Flipped Assignment 10

Group 5

2022/3/1

Input Data

```
setwd('C:/Users/ndong/OneDrive - Texas Tech University/IE 5344 Statistical Data Analysis/Flipped Assignm
data <- read.csv('data-SoftDrinkDeliveryTime.csv', header = TRUE)
data <- data[,-1]
colnames(data) <- c('y', 'x1', 'x2')
head(data)
```

```
##       y x1  x2
## 1 16.68  7 560
## 2 11.50  3 220
## 3 12.03  3 340
## 4 14.88  4  80
## 5 13.75  6 150
## 6 18.11  7 330
```

Part a.

```
fit <- lm(y~. + I(x1^2) + I(x2^2) + x1:x2,data)
summary(fit)
```

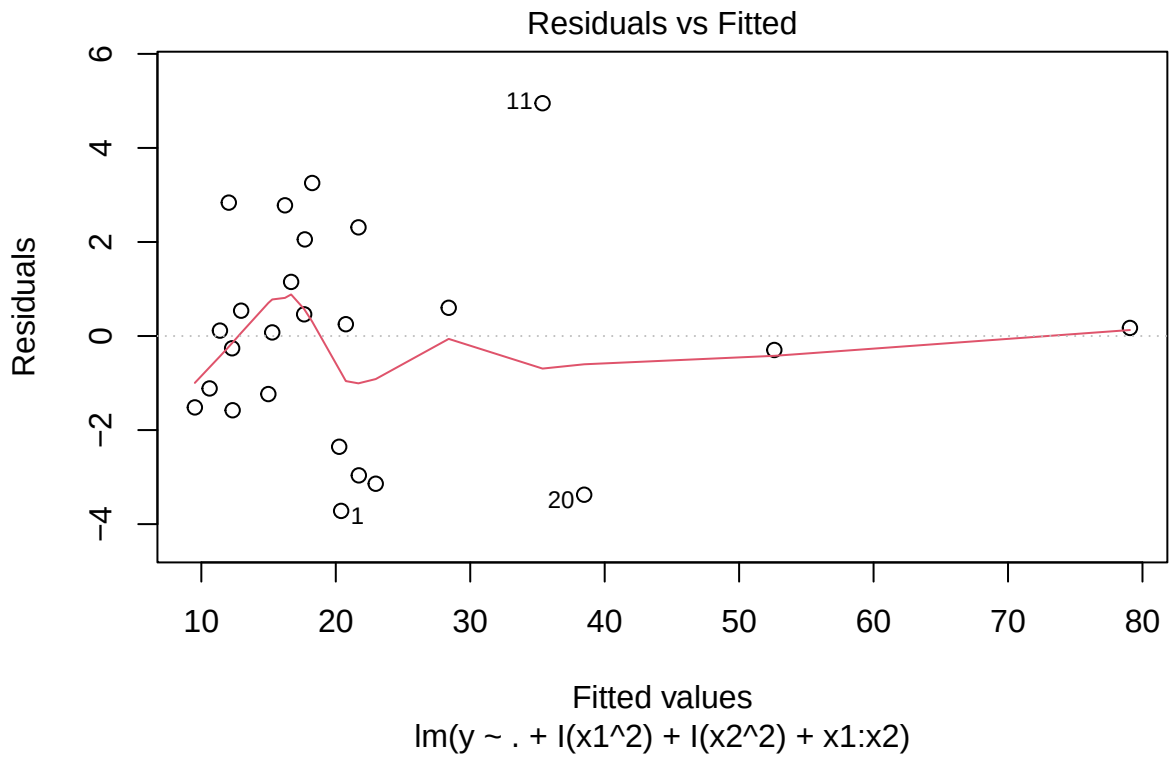
```
##
## Call:
## lm(formula = y ~ . + I(x1^2) + I(x2^2) + x1:x2, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.7187 -1.5177  0.1156  1.1509  4.9508
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.669e+00  1.570e+00   4.248 0.000435 ***
## x1          1.292e+00  3.706e-01   3.487 0.002469 **
## x2          7.832e-04  6.221e-03   0.126 0.901133
## I(x1^2)     7.585e-04  2.812e-02   0.027 0.978758
## I(x2^2)     1.109e-05  1.587e-05   0.699 0.493091
## x1:x2       1.865e-04  1.186e-03   0.157 0.876670
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.512 on 19 degrees of freedom
```

```
## Multiple R-squared:  0.9793, Adjusted R-squared:  0.9738
## F-statistic: 179.6 on 5 and 19 DF,  p-value: 2.613e-15
```

So $\hat{y} = 2.341231 + 1.615907x_1 + 0.014385x_2$. The R^2 is 0.9596 and the adjusted one is 0.9559.

