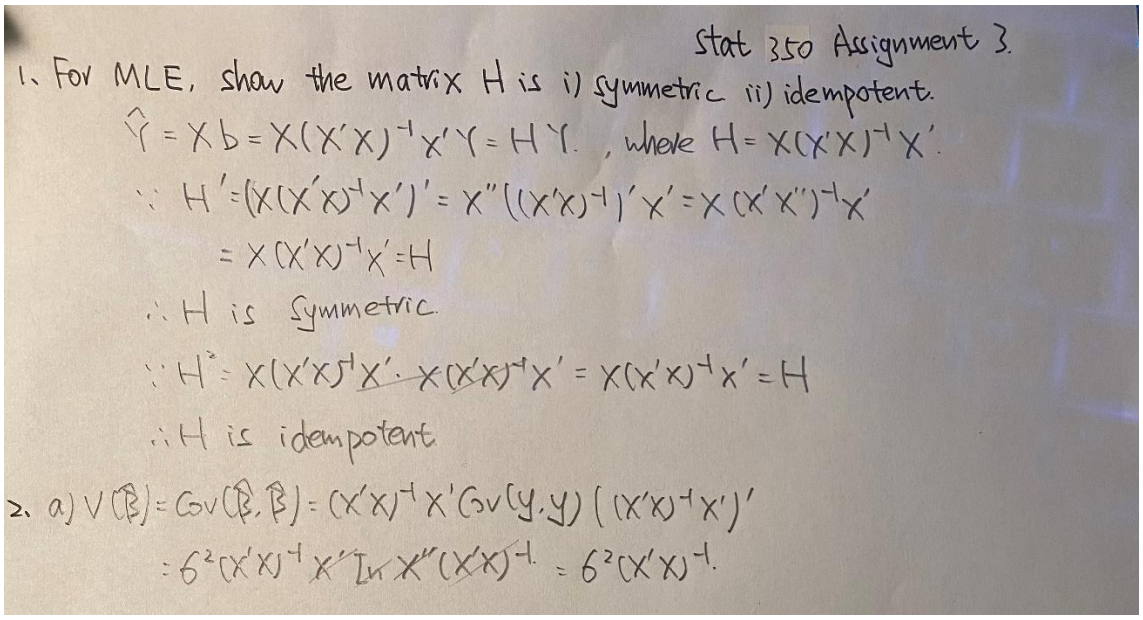
# STAT350

## Assignment3

Yuhan Zhang
301345627

Q1 +Q2 a



Stat 350 Assignment 3.

1. For MLE, show the matrix H is i) symmetric ii) idempotent.

$\hat{Y} = Xb = X(X'X)^{-1}X'Y = HY$. , where $H = X(X'X)^{-1}X'$.

$\because H' = (X(X'X)^{-1}X')' = X''((X'X)^{-1})'X' = X(X'X'')^{-1}X'$

$= X(X'X)^{-1}X' = H$

$\therefore H$ is symmetric.

$\because H^2 = X(X'X)^{-1}X' \cdot X(X'X)^{-1}X' = X(X'X)^{-1}X' = H$

$\therefore H$ is idempotent

2. a) $V(\hat{\beta}) = Cov(\hat{\beta}, \hat{\beta}) = (X'X)^{-1}X'Cov(y,y)((X'X)^{-1}X')'$

$= \sigma^2(X'X)^{-1}X'I_n X''(X'X)^{-1} = \sigma^2(X'X)^{-1}$.

