Tugas Individu Anreg K1

Bulan Cahyani Suhaeri G1401221030

2024-03-06

```
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 4.3.3
```

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

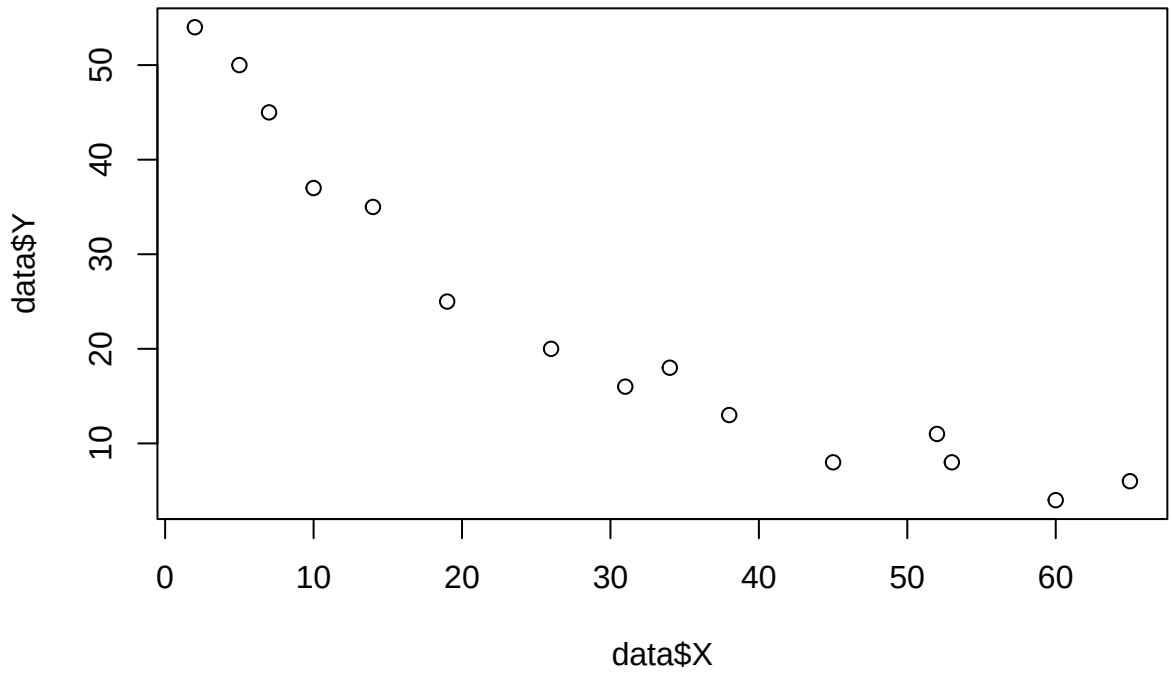
MEMBACA DATA

```
library(readxl)
data <- read_xlsx("C:/Users/bulbu/Documents/DataIndividuAnreg.xlsx")
data
```

```
## # A tibble: 15 x 2
##        X     Y
##    <dbl> <dbl>
##  1     2    54
##  2     5    50
##  3     7    45
##  4    10    37
##  5    14    35
##  6    19    25
##  7    26    20
##  8    31    16
##  9    34    18
## 10    38    13
## 11    45     8
## 12    52    11
## 13    53     8
## 14    60     4
## 15    65     6
```

#EKSPLORASI DATA

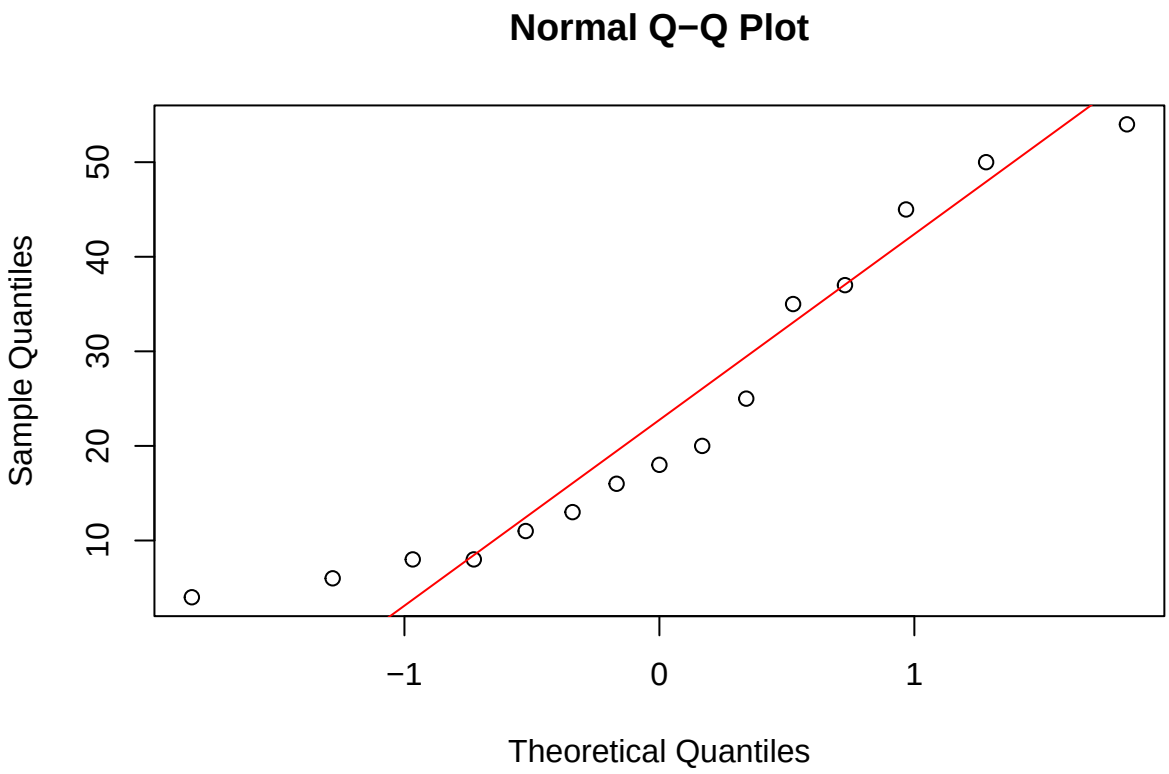
```
plot(x = data$X, y = data$Y)
```



Dari hasil Scatter Plot di atas menunjukan jika Y dan X tidak berhubungan linear. Jika tidak memenuhi asumsi-asumsi yang ada, maka data harus ditransformasi. Transformasi yang diperlukan jika scatter plotnya berbentuk seperti di atas, maka trasformasi yang dilakukan adalah dengan memperkecil x (mengakarkan X) atau memperkecil y (mengakarkan Y).

#UJI KENORMALAN

```
qqnorm(data$Y)
qqline(data$Y, col = "red")
```




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

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

Data yang diketahui menyebar normal dibuktikan dari hasil shapiro test yang lebih dari 0.05 walaupun hasil dari qq plot cenderung memiliki asumsi bahwa data tersebut tidak menyebar normal.