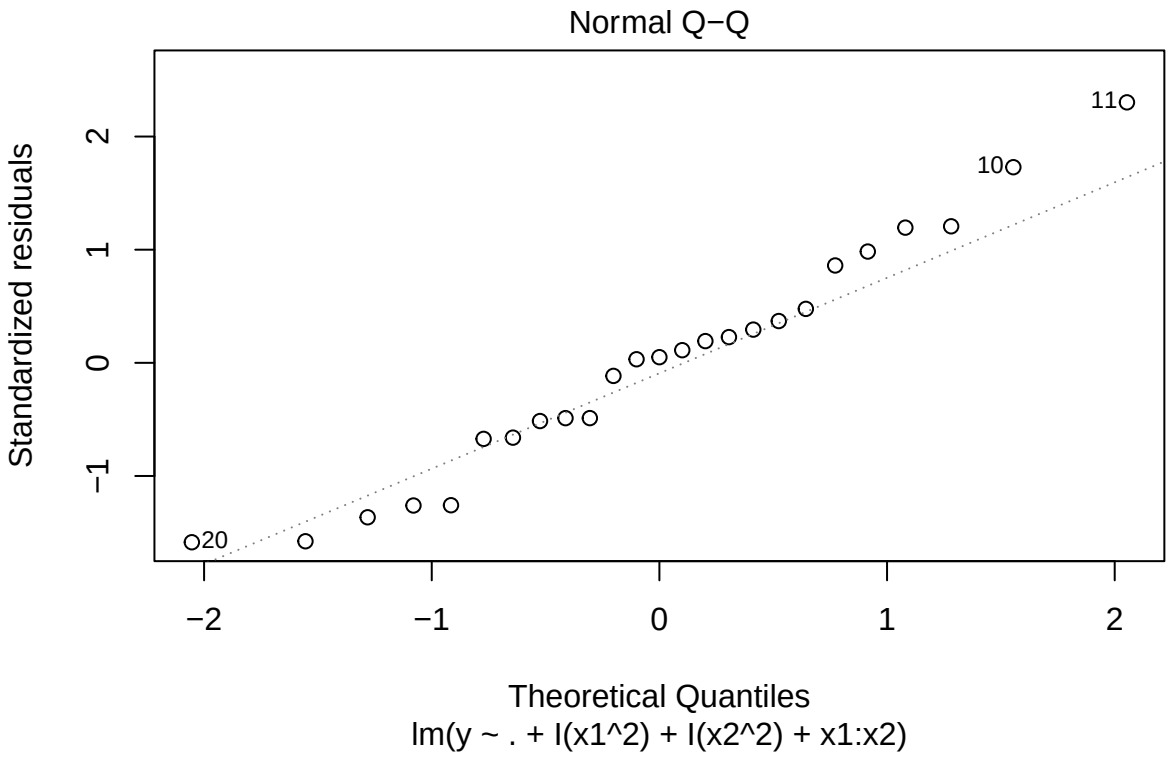
Part b.

```
plot(fit,1)
```



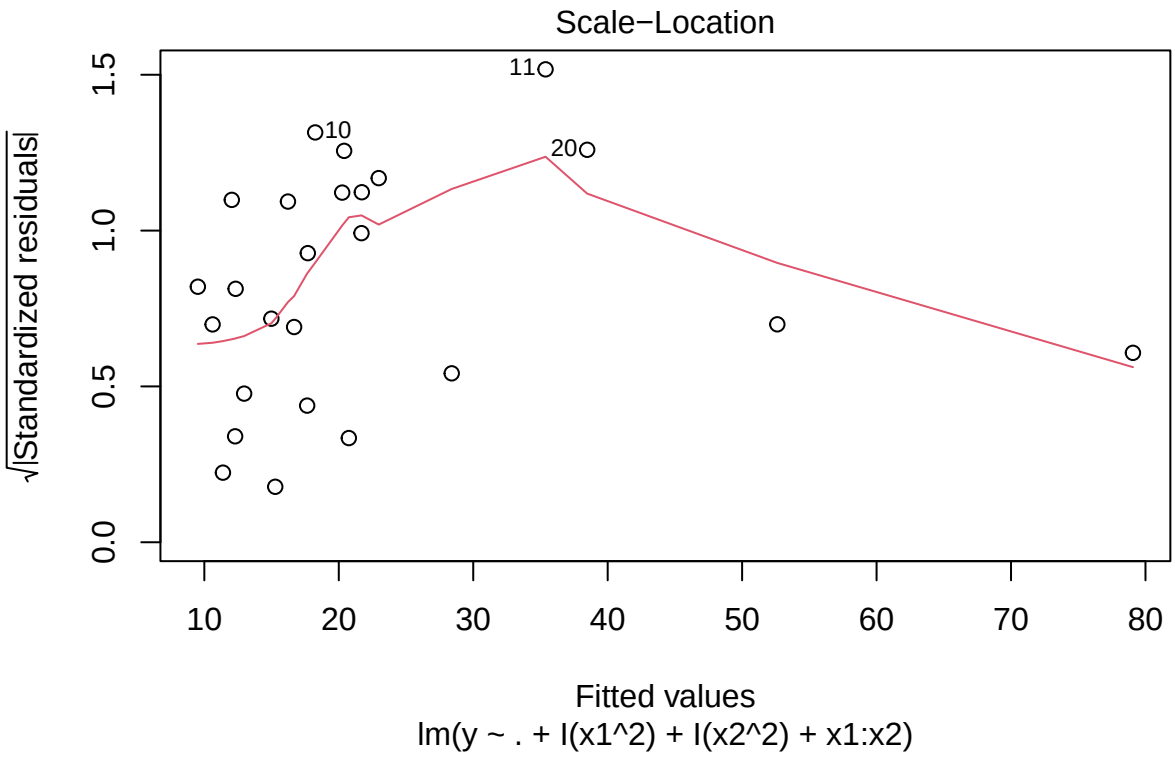
From the first figure, there is a pattern implying nonlinearity.

