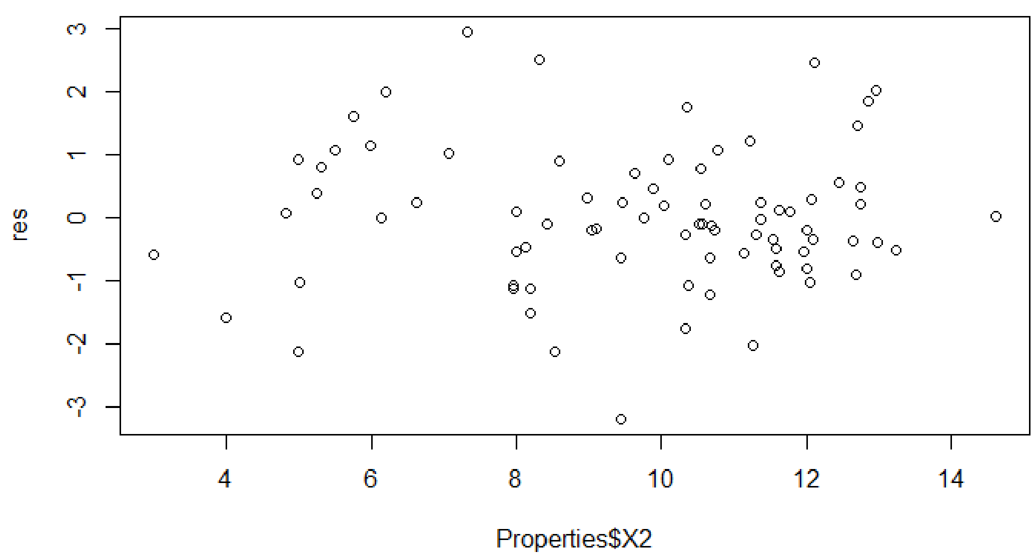
```
plot(fitted(model),res)
```



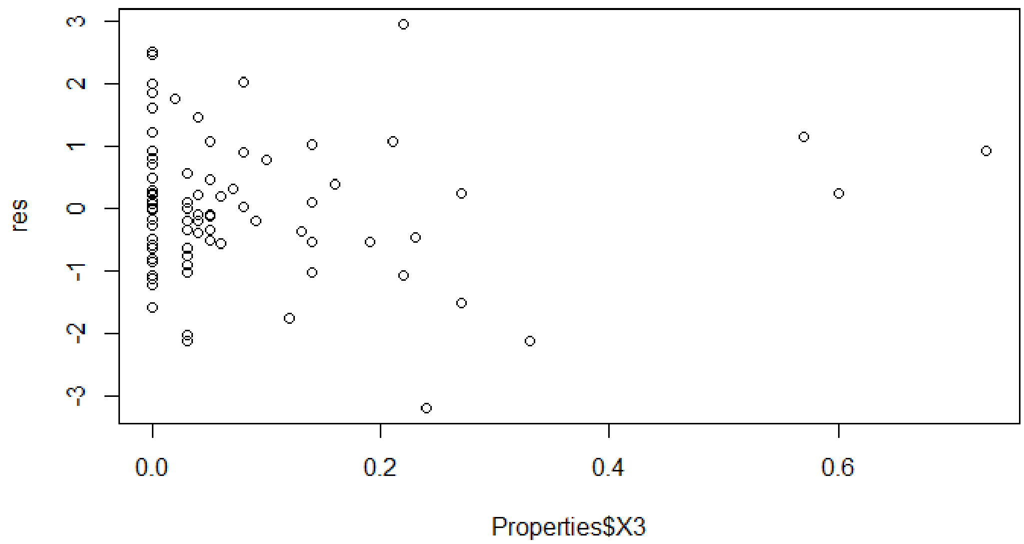

```
plot(Properties$X1,res)
```



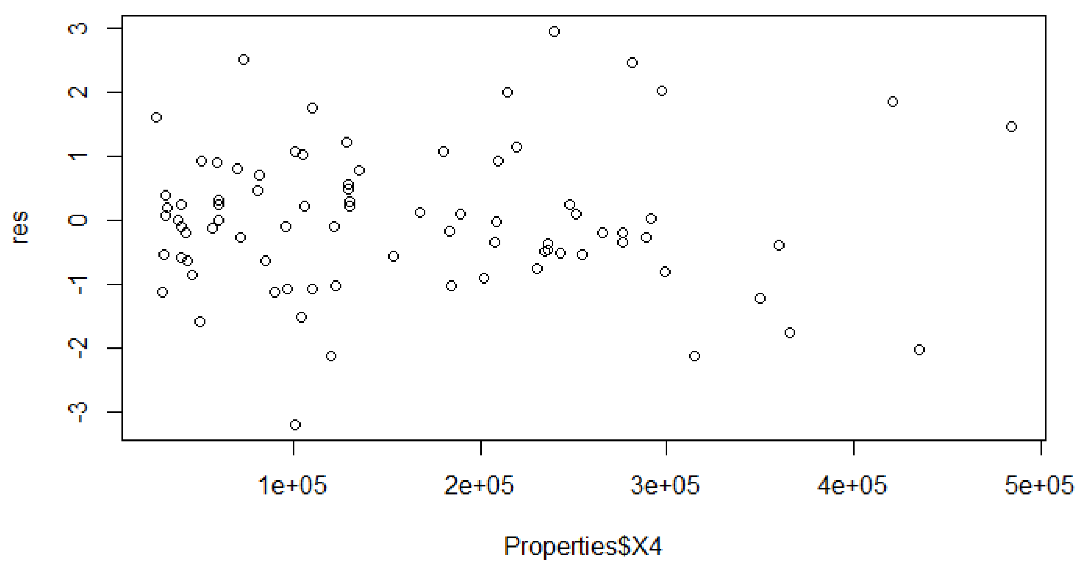
plot(Properties$X2,res)