#DEKLARASI MODEL REGRESI

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

```
##
## Call:
## lm(formula = Y ~ X, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1628 -4.7313 -0.9253  3.7386  9.0446
##
## Coefficients:
```

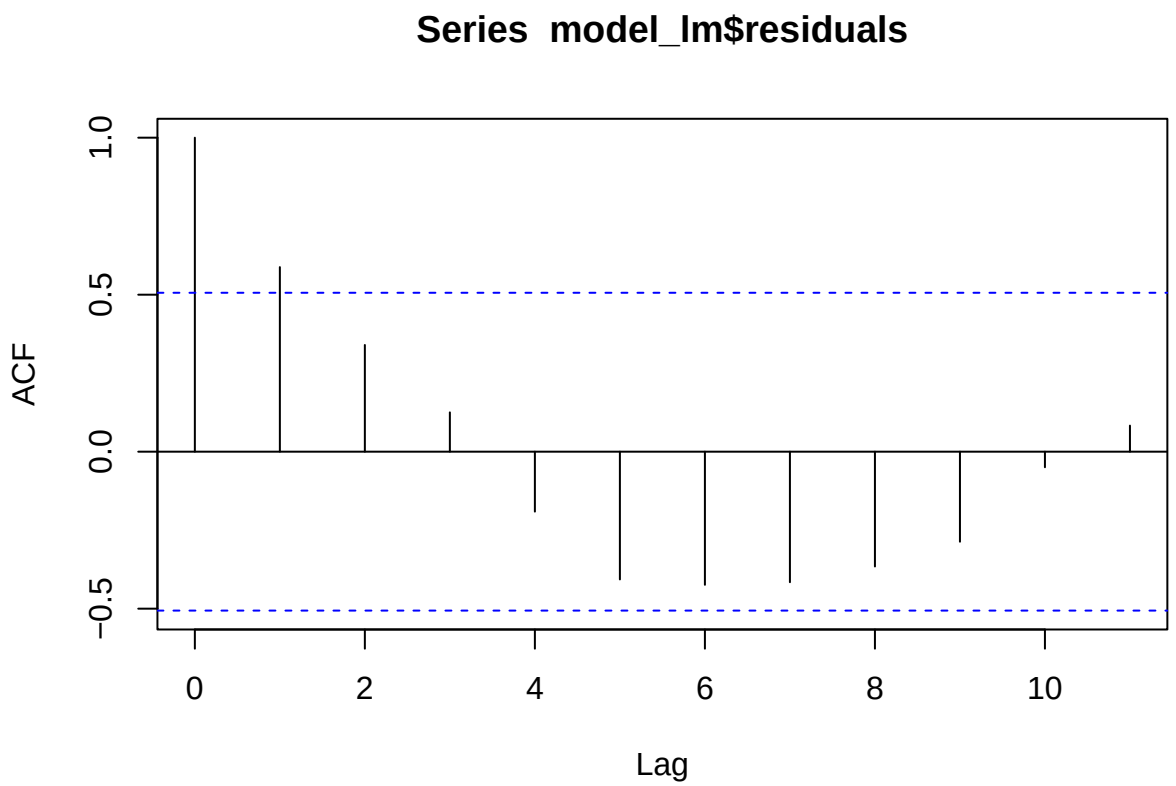
```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46.46041    2.76218   16.82 3.33e-10 ***
## X           -0.75251    0.07502  -10.03 1.74e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.891 on 13 degrees of freedom
## 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_lm
```

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

#AUTO KORELASI

```
acf(model_lm$residuals)
```



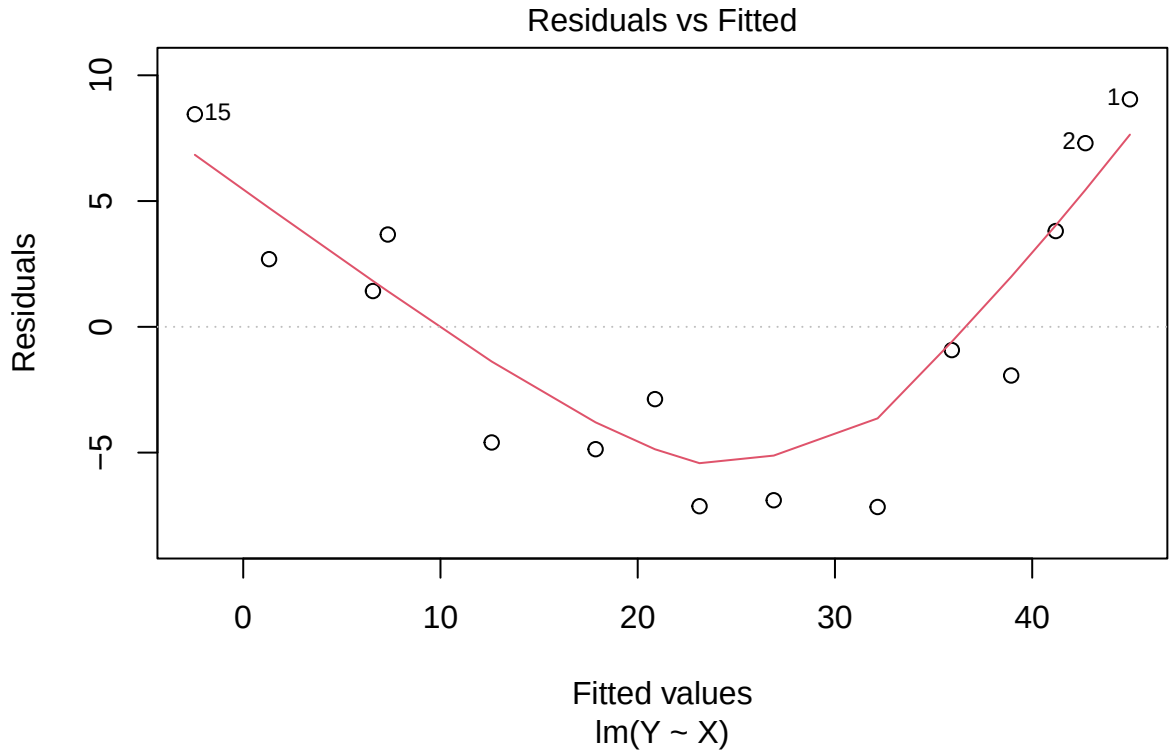
```
dwtest(model_lm)
```

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

Nilai p-value < 0.05 tolak H0, maka pada taraf nyata 95% terdapat autokorelai pada residual. Sehingga uji asumsi autokorelasi tidak terpenuhi.