```
plot(fit,2)
```



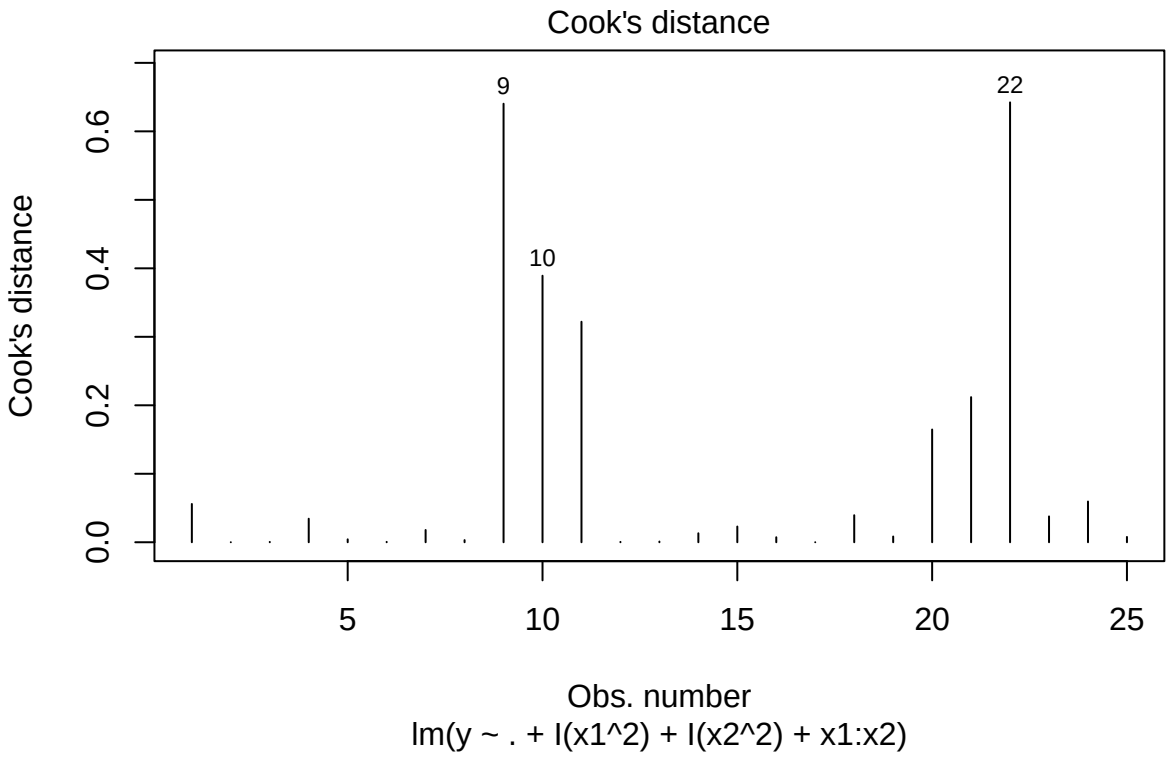
From the second figure, it is light-tailed distribution.

```
plot(fit,3)
```