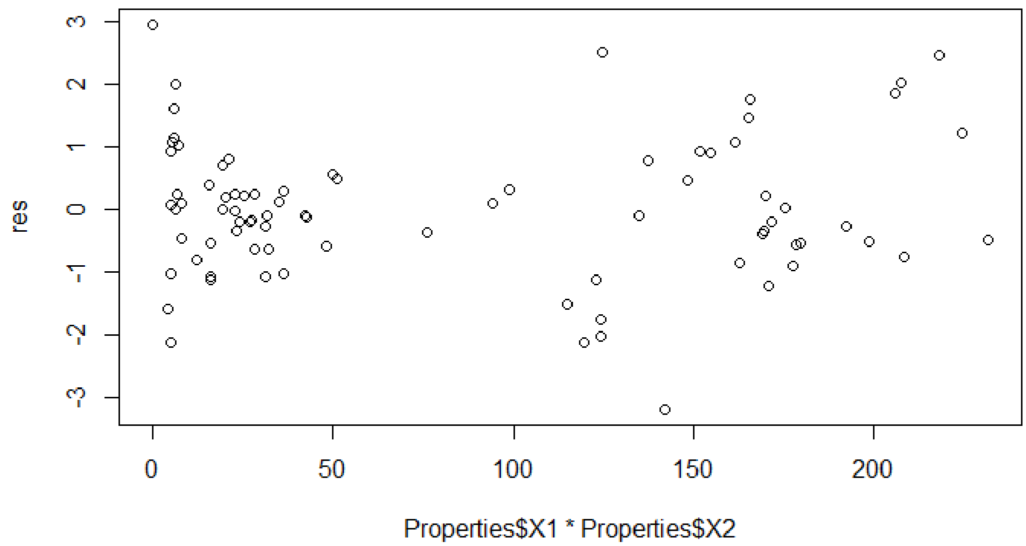
plot(Properties$X3,res)



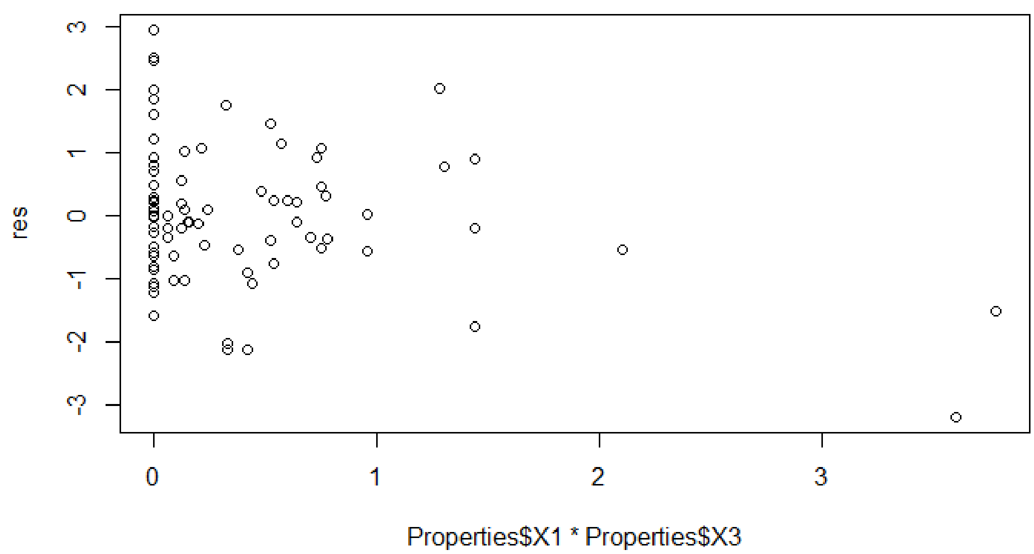
plot(Properties$X4,res)