#UJI HOMOSKEDASTISITAS

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



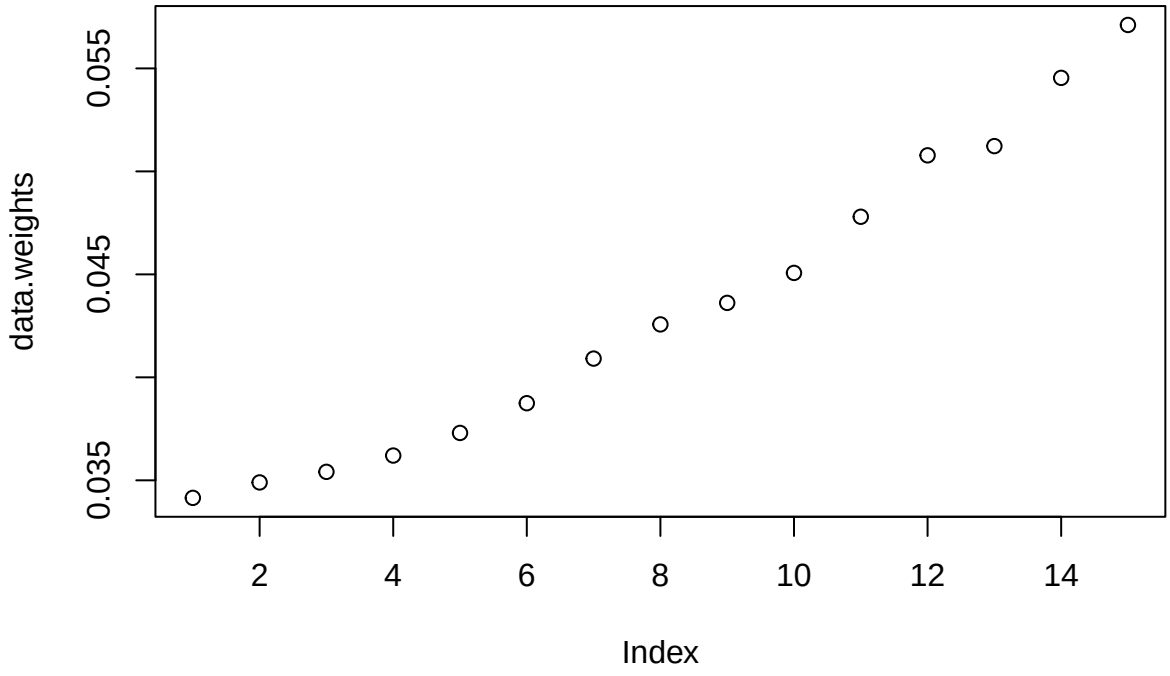
sisaan disekitar 0 menandakan nilai harapan galat sama dengan nol. Lebar pita sama untuk setiap nilai dugaanyang menandakan ragam homogen. Namun, karena bentuk pola tebaran sisaan vs yduga berpola, yaitu pola kurva, maka tidak memenuhi asumsi kondisi Gauss-Markov.

#TRANSFORMASI WLS

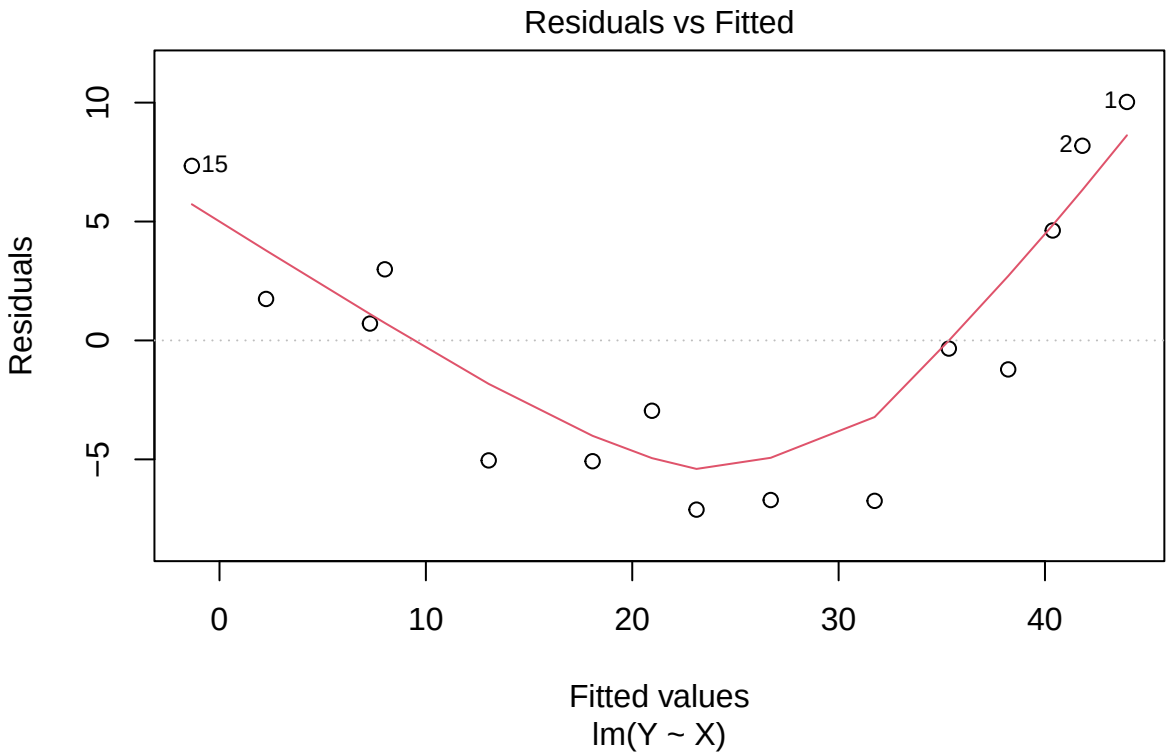
```
resid_abs <- abs(model_lm$residuals)
fitted_val <- model_lm$fitted.values
fit <- lm(resid_abs ~ fitted_val, data)
data.weights <- 1 / fit$fitted.values^2
data.weights
```

