Tugas Individu Analisis Regresi

Ardelia Rahma Febyan_G1401221029

2024-03-06

Inisialisasi Library

```
library (readxl)
library (tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## Warning: package 'lubridate' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library (ggridges)
```

```
## Warning: package 'ggridges' was built under R version 4.3.2
```

```
library (GGally)
```

```
## Warning: package 'GGally' was built under R version 4.3.2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library (plotly)
```

```
## Warning: package 'plotly' was built under R version 4.3.2
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library (dplyr)
library (lmtest)
```

```
## Warning: package 'lmtest' was built under R version 4.3.3
```

```
## Loading required package: zoo
```

```
## Warning: package 'zoo' was built under R version 4.3.2
```

```
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library (stats)
```

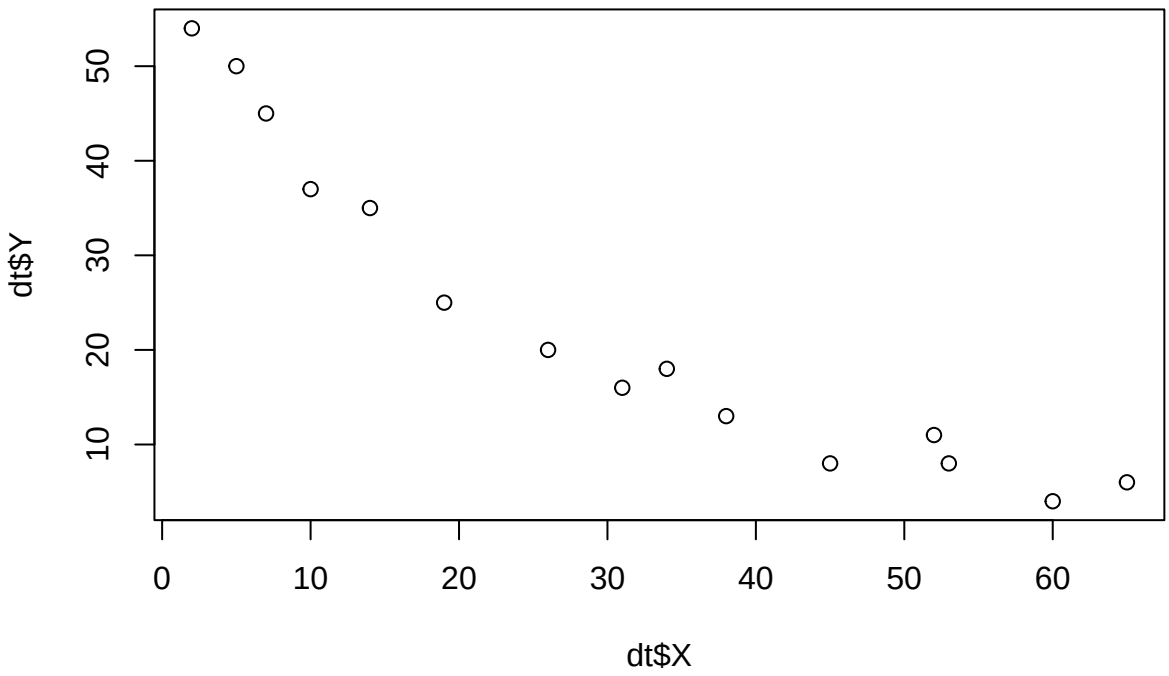
Input Data

```
dt<-read_xlsx("C:/Users/user/Downloads/Tugas Individu Anreg/Tugas Individu Analisis Regresi.xlsx")
str(dt)
```

```
## tibble [15 x 3] (S3: tbl_df/tbl/data.frame)
##  $ NO: num [1:15] 1 2 3 4 5 6 7 8 9 10 ...
##  $ X : num [1:15] 2 5 7 10 14 19 26 31 34 38 ...
##  $ Y : num [1:15] 54 50 45 37 35 25 20 16 18 13 ...
```

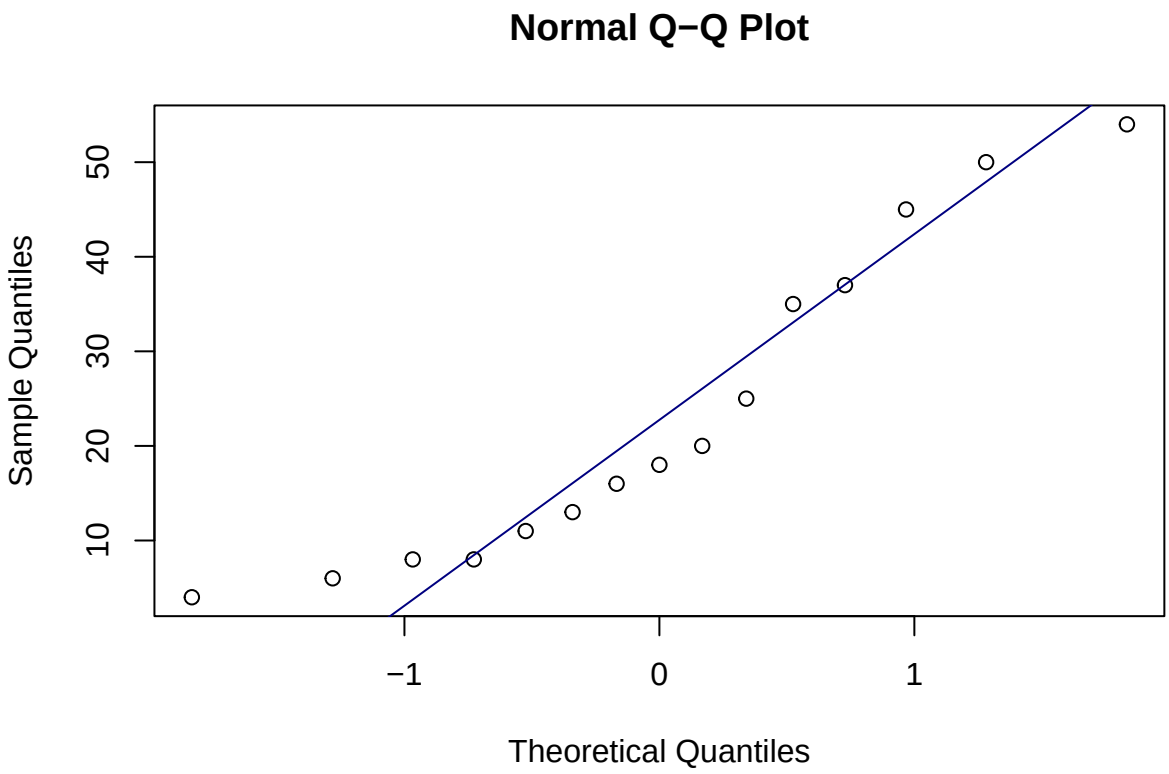
Scatter Plot

```
plot(x=dt$X, y=dt$Y)
```



Uji Normalitas

```
qqnorm(dt$Y)
qqline(dt$Y, col="navy")
```




```
shapiro.test(dt$Y)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  dt$Y
## W = 0.89636, p-value = 0.08374
```

Intepretasi

Hasil menunjukkan p-value = 0.08374. Karena p-value lebih dari 0.05, maka dapat disimpulkan bahwa data menyebar normal walaupun menurut qqplot data tidak menyebar normal.