Q2 b

```
> sigma_sq1 = 2  #sigma square =2 is given in the question
> sigma_sq1*solve(t(X) %*% X)
                        x1           x2
    7.575272 -1.519895879 -0.721752043
x1 -1.519896  0.434863250  0.008757325
x2 -0.721752  0.008757325  0.231715454
> sigmasq_2 = 2.703  #ANOVA shown the estimate of MSE is 2.703
> sigmasq_2*solve(t(X) %*% X)
                        x1          x2
   10.2379804 -2.05413928 -0.97544789
x1 -2.0541393  0.58771768  0.01183553
x2 -0.9754479  0.01183553  0.31316344
> vcov(fit1)
             (Intercept)          x1          x2
(Intercept)  10.2386726 -2.05427817 -0.97551384
x1           -2.0542782  0.58775742  0.01183633
x2           -0.9755138  0.01183633  0.31318461
>
```

The estimate is the first graph when using sigma=2, which is given in the question.

(The estimate is the second graph when using sigma=2.703, which is the MSE shown by the ANOVA table. The estimate is the third graph when just calculating vcov using the data.)

Q2 c

```
> 2*(1-hii[1])
[1] 1.316208
> 2*(1-hii[3])
[1] 1.295119
>
```

The variance is 1.3162 and 1.2951.

Q2 d

```
> -sigma_sq1*H[1,3]
[1] -0.4239149
```

The covariance is -0.4239.

Q3 a

```
> fit2  <- lm(location~. , data = mydata)
> round(vcov(fit2),5)
            (Intercept) thickness   energy       flux     gamma     opacity
(Intercept)   935.23953  -8.00881 -0.16376  -62.60013 -19.31702  -509.01639
thickness      -8.00881   0.39997 -0.00011    0.04180   0.13247     3.18406
energy         -0.16376  -0.00011  0.00005   -0.00439  -0.00112    -0.01784
flux          -62.60013   0.04180 -0.00439   53.93637  -0.28566  -102.59371
gamma         -19.31702   0.13247 -0.00112   -0.28566  20.30818    18.19048
opacity      -509.01639   3.18406 -0.01784 -102.59371  18.19048 10520.06100
> vcov(fit2)[2,2]
[1] 0.3999662
```

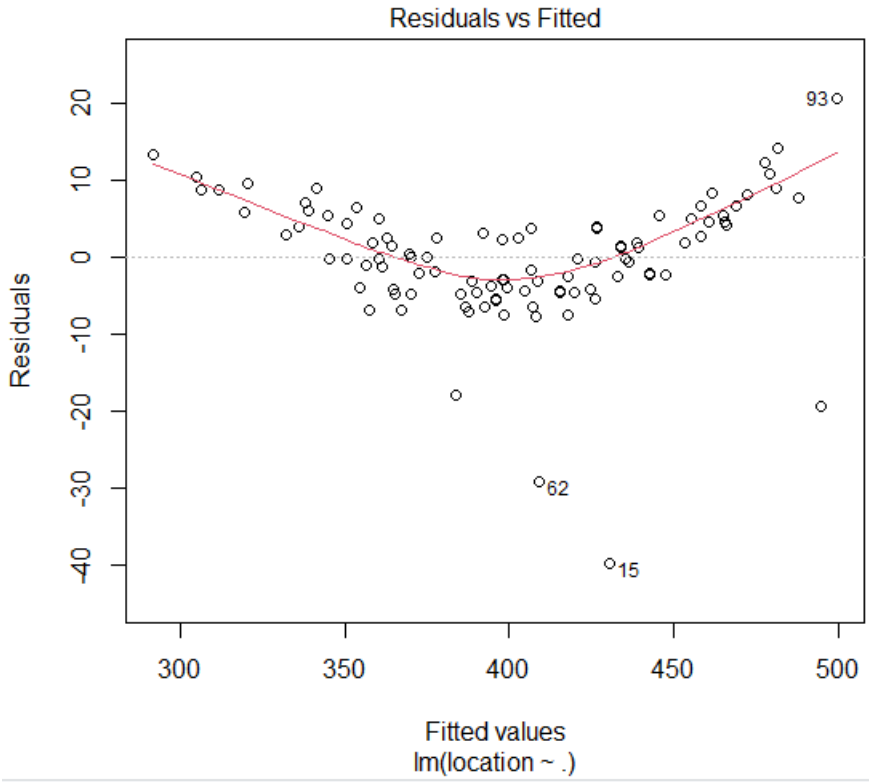
The estimated variance is 0.39997.