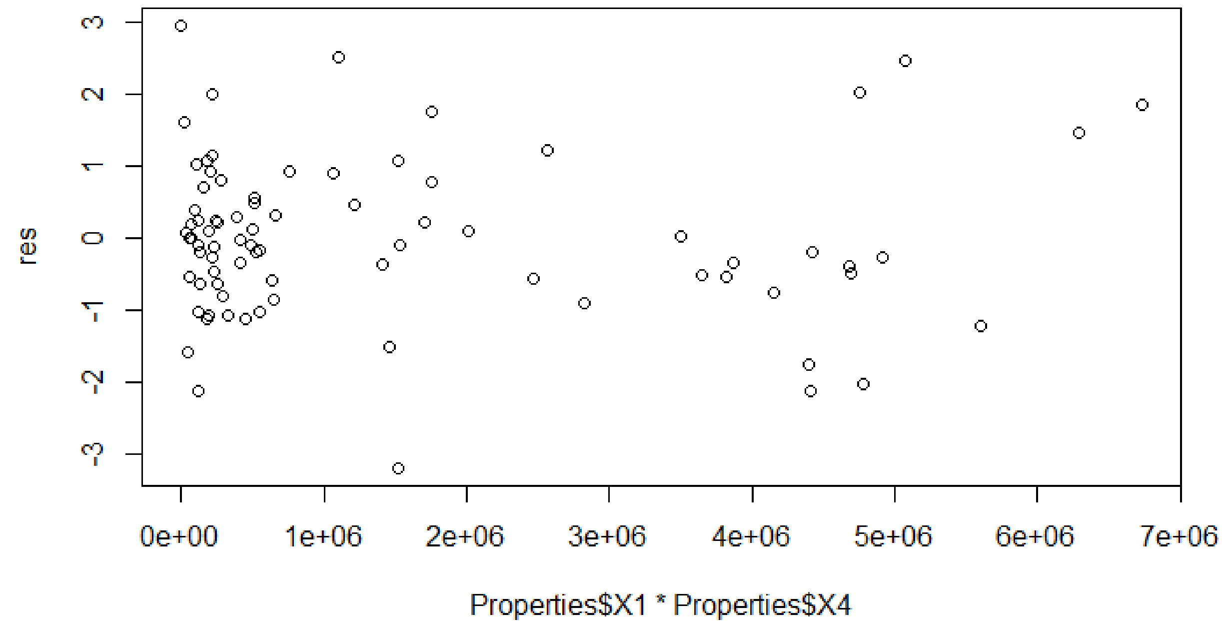
plot(Properties$X1*Properties$X2,res)



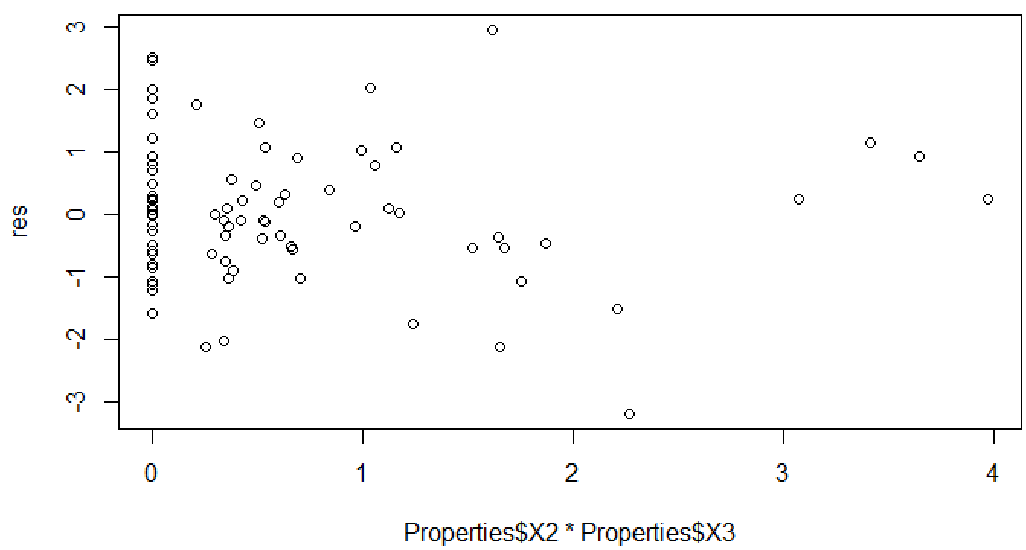
plot(Properties$X1*Properties$X3,res)