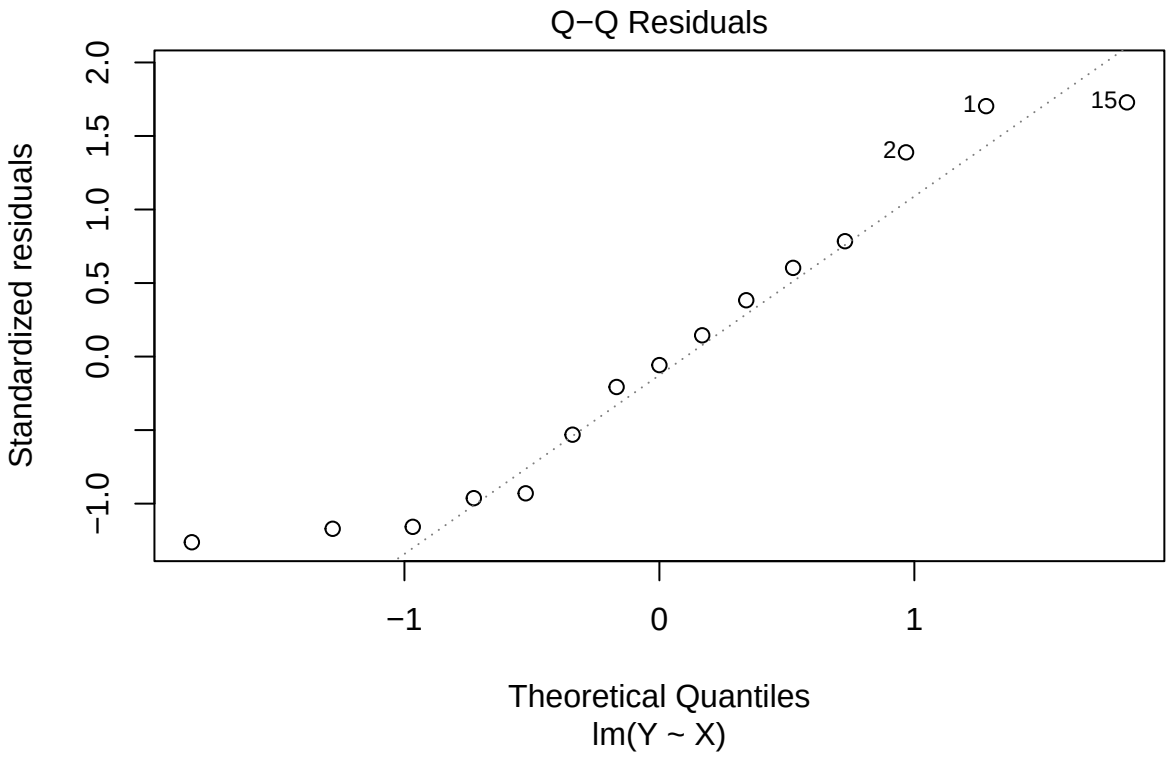
```
##          1          2          3          4          5          6          7
## 0.03414849 0.03489798 0.03541143 0.03620311 0.03730067 0.03874425 0.04091034
##          8          9         10         11         12         13         14
## 0.04257072 0.04361593 0.04507050 0.04779711 0.05077885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

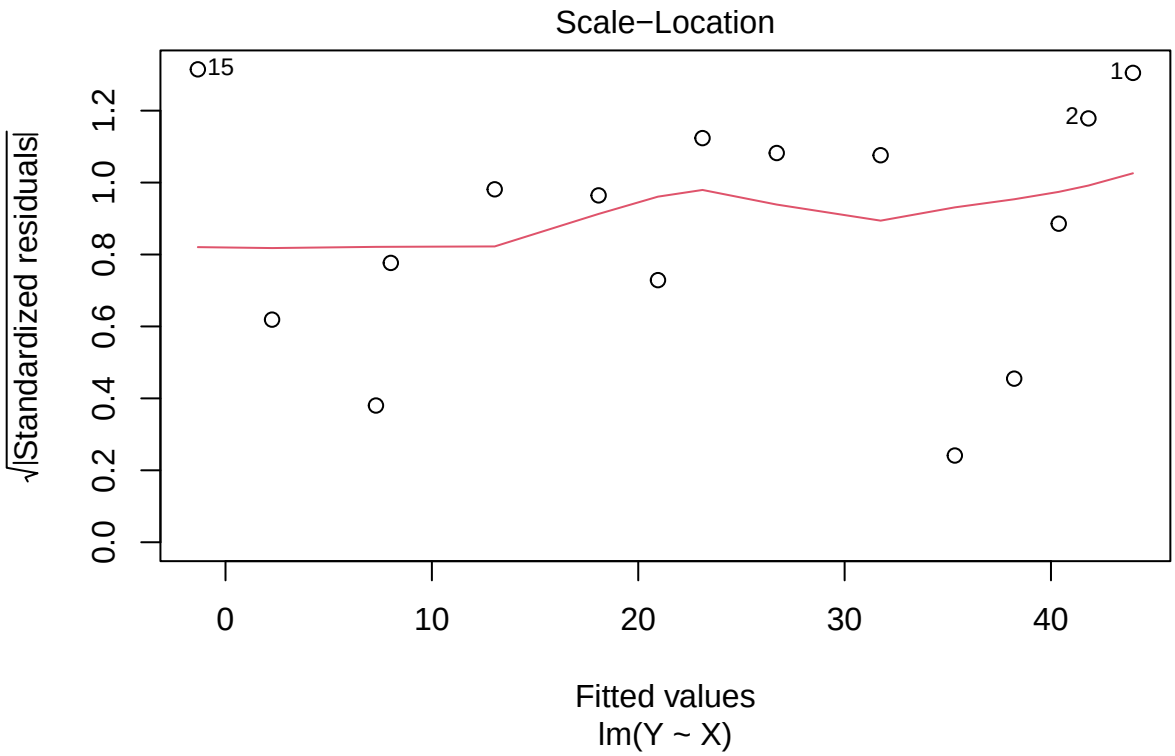
```
plot(data.weights)
```

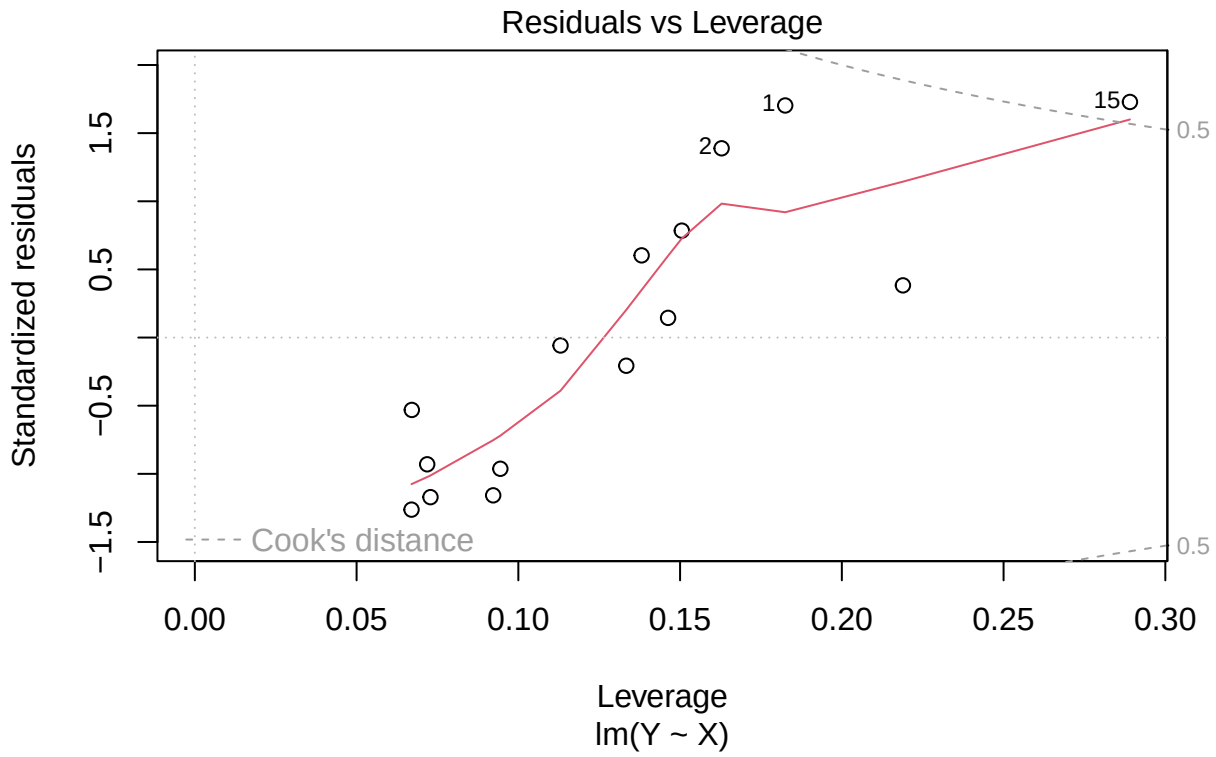



```
model_weighted <- lm(Y~X, data = data, weights = data.weights)
plot(model_weighted)
```


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_weighted)
```

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

#UJI FORMAL ASUMSI GAUSS MARCOV ##1. Nilai Harapan sisaan tidak sama dengan nol

```r
t.test(model_lm$residuals,mu = 0,conf.level = 0.95)
```

```
##
##  One Sample t-test
##
## data:  model_lm$residuals
## t = -4.9493e-16, df = 14, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -3.143811  3.143811
## sample estimates:
##     mean of x
## -7.254614e-16
```

Dengan menggunakan t-test, kesimpulannya terima H0 karena p-value > 0.05. Jadi, benar bahwa nilai harapan sisaan sama dengan nol.