Pemodelan Regresi Linear

```
model<-lm(formula=Y~X, data=dt)
summary(model)
```

```
##
## Call:
## lm(formula = Y ~ X, data = dt)
##
```

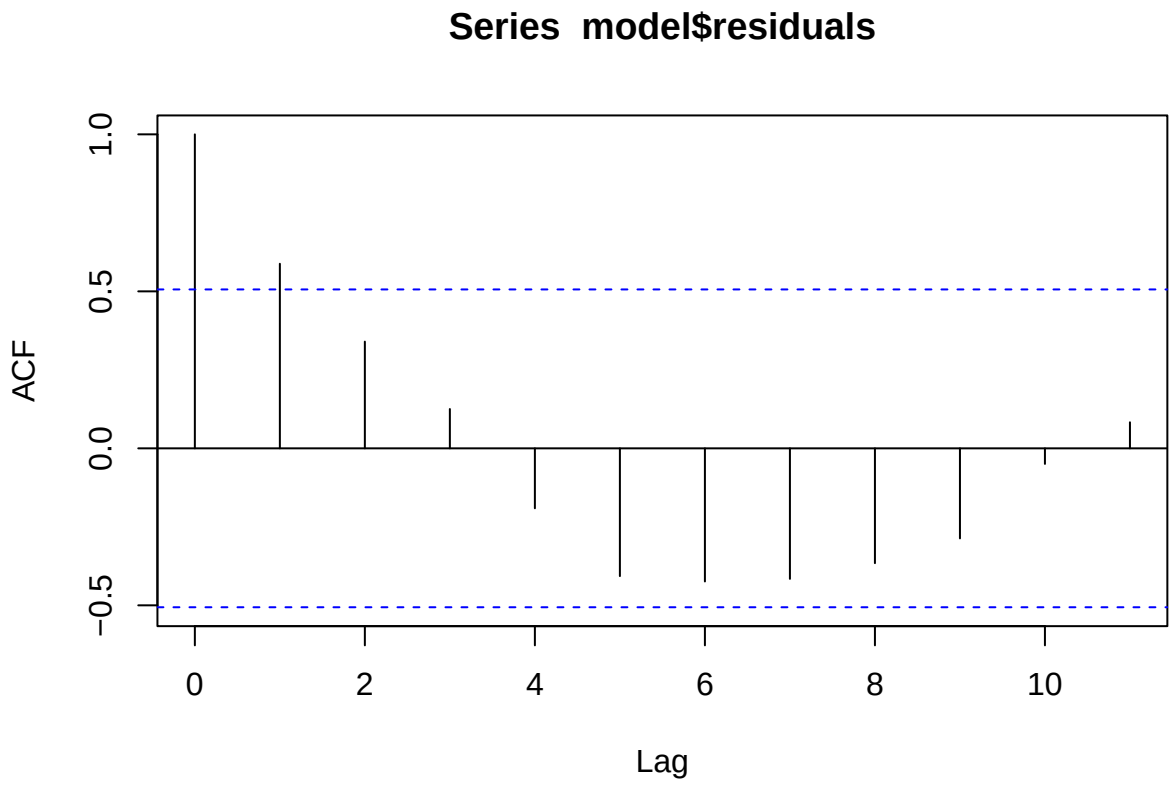
```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1628 -4.7313 -0.9253  3.7386  9.0446
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46.46041    2.76218   16.82 3.33e-10 ***
## X           -0.75251    0.07502  -10.03 1.74e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.891 on 13 degrees of freedom
## Multiple R-squared:  0.8856, Adjusted R-squared:  0.8768
## F-statistic: 100.6 on 1 and 13 DF,  p-value: 1.736e-07
```

```
model
```

```
##
## Call:
## lm(formula = Y ~ X, data = dt)
##
## Coefficients:
## (Intercept)            X
##     46.4604      -0.7525
```

Uji Autokorelasi

```
acf(model$residuals)
```




```
dwtest(model)
```

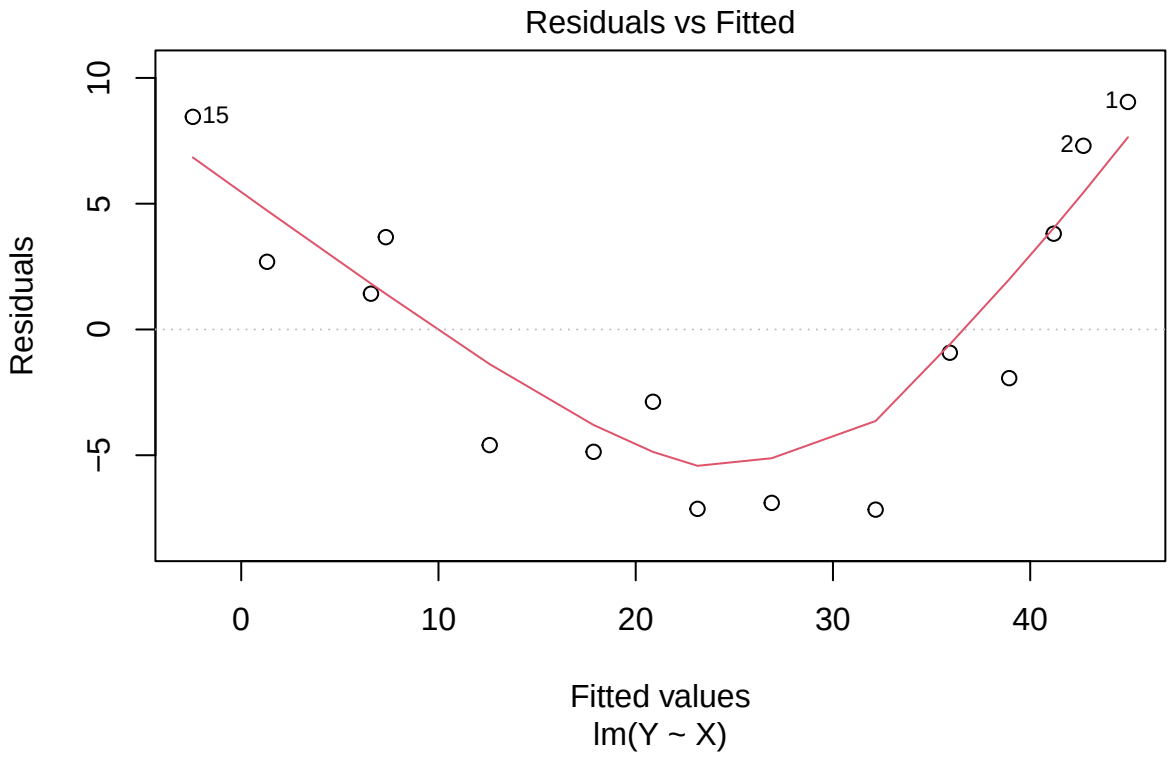
```
##
##  Durbin-Watson test
##
## data:  model
## DW = 0.48462, p-value = 1.333e-05
## alternative hypothesis: true autocorrelation is greater than 0
```

Intepretasi

Hasil menunjukkan bahwa autokorelasi pada lag 1 sekitar 0.6 dam lag 2 sekitar 0.4. Uji Durbin-Watson juga menunjukkan bahwa p-value = 1.333e-05 yang kurang dari 0.05. Nilai autokorelasi yang melebihi selang kepecrayaan menunjukkan bahwa dapat disimpulkan terdapat autokorelasi pada lag 1 dan 2 secara signifikan. Karena terdapat autokorelasi, maka tidak terpenuhinya asumsi Gauss-Makrov.

Uji Ragam Galat Homogen (Homoskedastisitas)

```
plot(model, which = 1)
```



Intepretasi

Dapat dilihat bahwa ragam galat cenderung meningkat seiring dengan nilai prediksi. Hal ini menunjukkan bahwa terdapat homoskedastisitas