From the third figure, data No.9 looks like an outlier.

```
plot(fit,4)
```

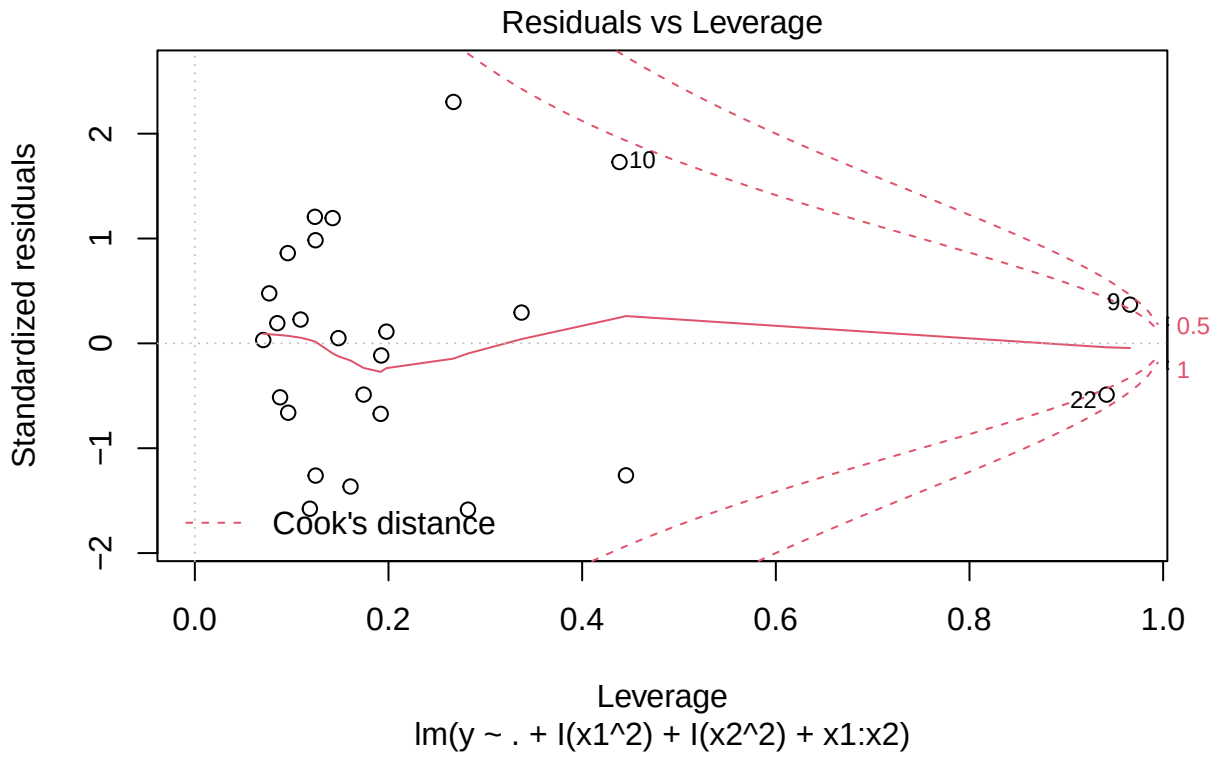


From the fourth figure, data No.9 is the only point whose Cook's distance is greater than 1.

```
plot(fit,5)
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```

```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



From the fifth figure, we can have the same conclusion as that from figure 4.