##2. Ragam Sisaan Homogen H0: Ragam sisaan homogen H1: Ragam sisaan tidak homogen

```r
kehomogenan = lm(formula = abs(model_lm$residuals) ~ X, # y: abs residual
data = data)
summary(kehomogenan)
```

```
##
## Call:
## lm(formula = abs(model_lm$residuals) ~ X, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2525 -1.7525  0.0235  2.0168  4.2681
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.45041    1.27241   4.284  0.00089 ***
## X           -0.01948    0.03456  -0.564  0.58266
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.714 on 13 degrees of freedom
## Multiple R-squared:  0.02385,    Adjusted R-squared:  -0.05124
## F-statistic: 0.3176 on 1 and 13 DF,  p-value: 0.5827
```

Kesimpulan yang dapat ditarik adalah tak tolak atau terima H0 karena p-value $>$ alpha $= 0.05$. Maka, ragam sisaan homogen.

##3. Sisaan Saling Bebas

H0: Sisaan saling bebas H1: Sisaan tidak saling bebas.

```r
dwtest(model_lm)
```

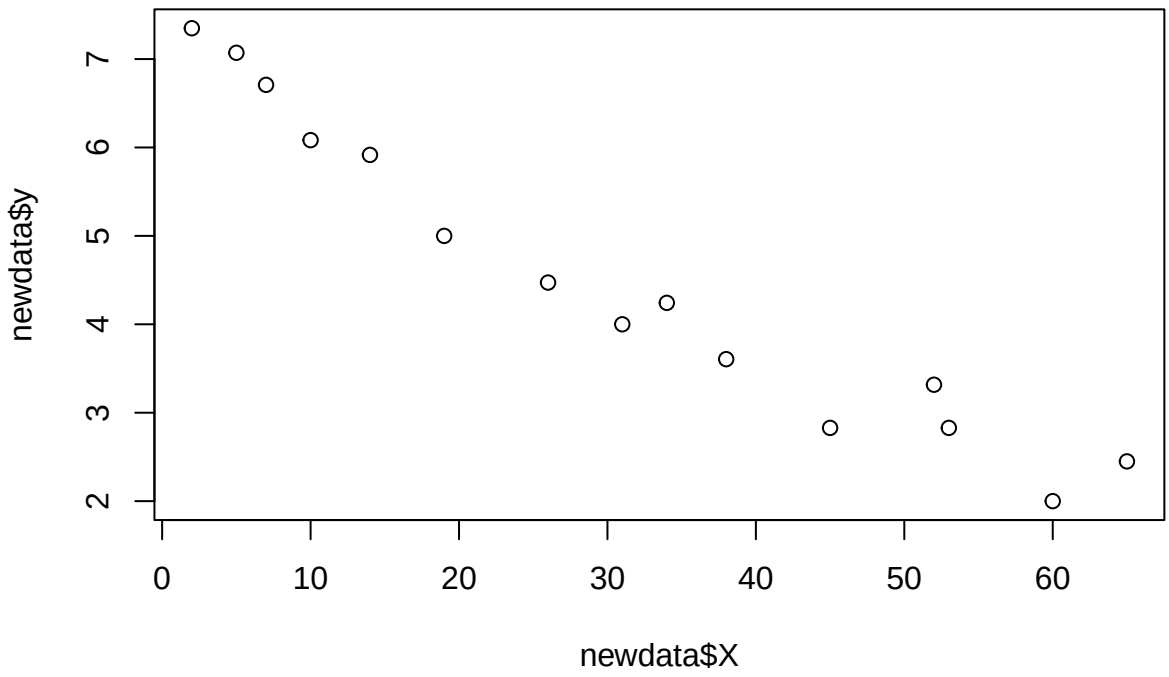
```
##
##  Durbin-Watson test
```

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

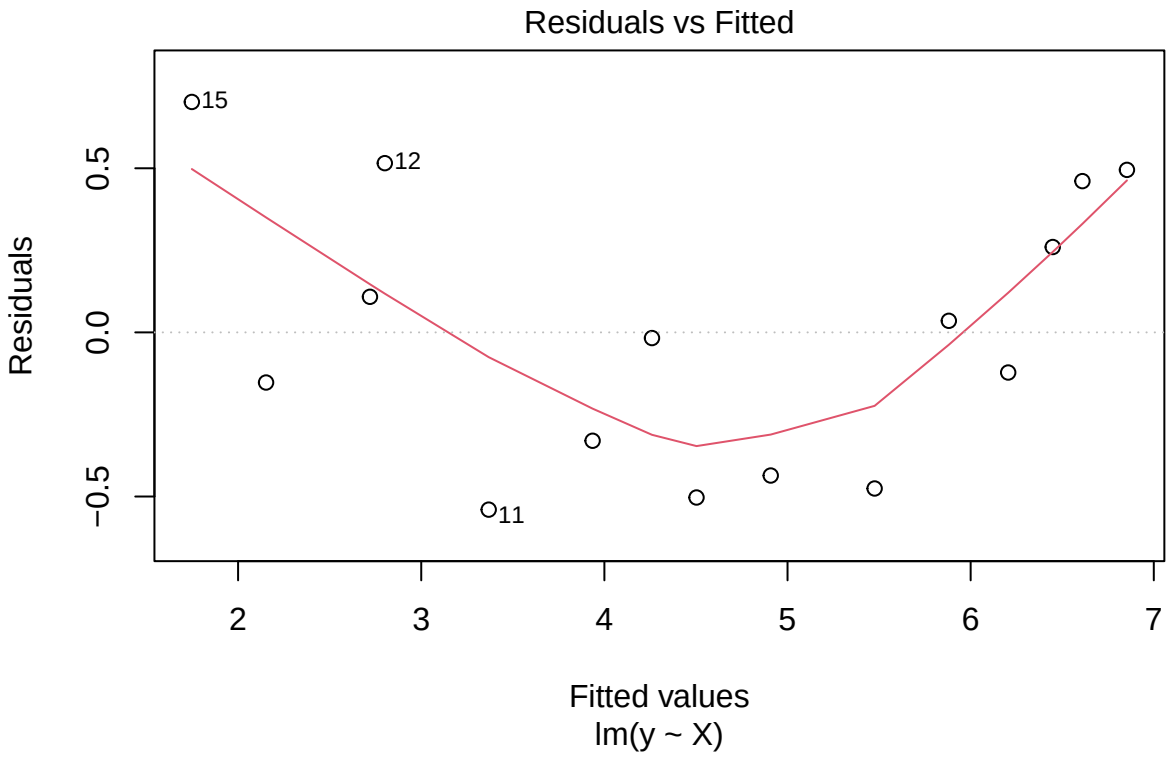
Karena p-value = 1.333e-05 (pada DW test) < alpha = 0.05, maka tolak H0, sisaan tidak saling bebas,asumsi tidak terpenuhi (ada autokorelasi).