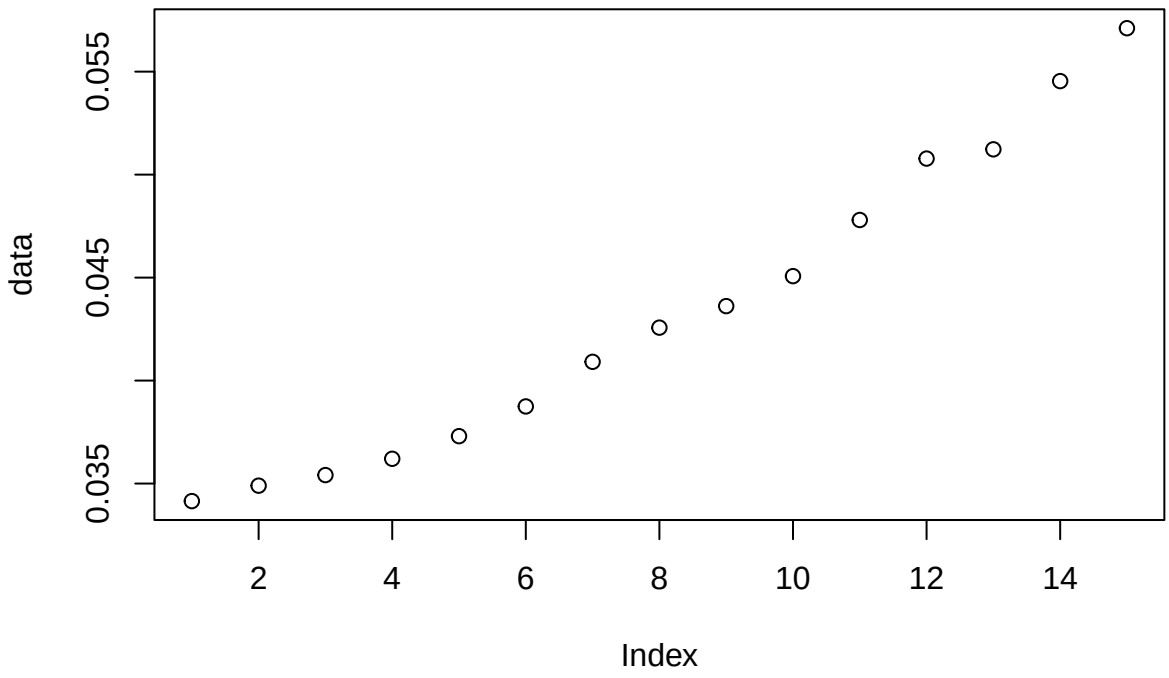
Transformasi Data

```
residual<-abs(model$residuals)
fitted<-model$fitted.values
fit<-lm(residual~fitted, dt)
data<-(1/(fit$fitted)^2)
data
```

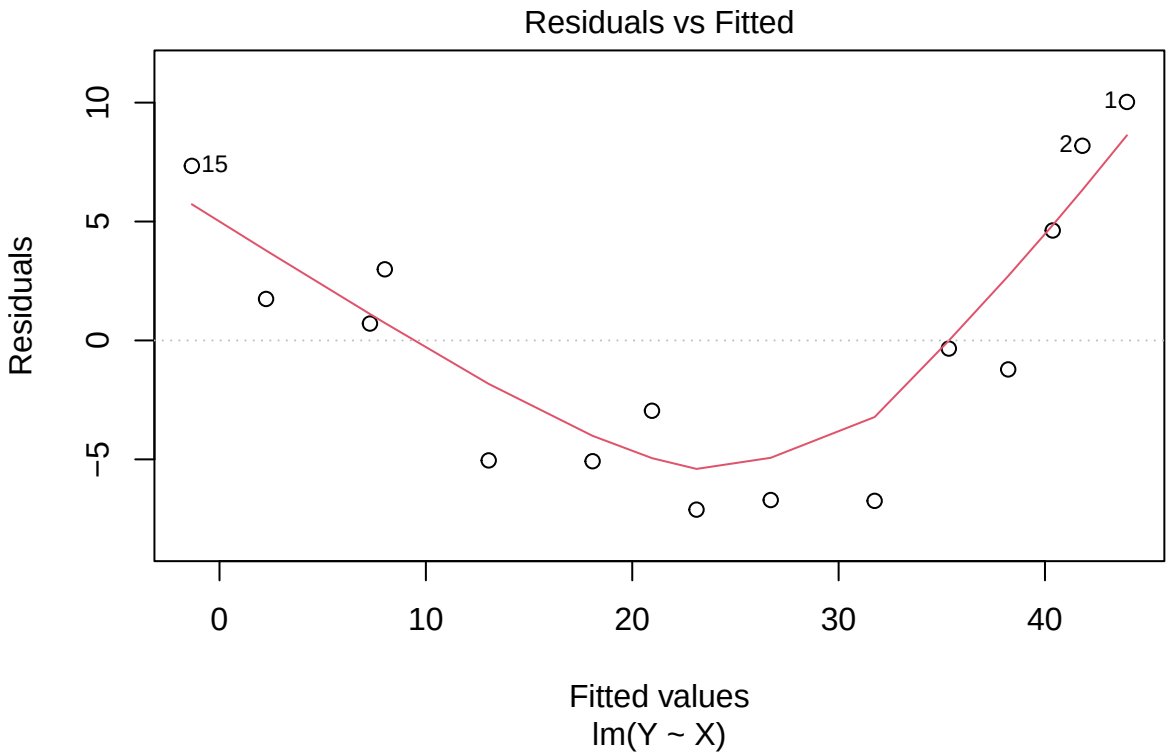
```
##          1          2          3          4          5          6          7
## 0.03414849 0.03489798 0.03541143 0.03620311 0.03730067 0.03874425 0.04091034
##          8          9         10         11         12         13         14
## 0.04257072 0.04361593 0.04507050 0.04779711 0.05077885 0.05122749 0.05454132
##         15
## 0.05710924
```