To be concluded, data No.9 is influential.

Part c.

As what we conclude in Part b, we remove data No.9:

```
data <- data[-9,]
```

Part d.

```
fit1 <- lm(y~.,data)
summary(fit1)
```

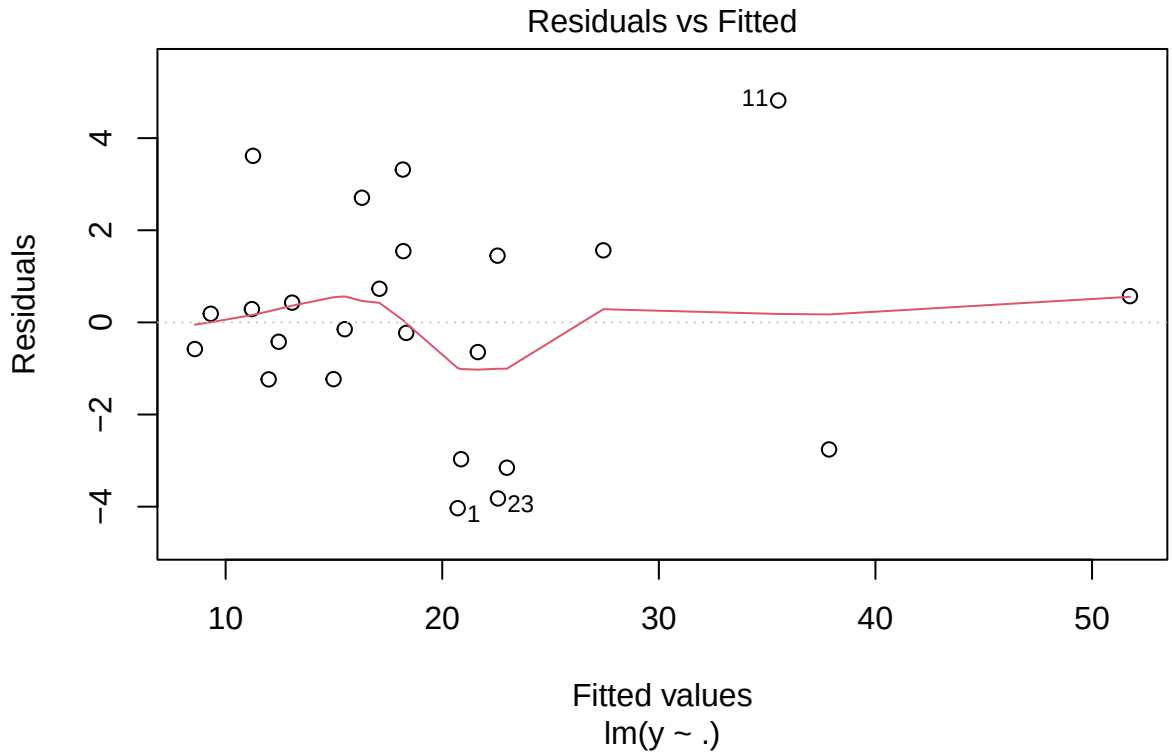
```
##
## Call:
## lm(formula = y ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.0325 -1.2331  0.0199  1.4730  4.8167
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.447238   0.952469   4.669 0.000131 ***
```

```
## x1          1.497691   0.130207  11.502 1.58e-10 ***
## x2          0.010324   0.002854   3.618 0.001614 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.43 on 21 degrees of freedom
## Multiple R-squared:  0.9487, Adjusted R-squared:  0.9438
## F-statistic: 194.2 on 2 and 21 DF,  p-value: 2.859e-14
```

So $\hat{y} = 4.447238 + 1.497691x_1 + 0.010324x_2$. The R^2 is 0.9487 and the adjusted one is 0.9438.