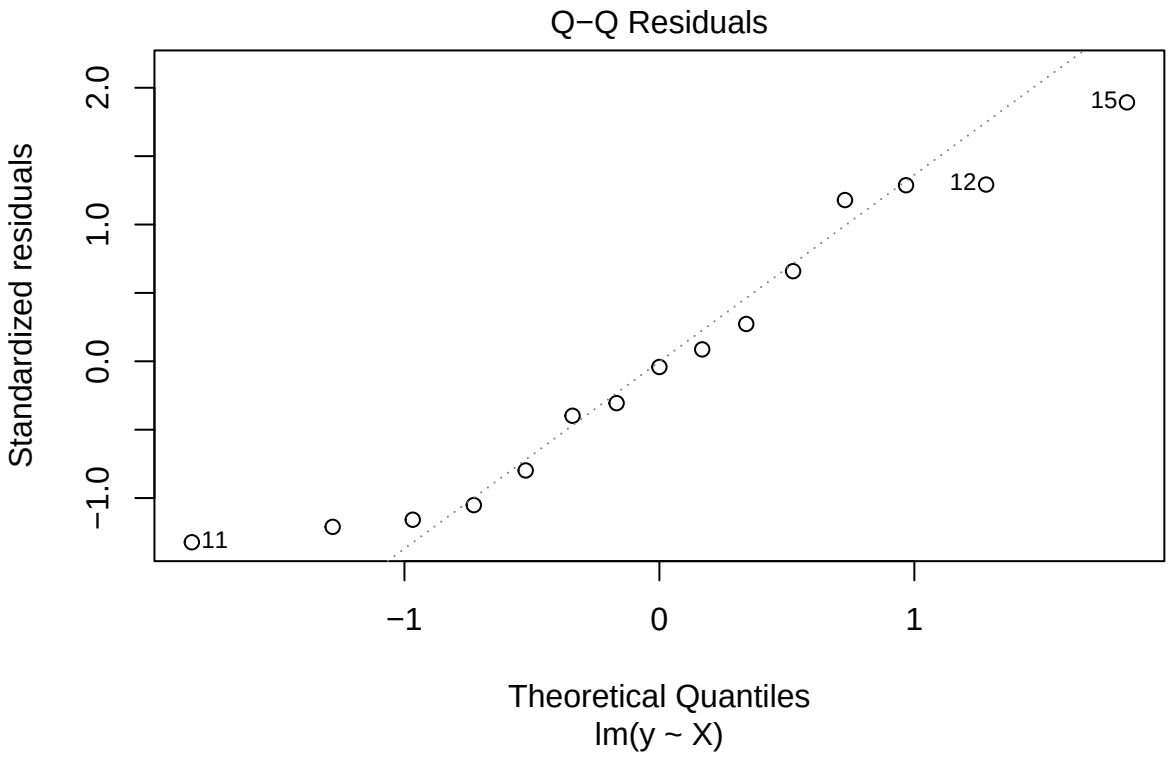
#TRANSFORMASI AKAR PADA X, Y , atau X dan Y

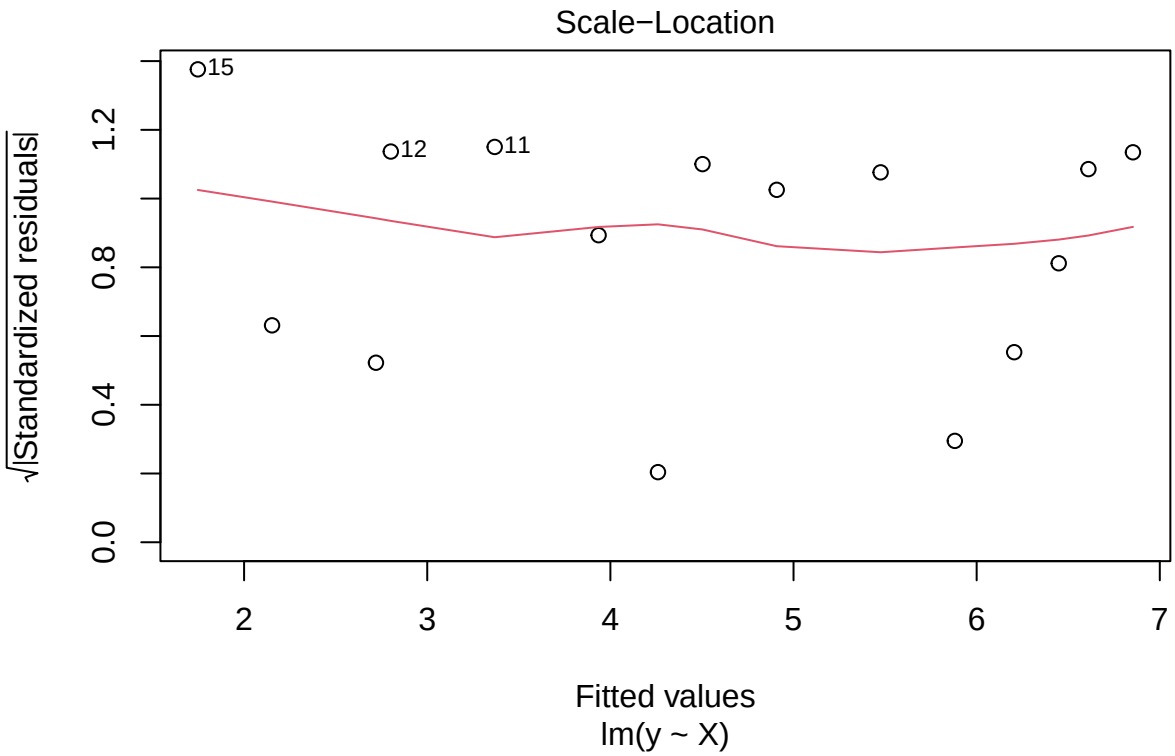
```
newdata <- data %>%
  mutate(y = sqrt(Y)) %>%
  mutate(x = sqrt(X))
model_sqrtx <- lm(y ~ X, data = newdata)
plot(x = newdata$X, y = newdata$y)
```