Q3 b

```
> vcov(fit2)[2,6]
[1] 3.18406
```

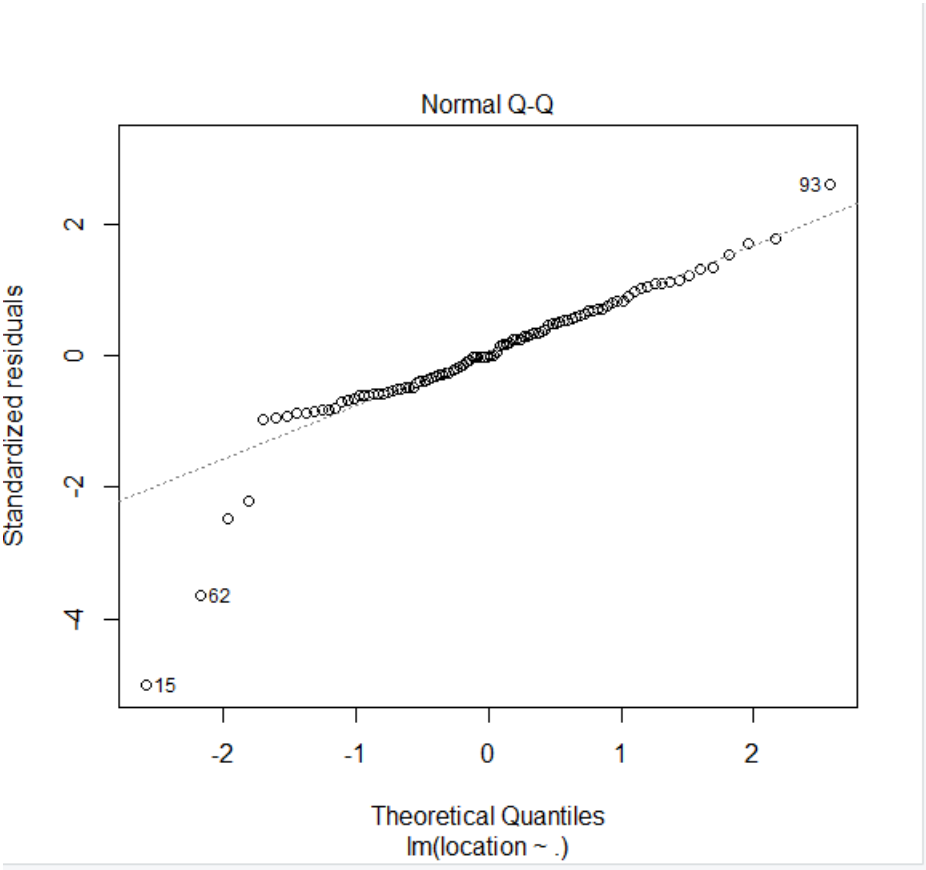
The estimated variance is 3.18406.

## Q3 c


Residuals vs Fitted
Residuals
Fitted values
lm(location ~ .)
93
62
15
20
10
0
-10
-20
-30
-40
300
350
400
450
500