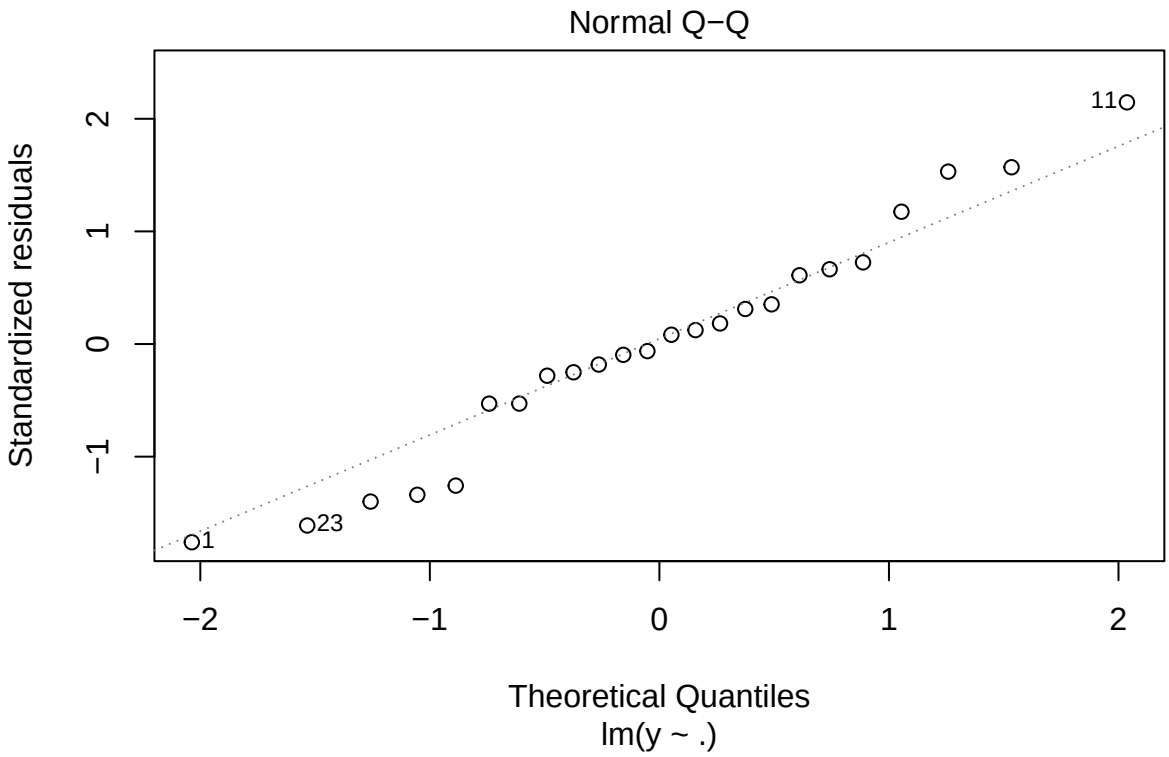
Part e.

```
plot(fit1,1)
```



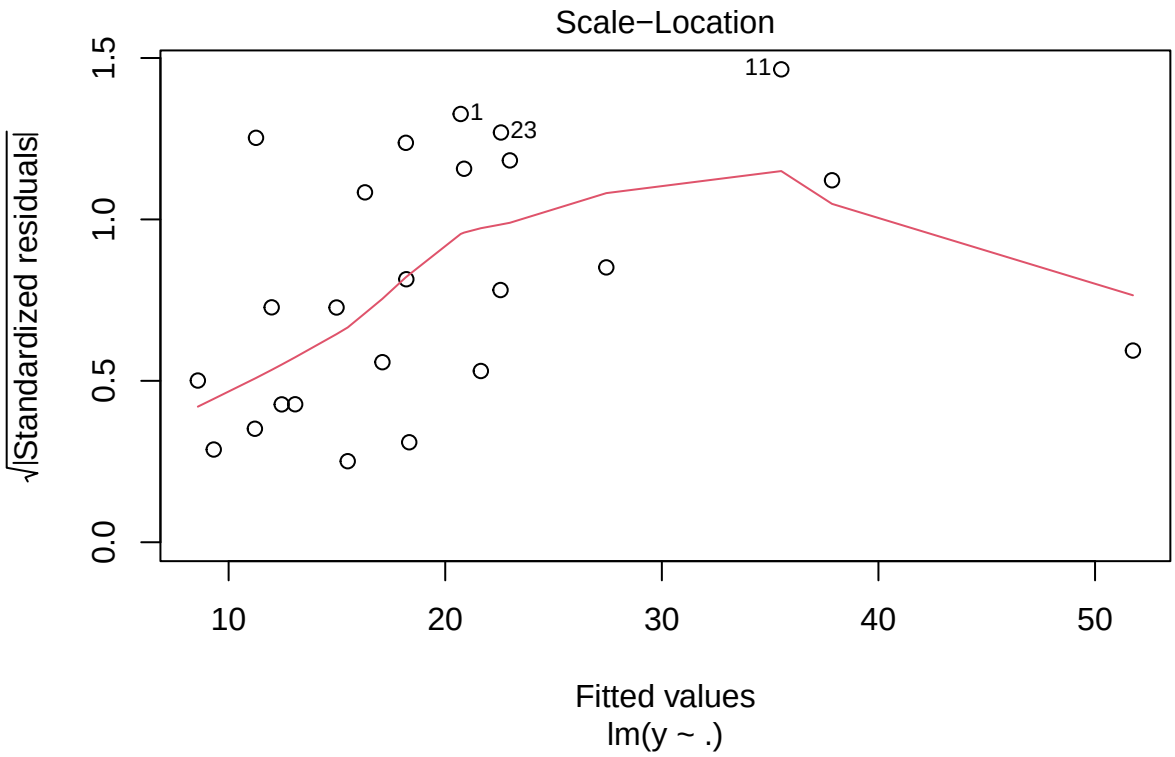
From the first figure, there is no pattern.