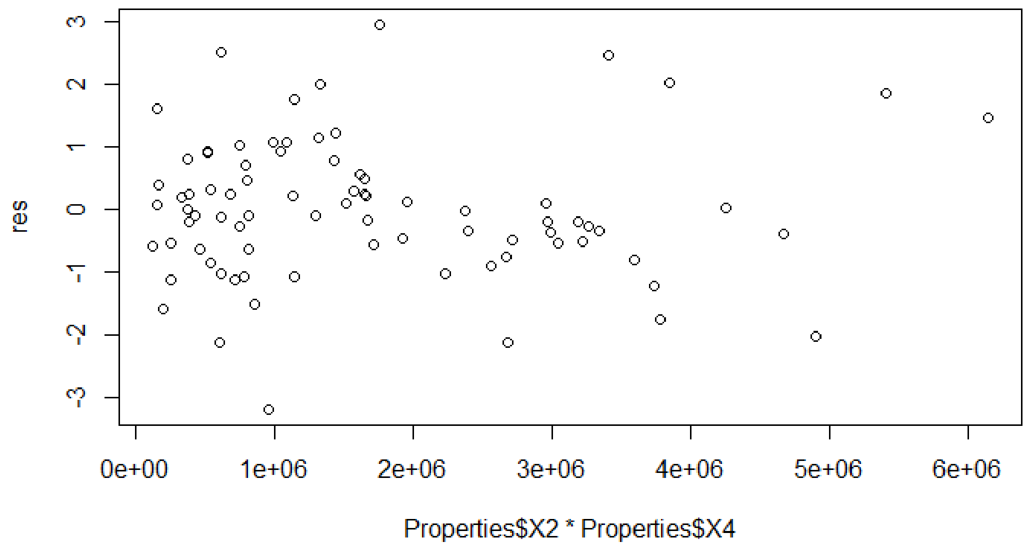
plot(Properties$X1*Properties$X4,res)



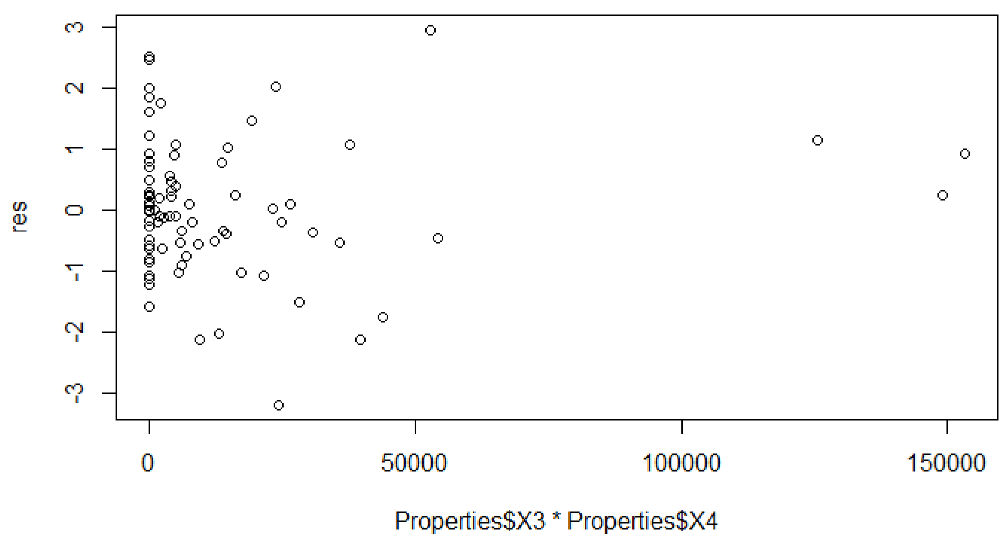
plot(Properties$X2*Properties$X3,res)



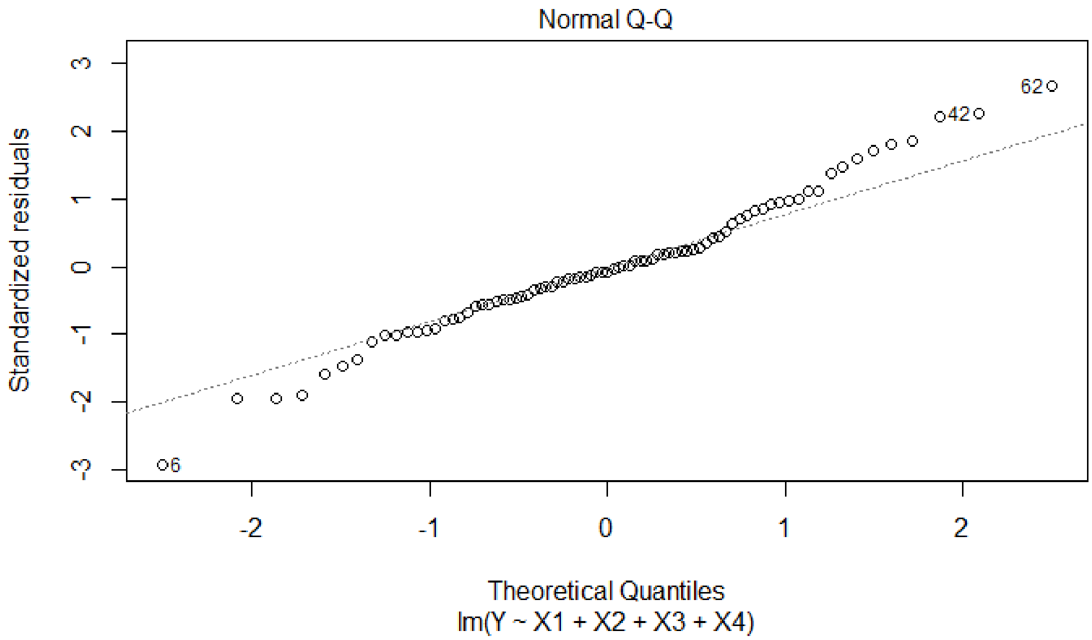
plot(Properties$X2*Properties$X4,res)



plot(Properties$X3*Properties$X4,res)



plot(model)