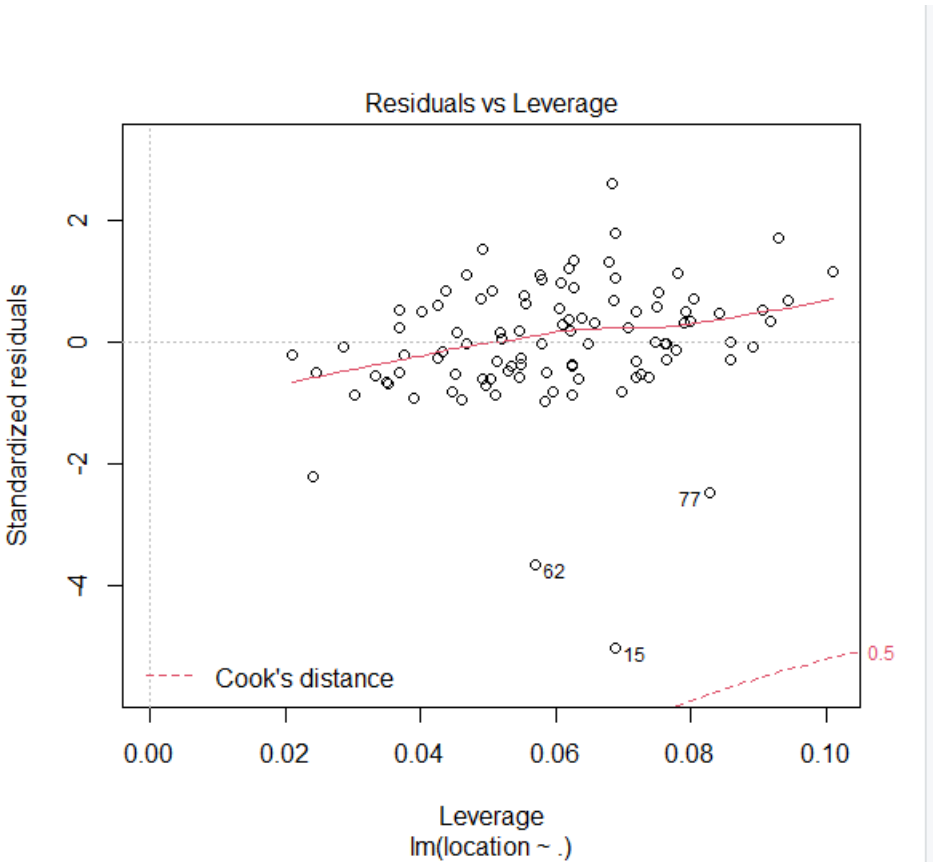
Plot (fit2, which=1)

i) The distribution has constant variance.
ii) The distribution is non-linear in this graph.
iii) There are some extreme outliers.




Plot (fit2, which=2)

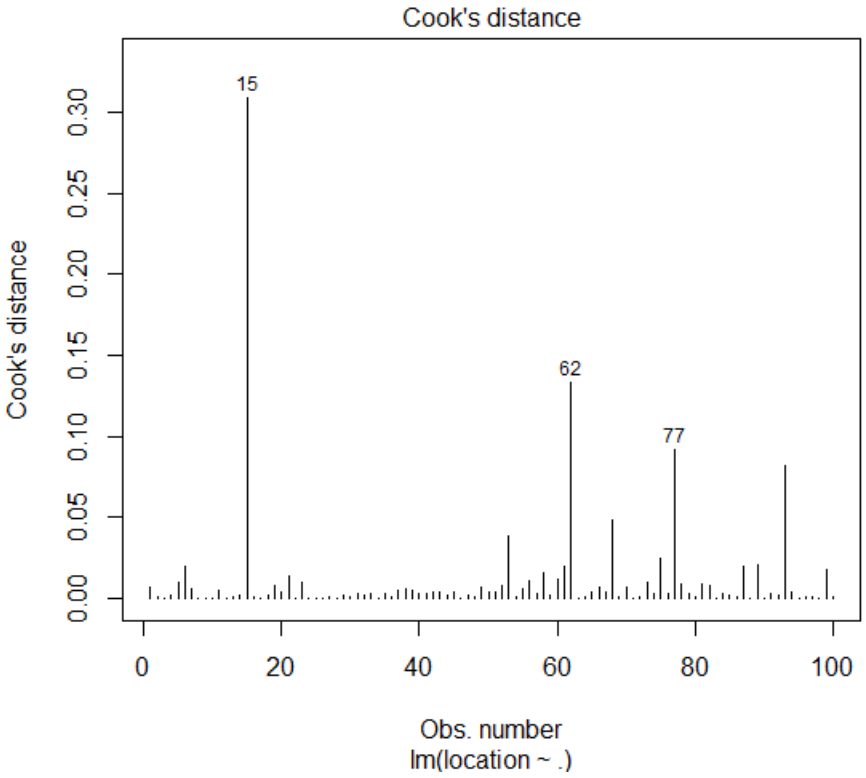
ii) The distribution is normal in general but obviously there are some extreme outliers