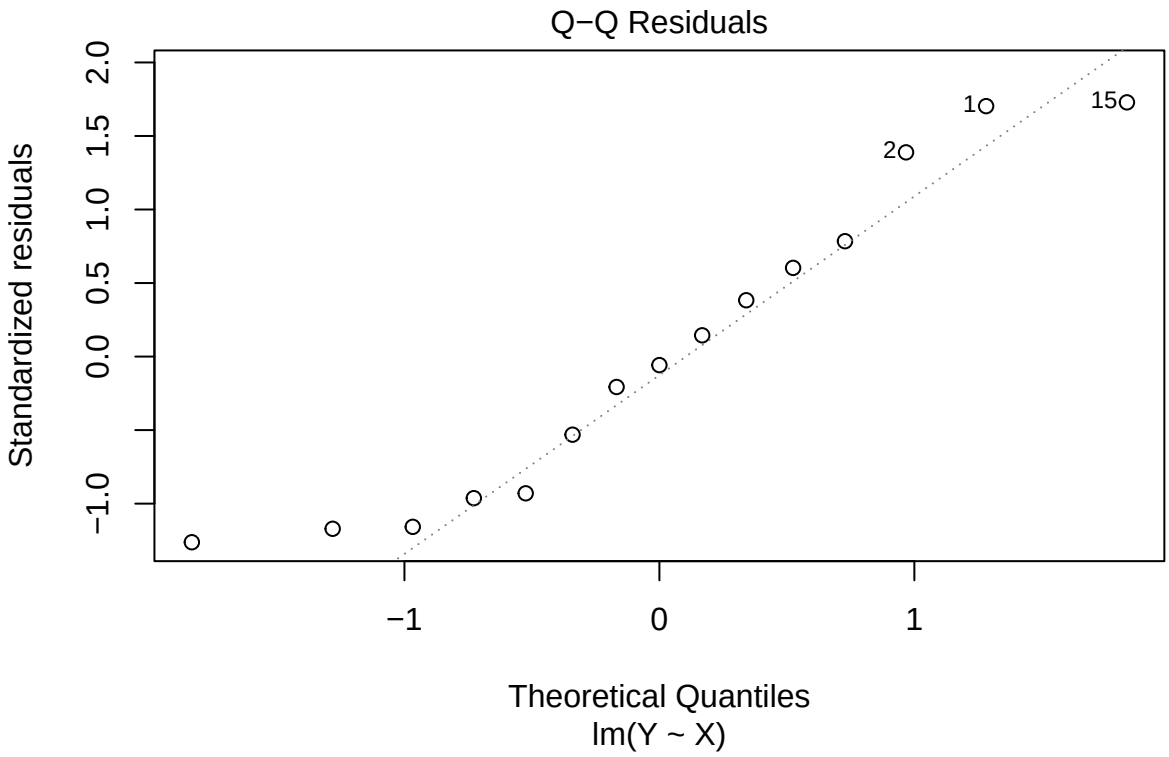
Scatter Plot

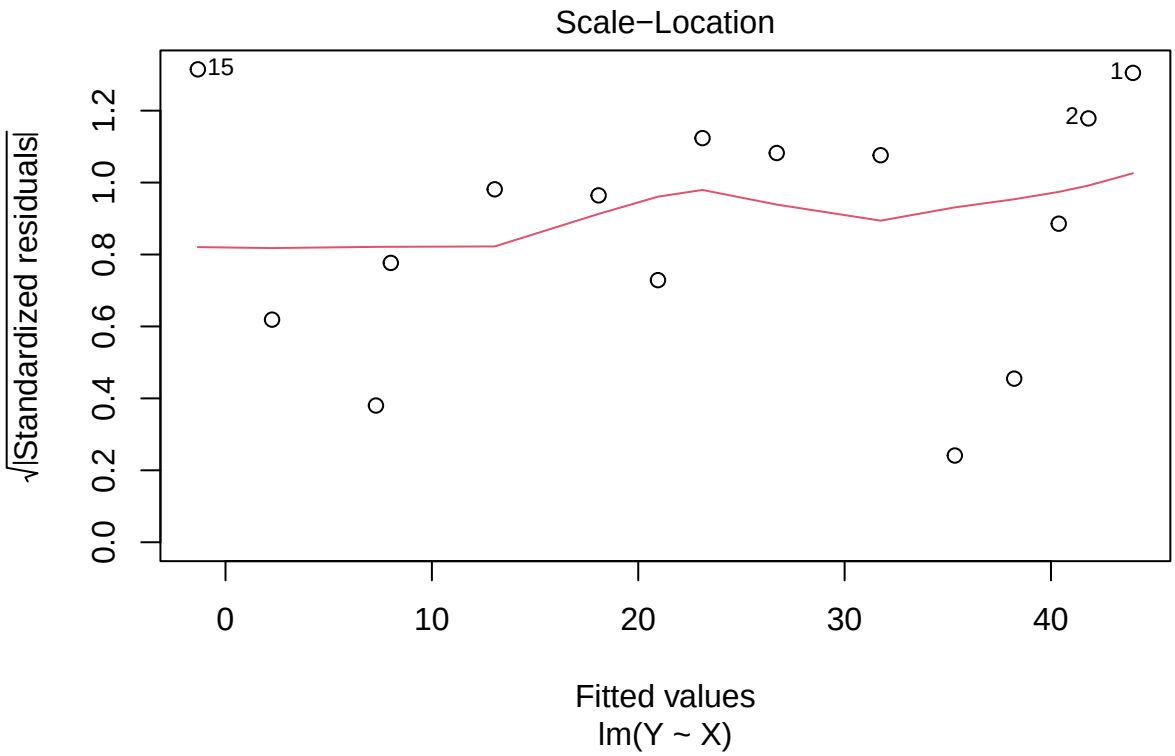
```
plot(data)
```

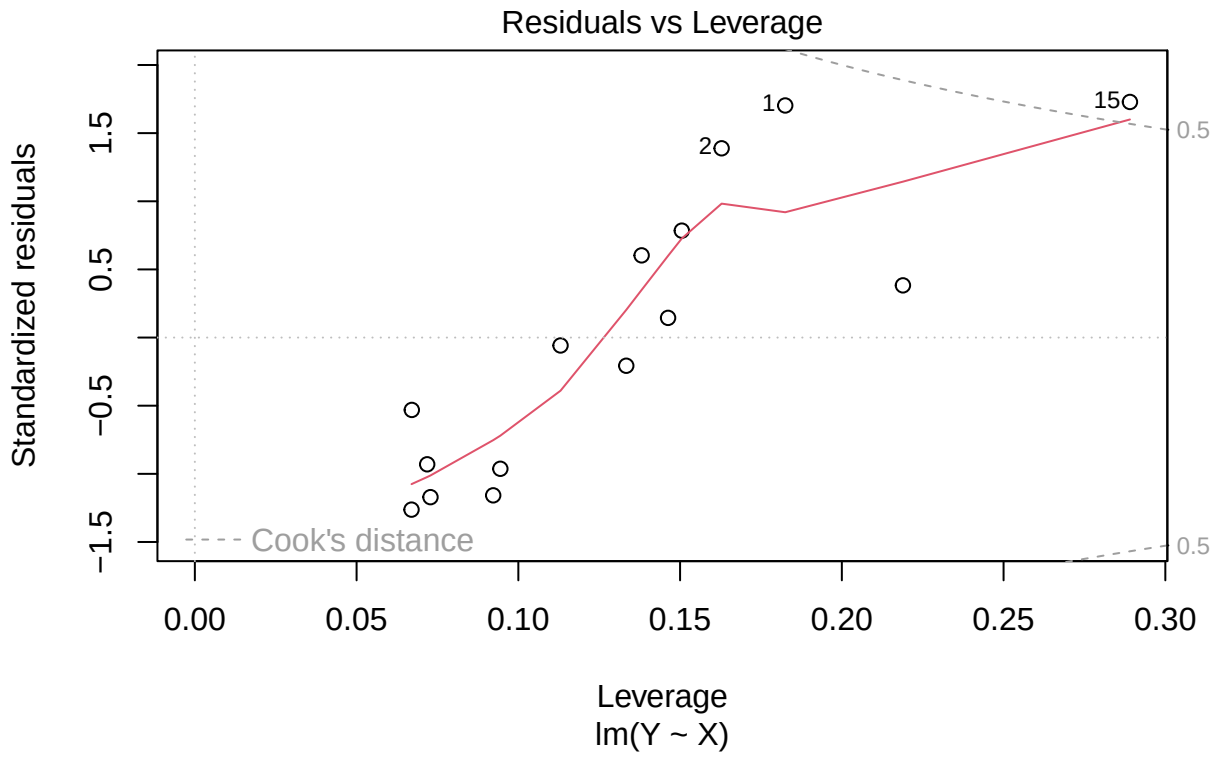



```
model_data <- lm(Y~X, data = dt, weights = data)
plot(model_data)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Y ~ X)
15
1
2
10
5
0
−5
0
10
20
30
40


Q−Q Residuals
Standardized residuals
−1.0
0.0
0.5
1.0
1.5
2.0
1
2
15
−1
0
1
Theoretical Quantiles
lm(Y ~ X)


Scale−Location
√|Standardized residuals|
0.0
0.2
0.4
0.6
0.8
1.0
1.2
15
1
2
0
10
20
30
40
Fitted values
lm(Y ~ X)