Looking at these plot versue fitted, it seems there is a linear relationship. None of the res vs predicto plots seem to have any linear or nonlinear relationship. The qq plot shows it is nomal with a fat tail.

f)

No, since it is linear

g)

```
Properties=cbind(Properties,fitted(model))
colnames(Properties)=c('Y','X1','X2','X3','X4','fitted')
attach(Properties)
Properties2=Properties[order(Properties$fitted),]
Properties3=data.frame(Properties2,Group = as.factor(rep(c("1","2"),times = c(40,41)))
bf.test(Y~Group,data = Properties3)
 Brown-Forsythe Test (alpha = 0.05)
-------------------------------------------------------------
 data : Y and Group

 statistic  : 73.41845
 num df     : 1
 denom df   : 78.21326
 p.value    : 7.282466e-13
```

Result : Difference is statistically significant.

P-value is less than .05 so error variance is constant

6.19

a)

H0: $\beta1 = \beta2 = \beta3 = \beta4 = 0$

Ha: not all $\beta k = 0$

```
 SSE=sum(model$residuals^2)
> MSE=SSE/76
> MSE
[1] 1.292508
> SST=var(Properties$Y)*(nrow(Properties-1))
> SSR=SST-SSE
> MSR=SSR/4
> MSR
[1] 35.32097
> Fstar=MSR/MSE
> Fstar
[1] 27.32747
> F=qf(.95,4,76)
> F
[1] 2.492049
```

Since Fstar is greater than F, we reject the null

```
pf(Fstar,4,76,lower.tail = FALSE)
[1] 4.573727e-14
```

P-value is close to 0

b)

```
> confint(model,level = 1-.05/4)
                 0.625 %      99.375 %
(Intercept)  1.072186e+01  1.367931e+01
X1          -1.966396e-01 -8.742769e-02
X2           1.203875e-01  4.436456e-01
X3          -2.161312e+00  3.399999e+00
X4           4.381297e-06  1.146731e-05
```

$-.1966 \leq \beta1 \leq -.0874$.

$.1204 \leq \beta2 \leq .4436$

$-2.1613 \leq \beta 3 \leq 3.3999$.
$.0000044 \leq \beta 4 \leq .0000114$

c)

```
> R2=SSR/SST
> R2
[1] 0.5898762
```

6.20

```
f=file.choose()
Properties4=read.table(f)
colnames(Properties4)=c('X1','X2','X3','X4')

predict(model,Properties4,interval="confidence",level = 1-.05/4)
       fit      lwr      upr
1 15.79813 15.08664 16.50962
2 16.02754 15.42391 16.63116
3 15.90072 15.33232 16.46913
4 15.84339 15.18040 16.50638
```

6.21

```
> f=file.choose()
> Properties6=read.table(f)
> colnames(Properties6)=c('X1','X2','X3','X4')
> predict(model,Properties6,interval = "predict",level = .95)
       fit      lwr      upr
1 15.14850 12.85249 17.44450
2 15.54249 13.24504 17.83994
3 16.91384 14.53469 19.29299
```

3

```
f=file.choose()
MathSal=read.table(f)
colnames(MathSal)=c('Y','X1','X2','X3')
```

a)

```
> model=lm(Y~X1+X2+X3,data = MathSal)
> model

Call:
```

```
lm(formula = Y ~ X1 + X2 + X3, data = MathSal)

Coefficients:
(Intercept)          X1           X2           X3
    21.0045       0.6169       0.2807       1.4354
```

Y=21.0045+.6169X1+.2807X2+1.4354X3

H0: β1 = β2 = β3 = 0
Ha: not all βk = 0

```
> SSE=sum(model$residuals^2)
> MSE=SSE/20
> MSE
[1] 6.031172
> SST=var(MathSal$Y)*(nrow(MathSal-1))
> SSR=SST-SSE
> MSR=SSR/3
> MSR
[1] 167.6826
> Fstar=MSR/MSE
> Fstar
[1] 27.80266
> F=qf(.95,3,20)
> F
[1] 3.098391
```

Since Fstar is greater than F, we reject the null

```
> pf(Fstar,3,20,lower.tail = FALSE)
[1] 2.46018e-07
```