plot(fit2 ,which=5)

v) no influential observation cooks distance is small smaller than 0.5

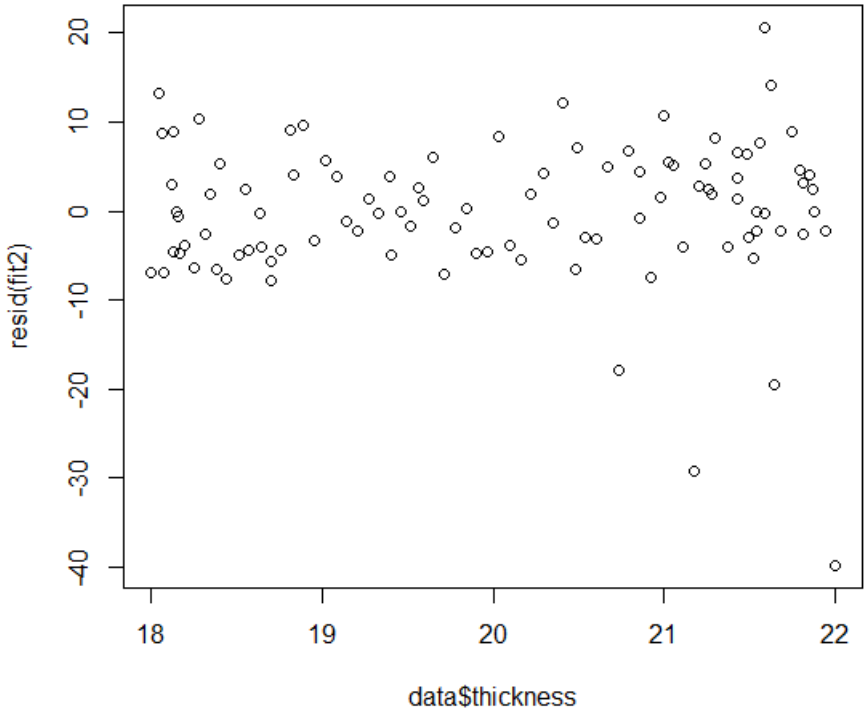