```r
summary(model_data)
```

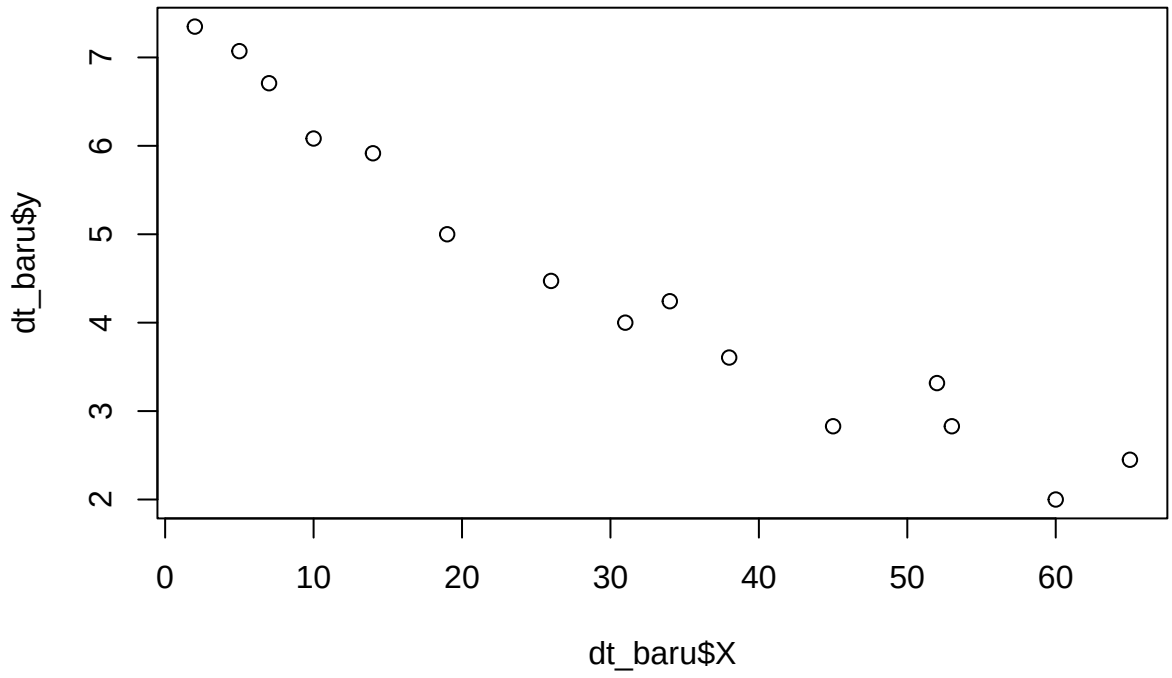
```
##
## Call:
## lm(formula = Y ~ X, data = dt, weights = data)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -1.46776 -1.09054 -0.06587  0.77203  1.85309
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.41058    2.90674  15.623 8.35e-10 ***
## X           -0.71925    0.07313  -9.835 2.18e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.204 on 13 degrees of freedom
## Multiple R-squared:  0.8815, Adjusted R-squared:  0.8724
## F-statistic: 96.73 on 1 and 13 DF,  p-value: 2.182e-07
```

Intepretasi

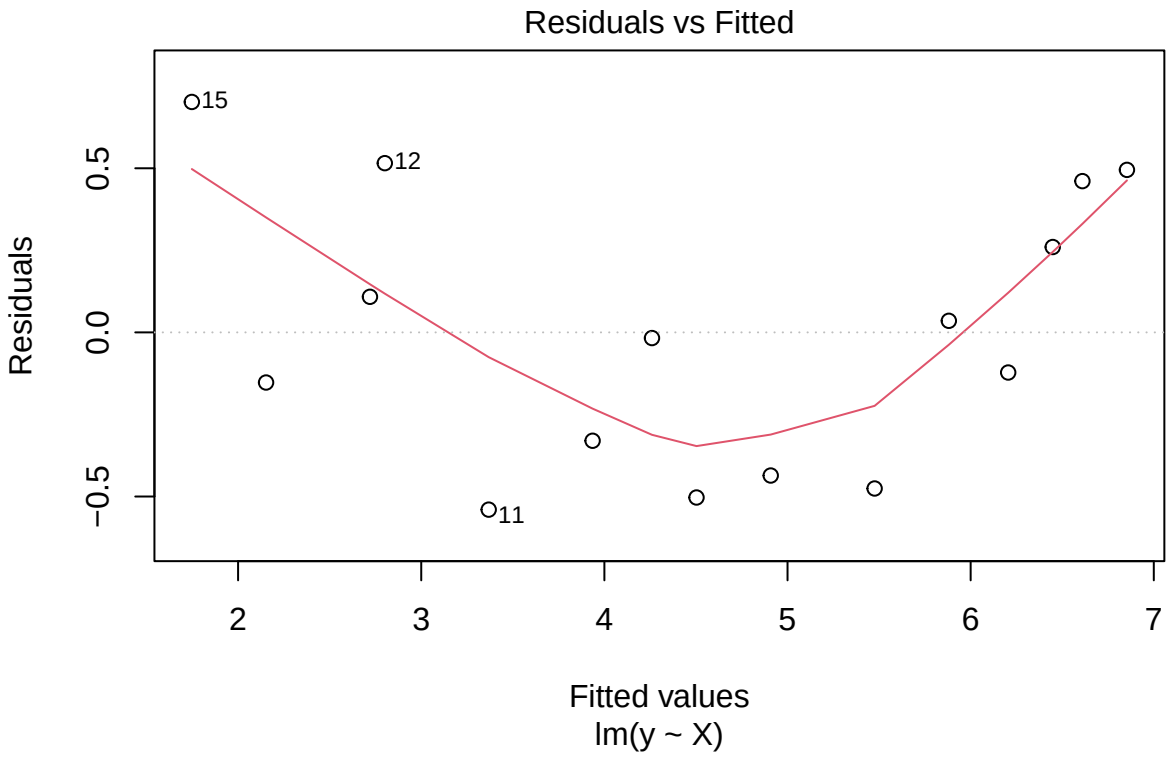
Dapat dilihat bahwa transformasi WLS masih belum efektif karena data masih belum memenuhi asumsi Gauss-Markov.

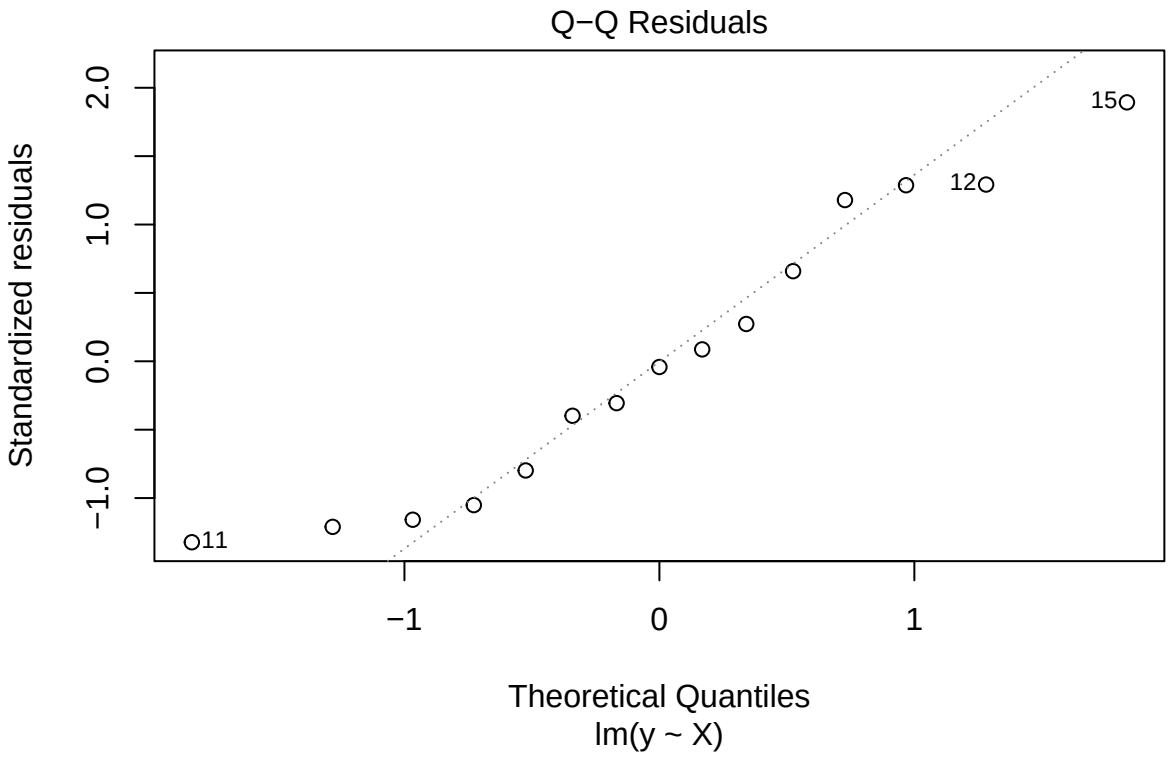
Transformasi Akar pada X dan Y

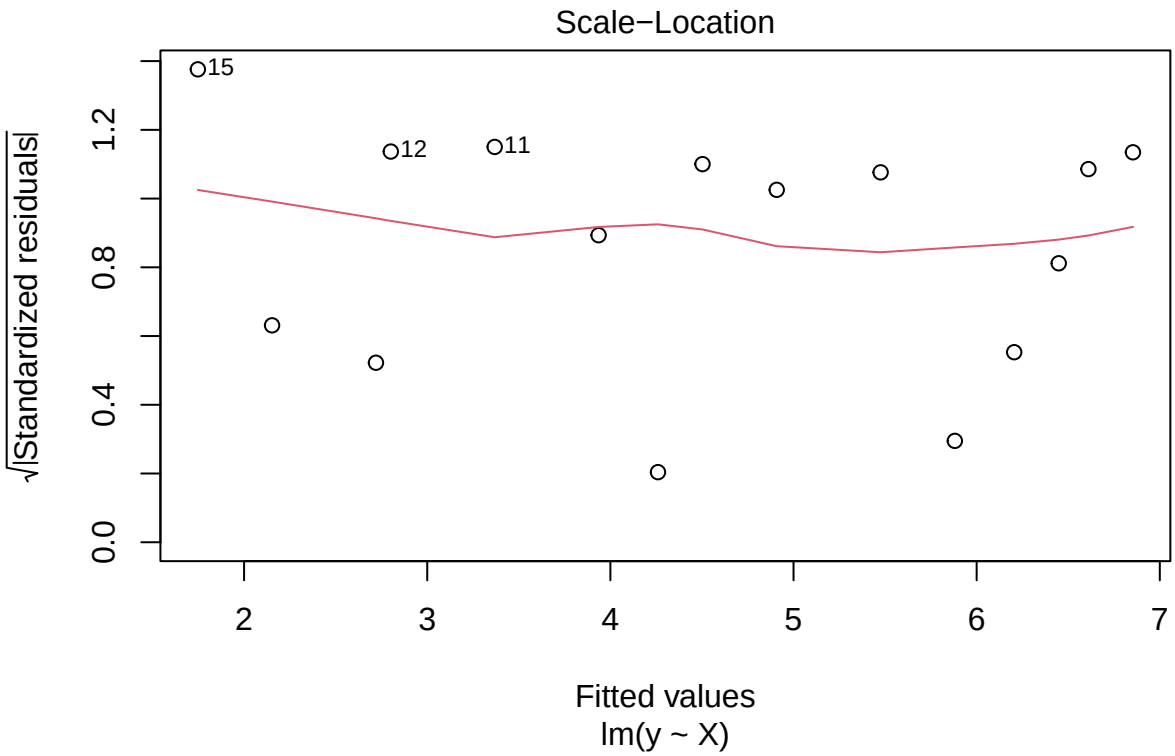
```
dt_baru<-dt %>%
  mutate(y=sqrt(Y)) %>%
  mutate(x=sqrt(X))
model_sqrt_x <- lm(y~X, data=dt_baru)
plot(x=dt_baru$X, y=dt_baru$y)
```

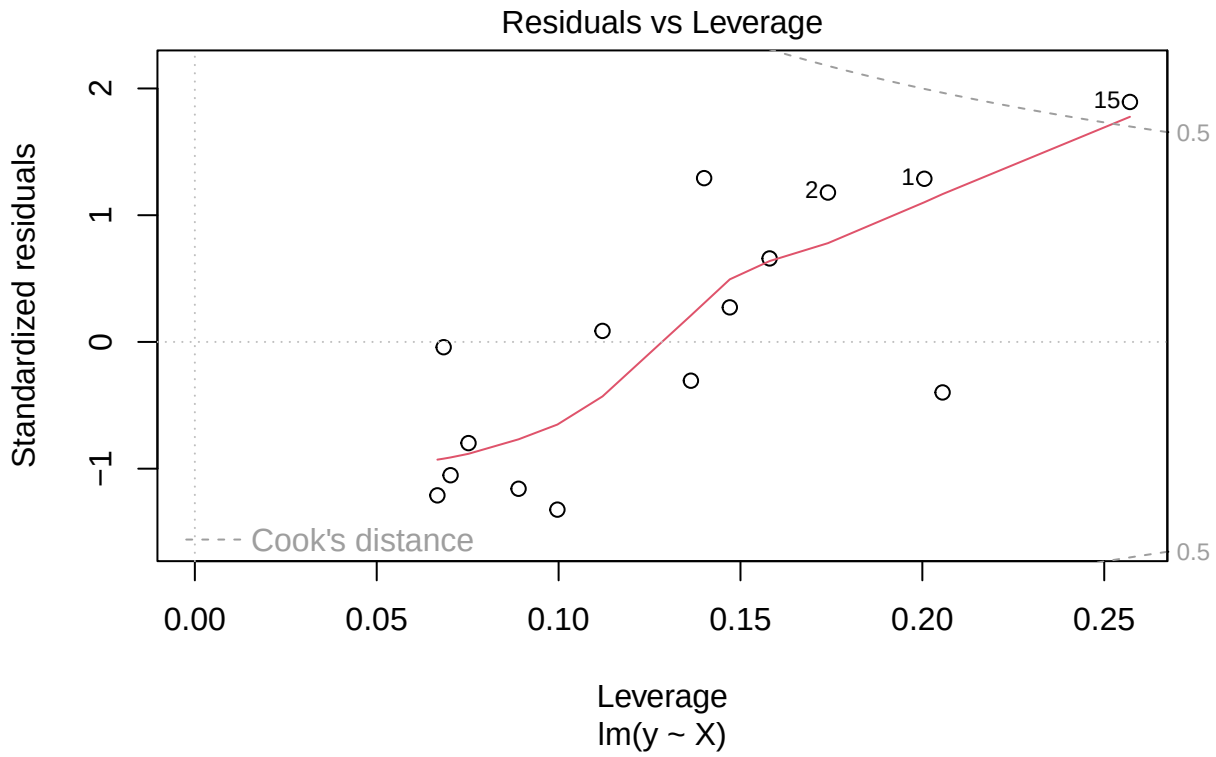