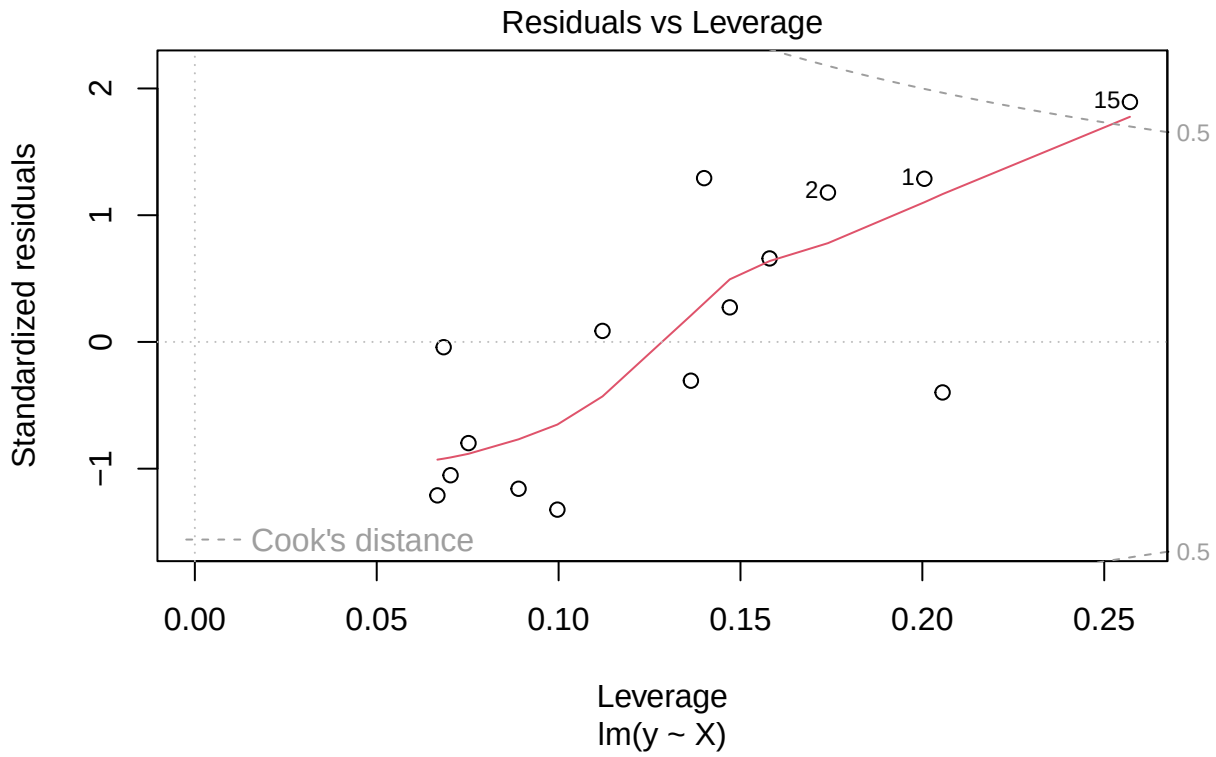



```
plot(model_sqrtx)
```


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_sqrtx)
```

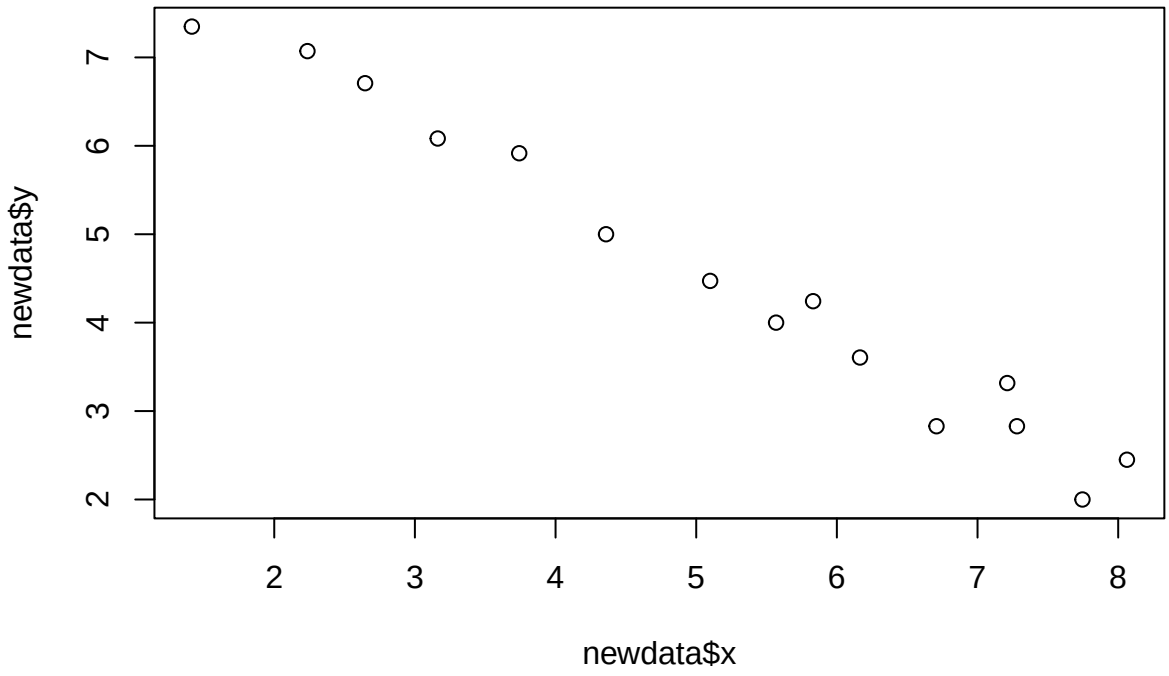
```
##
## Call:
## lm(formula = y ~ X, data = newdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53998 -0.38316 -0.01727  0.36045  0.70199
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.015455   0.201677   34.79 3.24e-14 ***
## X           -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 AUTO KORELASI MODEL REGRESI TRANSFORMASI