iii) no leverage points

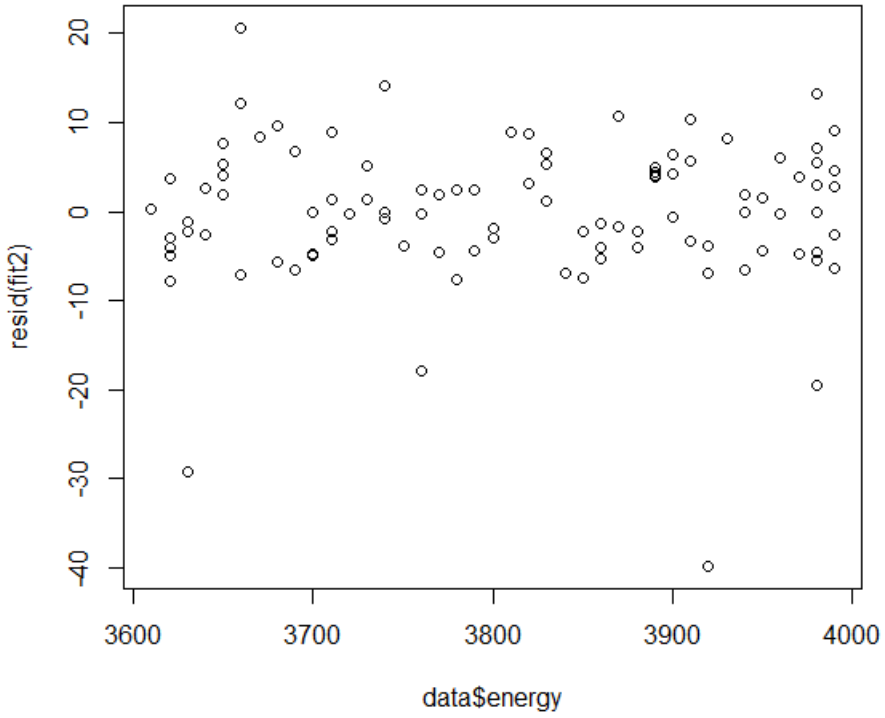


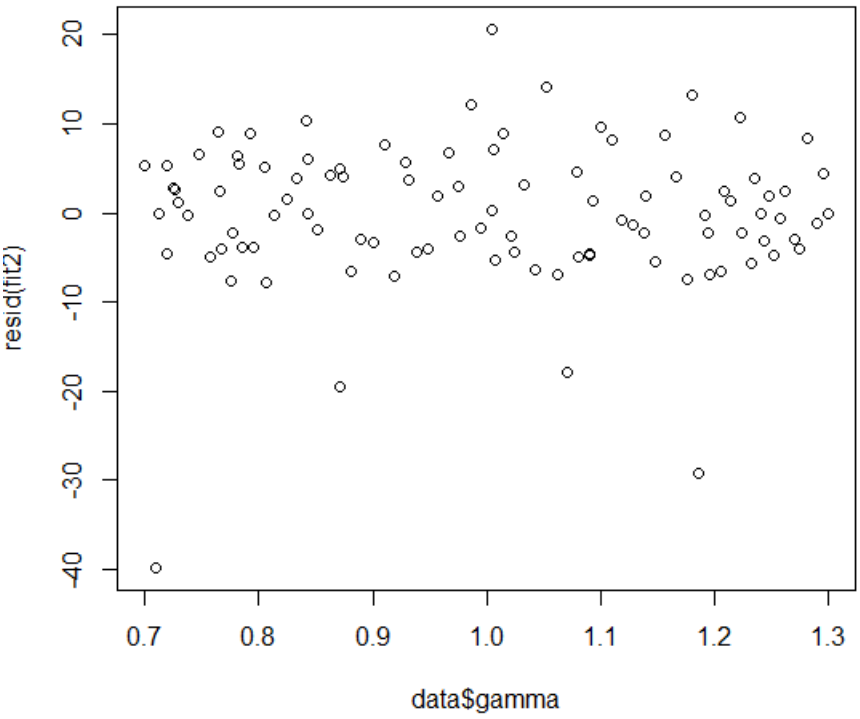

plot(fit2,which=4)