P-value is close to 0

```
> R2=SSR/SST
> R2
[1] 0.8065913
```

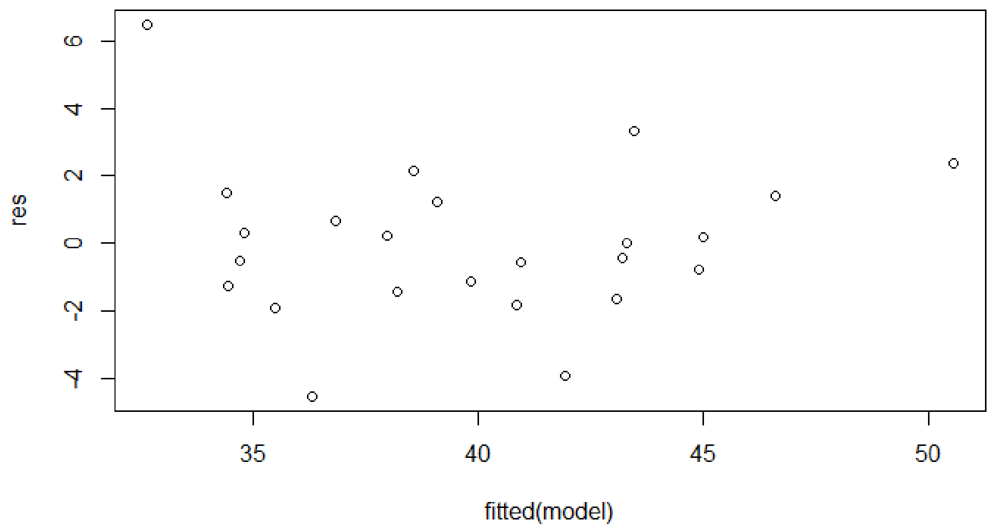
b)

```
> confint(model,level = .95)
                 2.5 %     97.5 %
(Intercept) 15.4252618 26.5837049
X1          -0.3110263  1.5447576
```

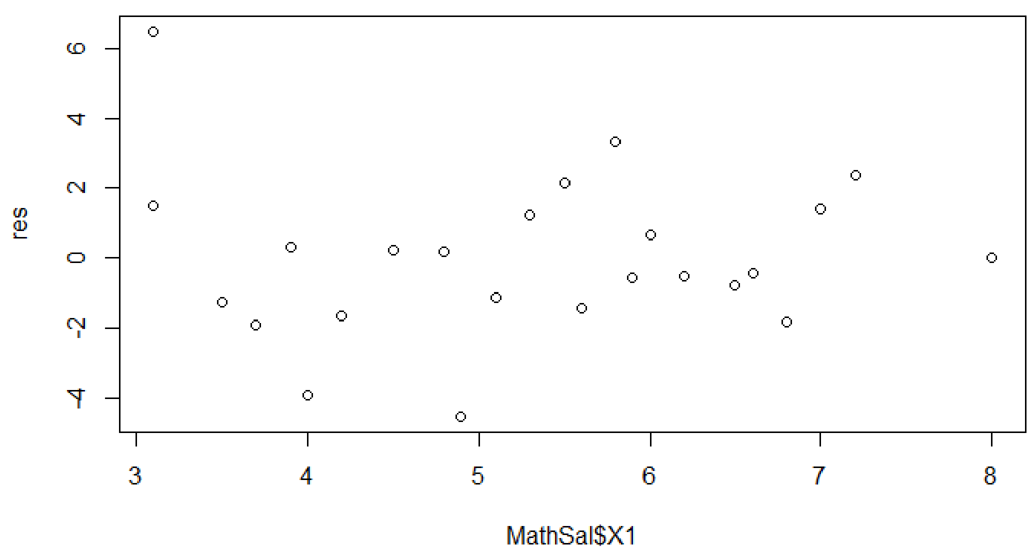
```
X2          0.1736472  0.3878391
X3          0.5409409  2.3297888

> summary(model)

Call:
lm(formula = Y ~ X1 + X2 + X3, data = MathSal)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5016 -1.2872 -0.2034  1.2687  6.4544 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 21.00448    2.67465   7.853 1.55e-07 ***
X1           0.61687    0.44483   1.387  0.18078    
X2           0.28074    0.05134   5.468 2.37e-05 ***
X3           1.43536    0.42878   3.348  0.00321 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.456 on 20 degrees of freedom
Multiple R-squared:  0.7982,  Adjusted R-squared:  0.7679 
F-statistic: 26.37 on 3 and 20 DF,  p-value: 3.748e-07
```

c)