```
plot(fit1,2)
```



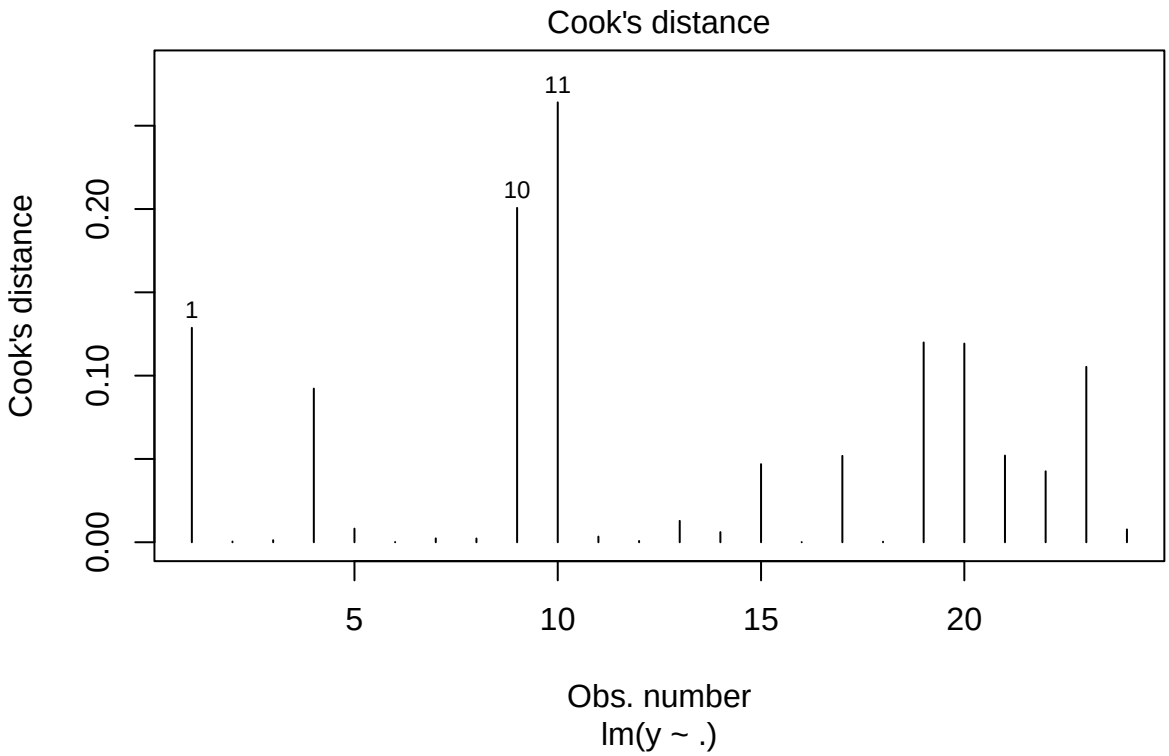
From the second figure, it is straight.

```
plot(fit1,3)
```