```
plot(model_sqrt_x)
```


Residuals vs Fitted
Residuals
0.5
0.0
−0.5
15
12
11
2
3
4
5
6
7
Fitted values
lm(y ~ X)


Q−Q Residuals
Standardized residuals
2.0
1.0
0.0
−1.0
15
12
11
−1
0
1
Theoretical Quantiles
lm(y ~ X)


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
15
12
11
2
3
4
5
6
7
Fitted values
lm(y ~ X)




```
summary(model_sqrt_x)
```

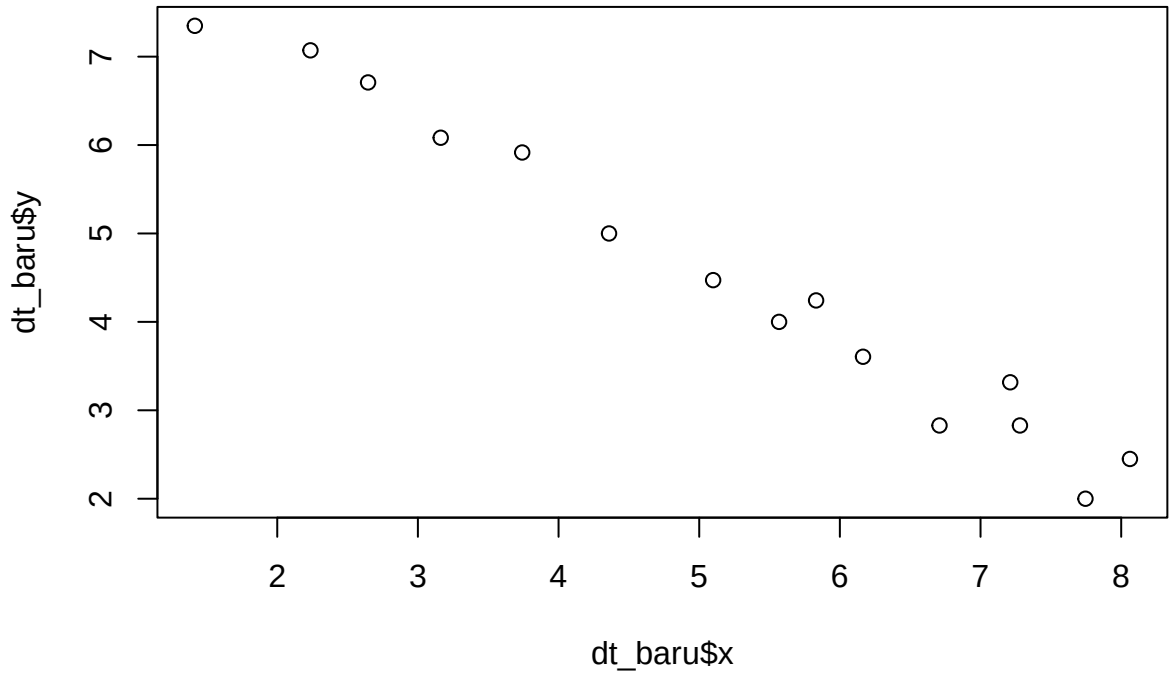
```
##
## Call:
## lm(formula = y ~ X, data = dt_baru)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53998 -0.38316 -0.01727  0.36045  0.70199
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.015455   0.201677   34.79 3.24e-14 ***