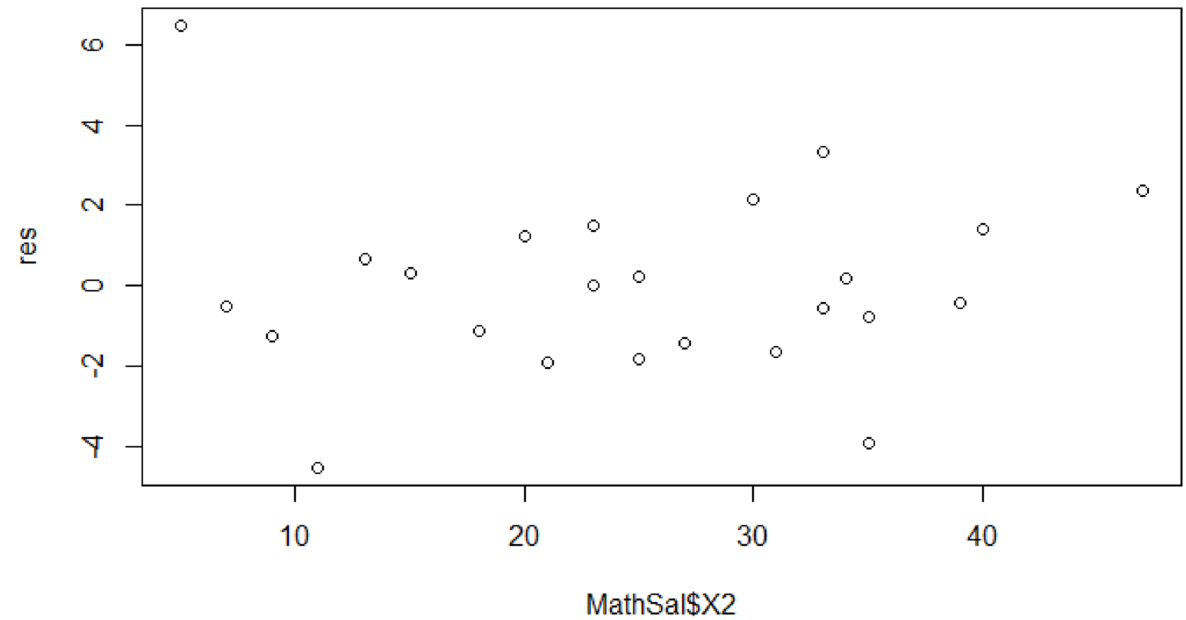
```
plot(fitted(model),res)
```



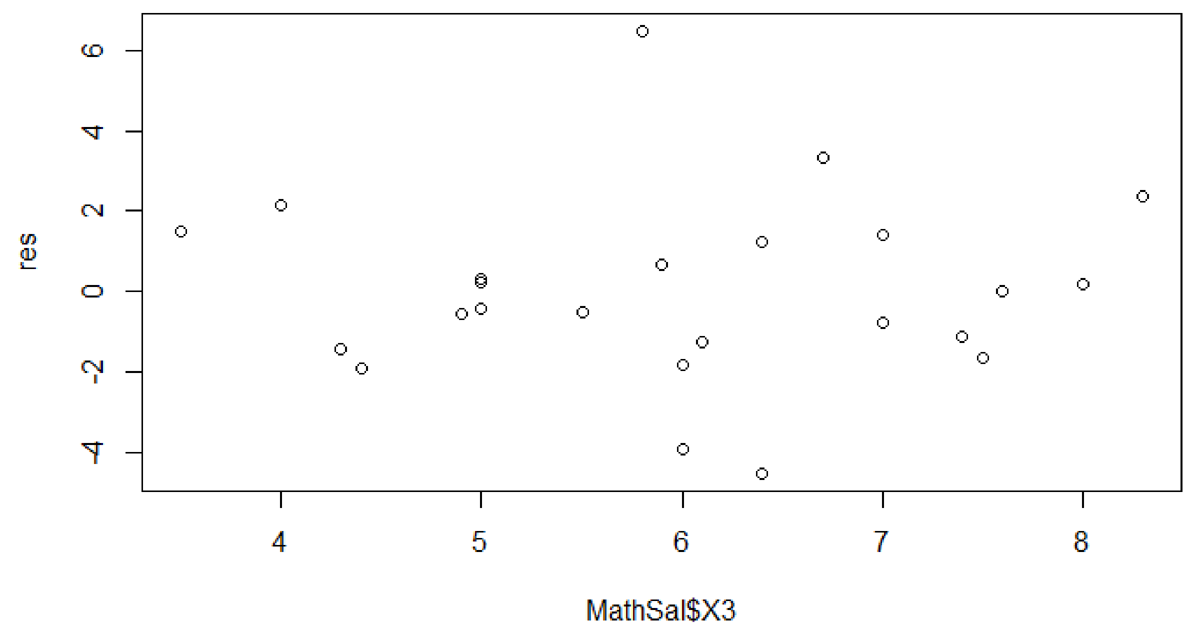
plot(MathSal$X1,res)



plot(MathSal$X2,res)