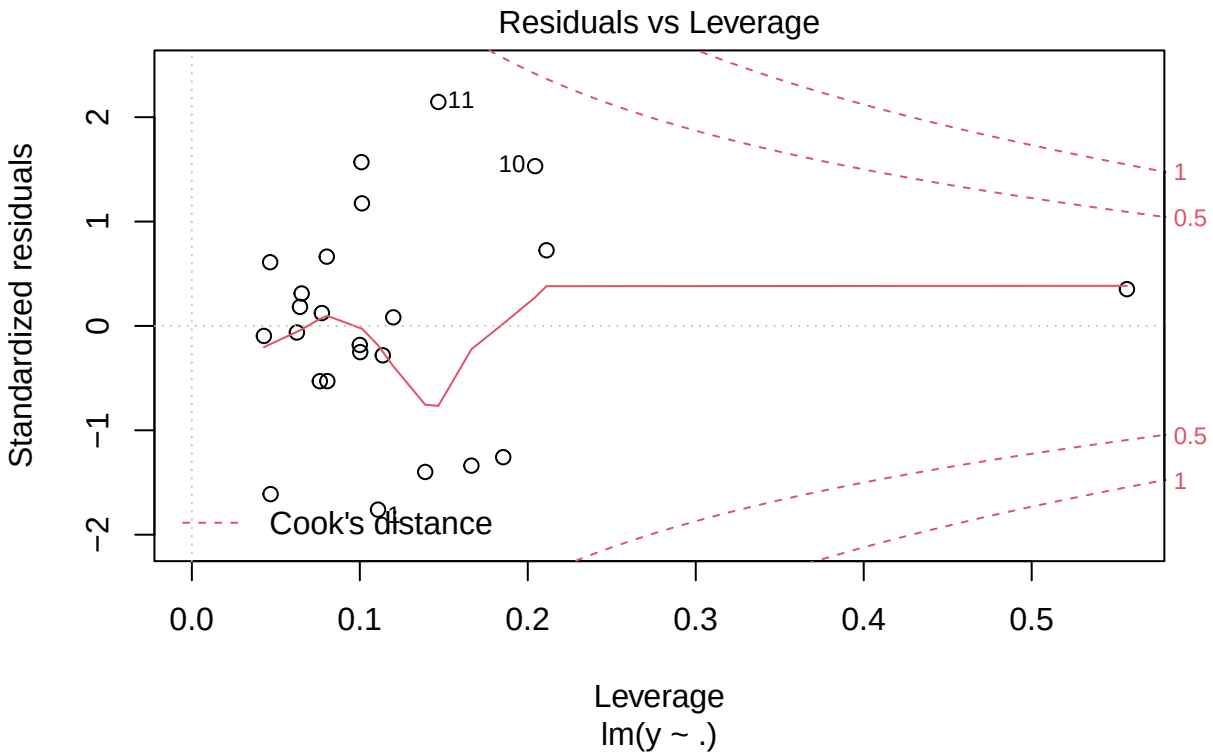
From the third figure, we don't think there is any outliers.

```
plot(fit1,4)
```



From the fourth figure, there is no point whose Cook's distance is greater than 1.

```
plot(fit1,5)
```



From the fifth figure, we can see one point that has high leverage and no points whose Cook's distance is greater than 1.

To be concluded, this model is adequate.

Part f.

```
H <- hatvalues(fit1)
H[which.max(H)]
```

```
##        22
## 0.5567143
```

Point No.22 has high leverage.

Part g.

That point No.22 is not influential because it's not an outlier. So, we don't remove it.