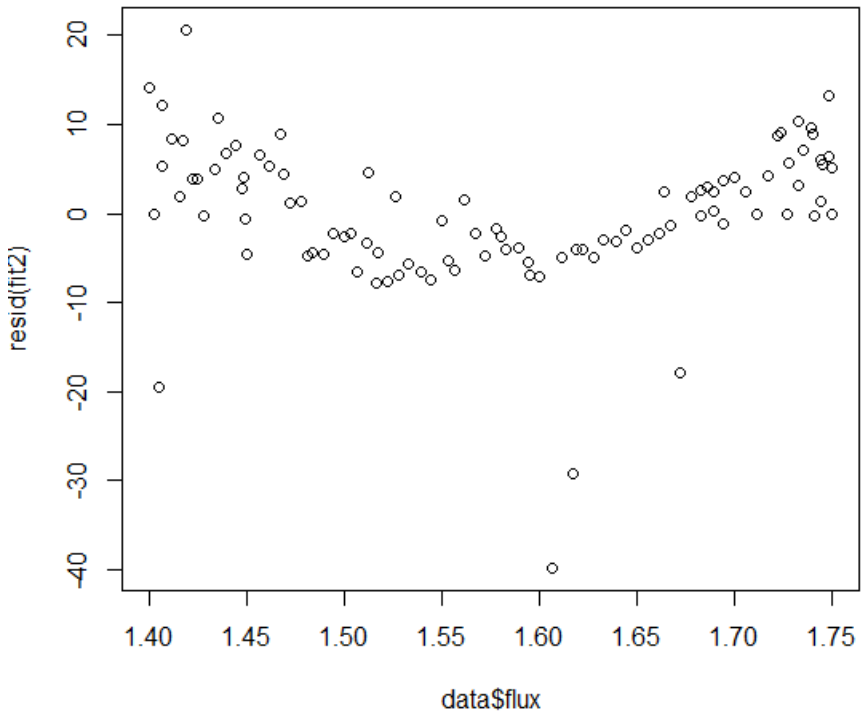
v) As the cook's distance are all smaller than 0.5. The cook's distance is small, so there is no influential observations

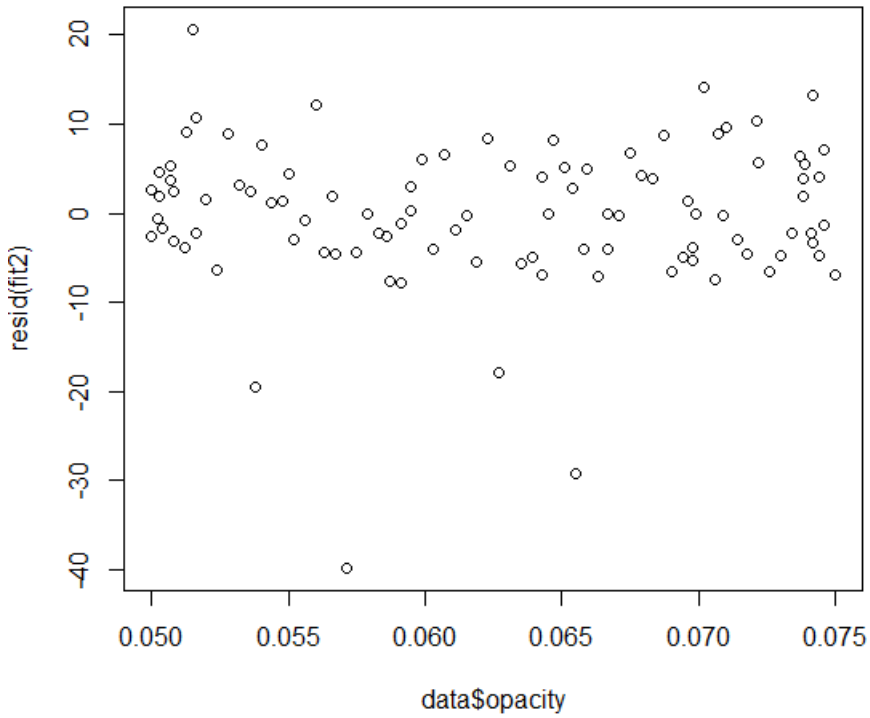
vi)




resid(fit2)
20
10
0
-10
-20
-30
-40
3600
3700
3800
3900
4000
data$energy


resid(fit2)
20
10
0
-10
-20
-30
-40
0.7
0.8
0.9
1.0
1.1
1.2
1.3
data$gamma


resid(fit2)
20
10
0
-10
-20
-30
-40
1.40
1.45
1.50
1.55
1.60
1.65
1.70
1.75
data$flux


resid(fit2)
20
10
0
-10
-20
-30
-40
0.050
0.055
0.060
0.065
0.070
0.075
data$opacity

The relationship between predictors and variables are linear expect for **flux.** Therefore, some transformation is needed for flux.