## X           -0.081045   0.005477  -14.80 1.63e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4301 on 13 degrees of freedom
## Multiple R-squared:  0.9439, Adjusted R-squared:  0.9396
## F-statistic: 218.9 on 1 and 13 DF,  p-value: 1.634e-09
```

Uji Korelasi terhadap Model Regresi Transformasi

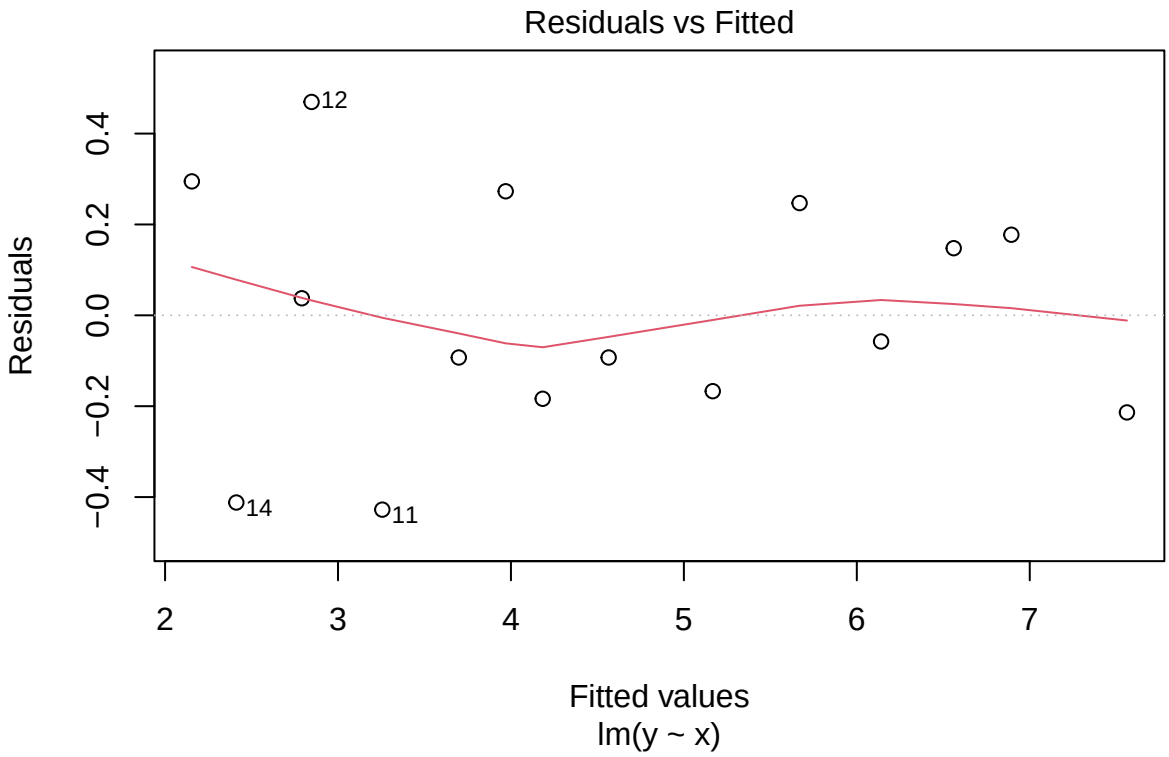
```
dwtest(model_sqrt_x)
```

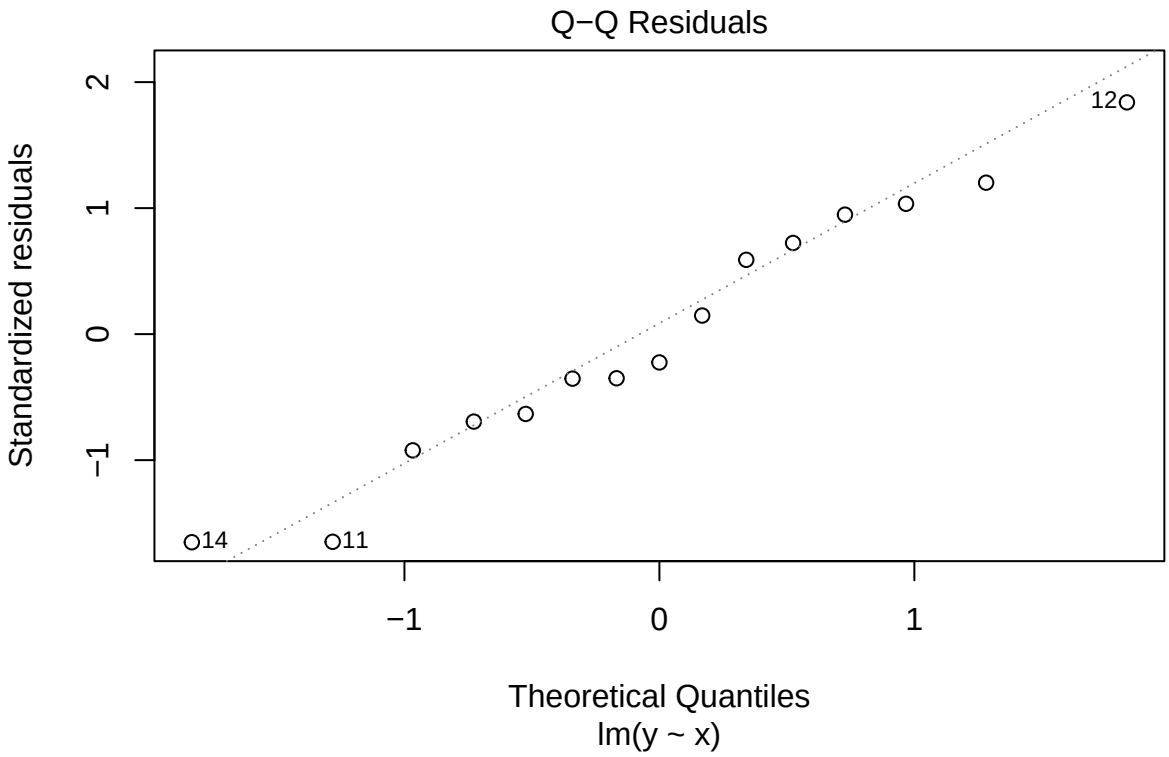
```
##
##  Durbin-Watson test
##
## data:  model_sqrt_x
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0
```

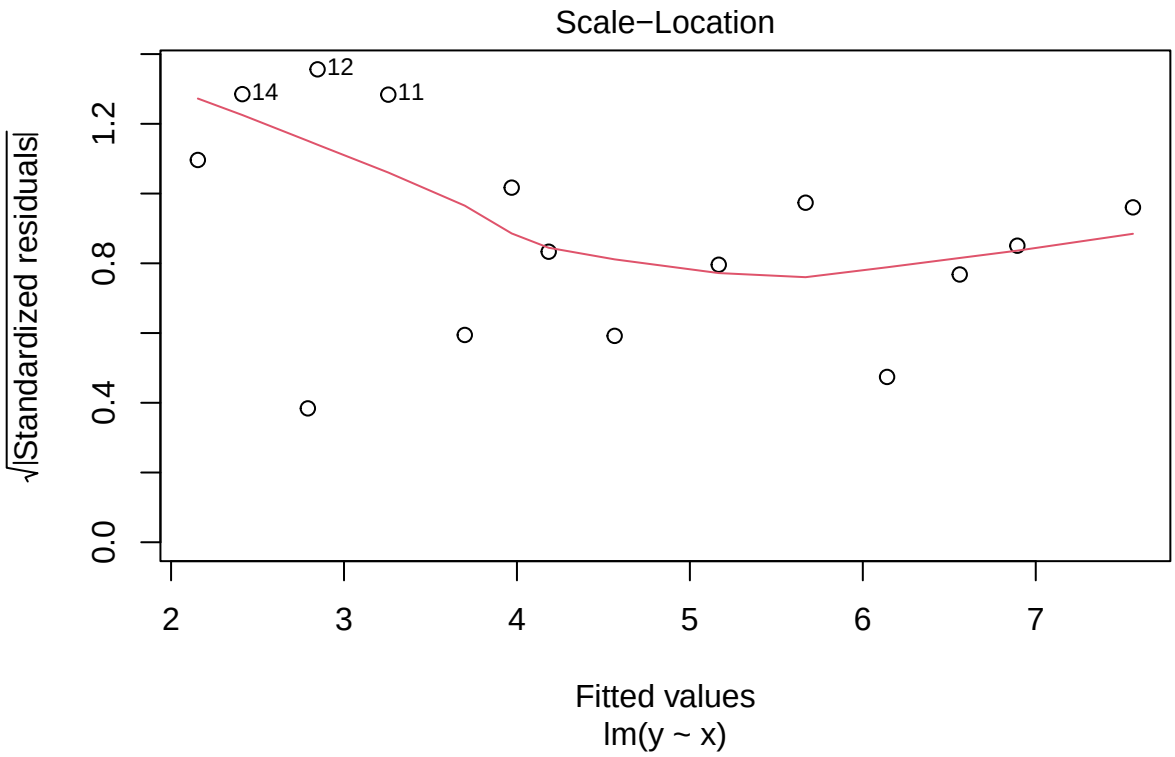
```
model_sqrt <- lm(y ~ x, data = dt_baru)
plot(x = dt_baru$x, y = dt_baru$y)
```

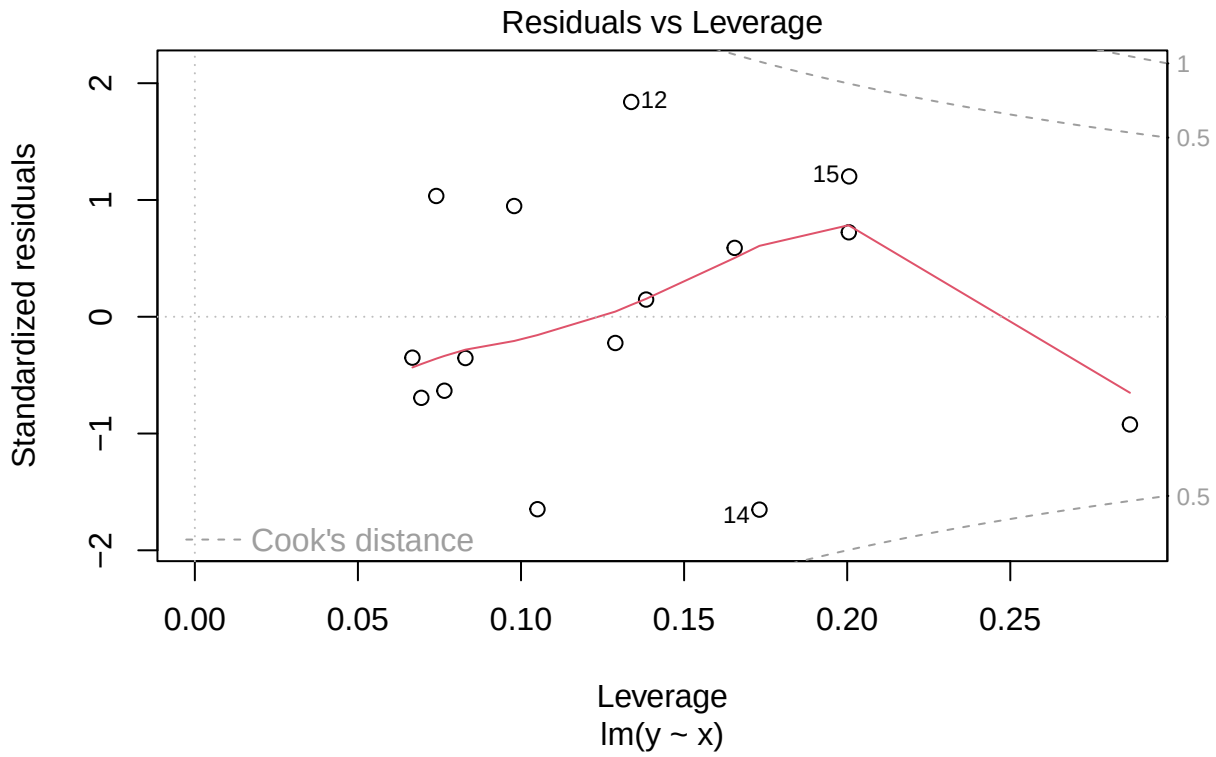