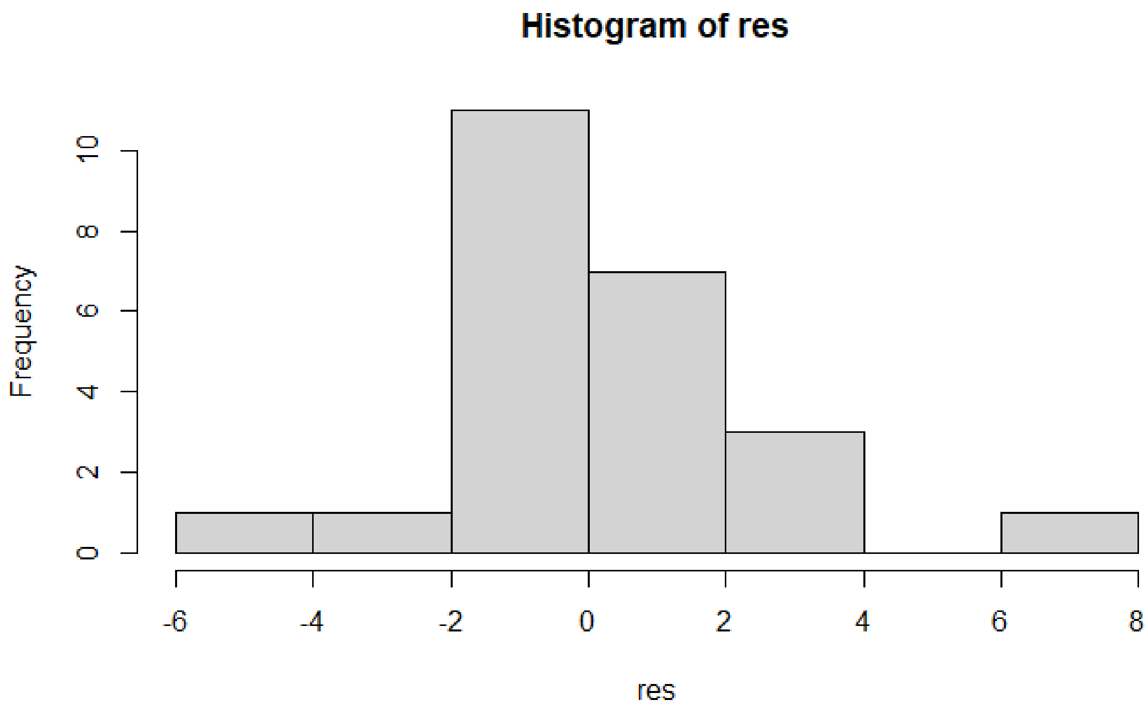
plot(MathSal$X3,res)



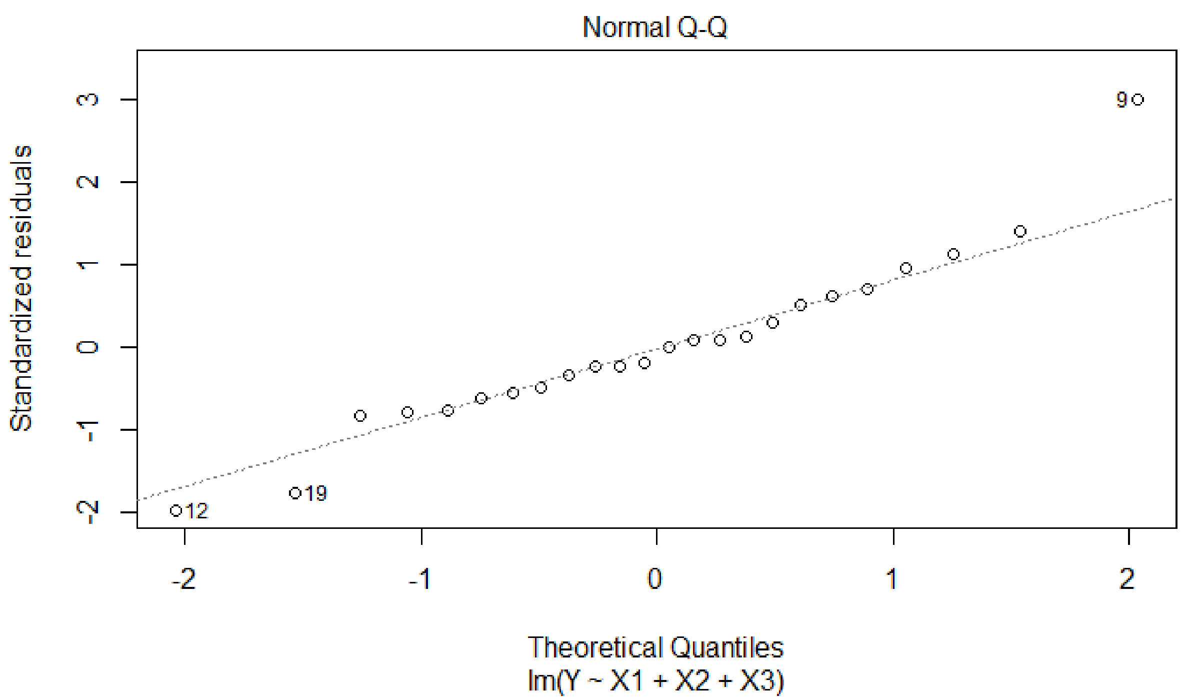
No, there are not any unusual patterns.

d)

hist(res)



plot(model)



Appears to be fairly normal

e)

```
predict(model,data.frame(X1=6.2,X2=8,X3=5.9),level = .95,interval = "prediction")
       fit      lwr      upr
1 35.54365 29.83829 41.24901
```