```
plot(model_sqrt)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
12
14
11
2
3
4
5
6
7
Fitted values
lm(y ~ x)


Q−Q Residuals
Standardized residuals
2
1
0
−1
12
14
11
−1
0
1
Theoretical Quantiles
lm(y ~ x)


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
14
12
11
2
3
4
5
6
7
Fitted values
lm(y ~ x)




```
summary(model_sqrt)
```

```
##
## Call:
## lm(formula = y ~ x, data = dt_baru)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42765 -0.17534 -0.05753  0.21223  0.46960
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.71245    0.19101   45.61 9.83e-16 ***
## x           -0.81339    0.03445  -23.61 4.64e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2743 on 13 degrees of freedom
## Multiple R-squared:  0.9772, Adjusted R-squared:  0.9755
## F-statistic: 557.3 on 1 and 13 DF,  p-value: 4.643e-12
```

Intepretasi

Hasil test Durbin-Watson menunjukkan adanya autokorelasi positif karena nilai DW test yang rendah dan p-value yang kurang dari 0.05 sehingga signifikan.

Uji Autokorelasi Model Regresi

```
dwtest(model_sqrt)
```

```
##
##  Durbin-Watson test
##
## data:  model_sqrt
## DW = 2.6803, p-value = 0.8629
## alternative hypothesis: true autocorrelation is greater than 0
```

Kesimpulan

Dapat disimpulkan bahwa nilai p-value yang besar dari 0.05 meunjukkan bahwa belum cukup bukti untuk tolak H0, maka tidak terdapat autokorelasi. Dari hasil transformasi diatas juga dapat disimpulkan bahwa transformasi akar Y dapat membuat pemodelan regresi linear lebih efektif. Model Regresi Linear setelah ditransformasi, yaitu:

$$Y^* = 8.71245 - 0.81339X^* + eY^* = \sqrt{Y}X^* = \sqrt{X}$$

Interpretasi model menunjukkan bahwa Y berkorelasi terbalik dengan akar kuadrat dari X, dengan hubungan yang bersifat kuadratik. Semakin besar nilai akar kuadrat dari X, semakin kecil rata-rata nilai Y, dengan tingkat penurunan yang semakin meningkat. Puncak kurva menunjukkan nilai rata-rata maksimum Y untuk nilai tertentu dari X. Konstanta 8.71245 mewakili nilai Y ketika X sama dengan 0. Koefisien -0.81339 merupakan koefisien regresi untuk variabel X. Nilai negatif menunjukkan hubungan terbalik antara Y dan akar kuadrat dari X. Dengan kata lain, semakin besar akar kuadrat dari X, semakin kecil nilai Y. Pangkat dua pada koefisien regresi menunjukkan bahwa hubungan antara Y dan X bersifat kuadratik. Ini berarti perubahan Y tidak proporsional dengan perubahan X, melainkan berubah dengan tingkat peningkatan yang semakin tinggi.