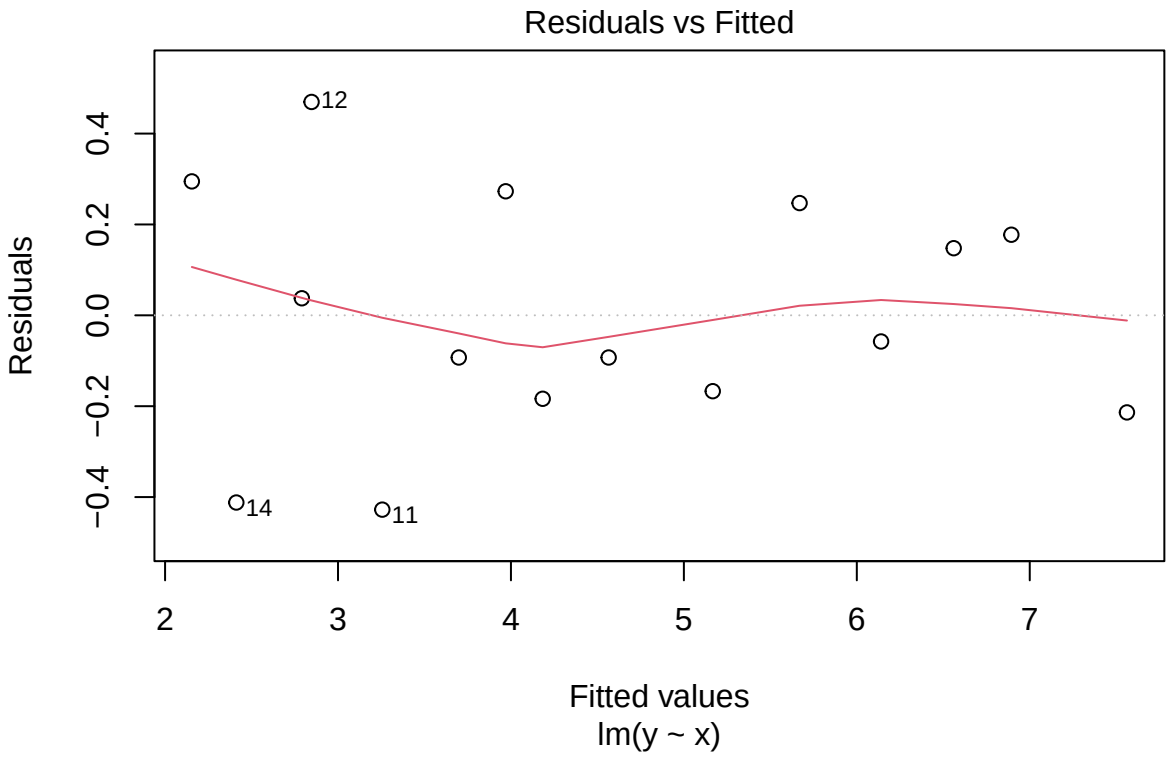
```
dwtest(model_sqrtx)
```

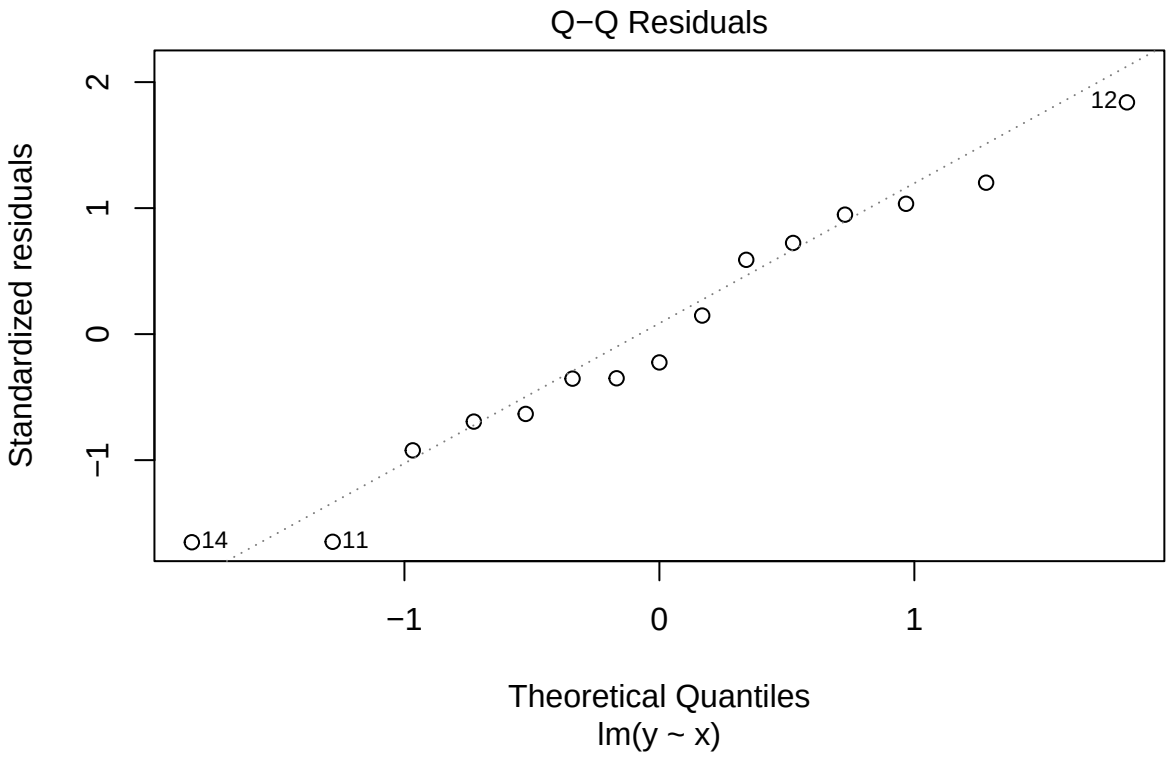
```
##
##  Durbin-Watson test
##
## data:  model_sqrtx
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0
```

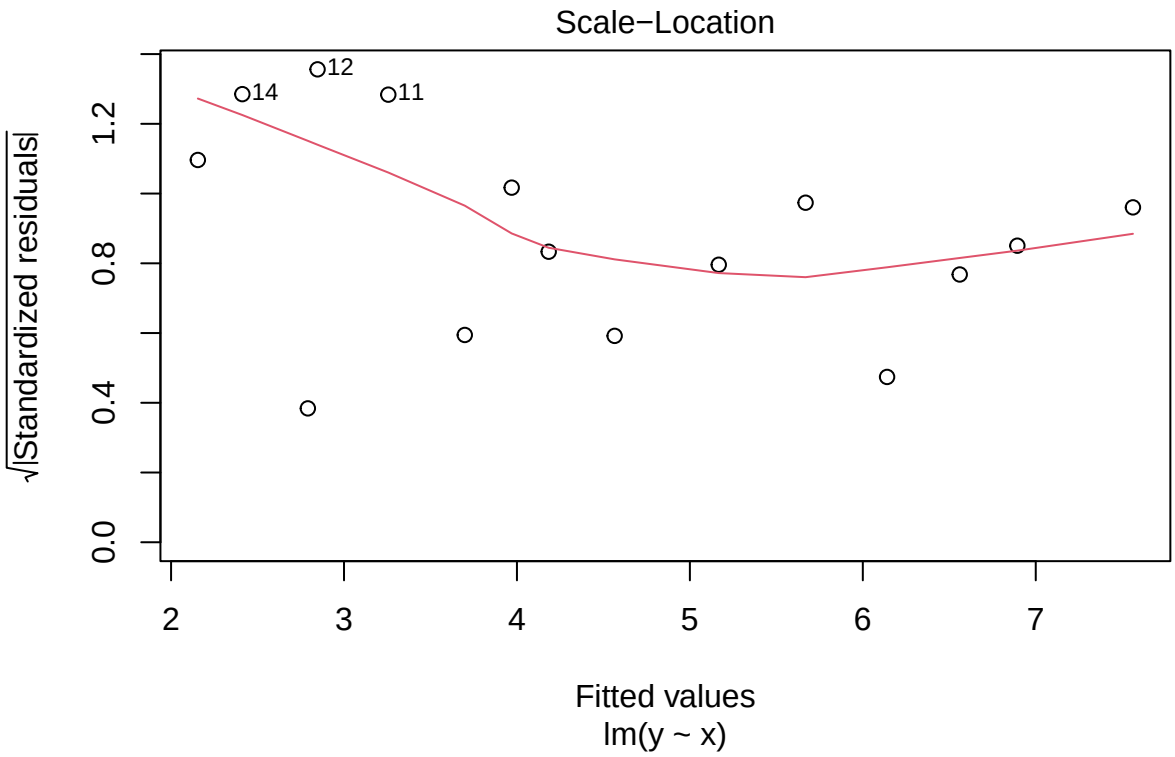
```
model_sqrt <- lm(y ~ x, data = newdata)
plot(x = newdata$x, y = newdata$y)
```

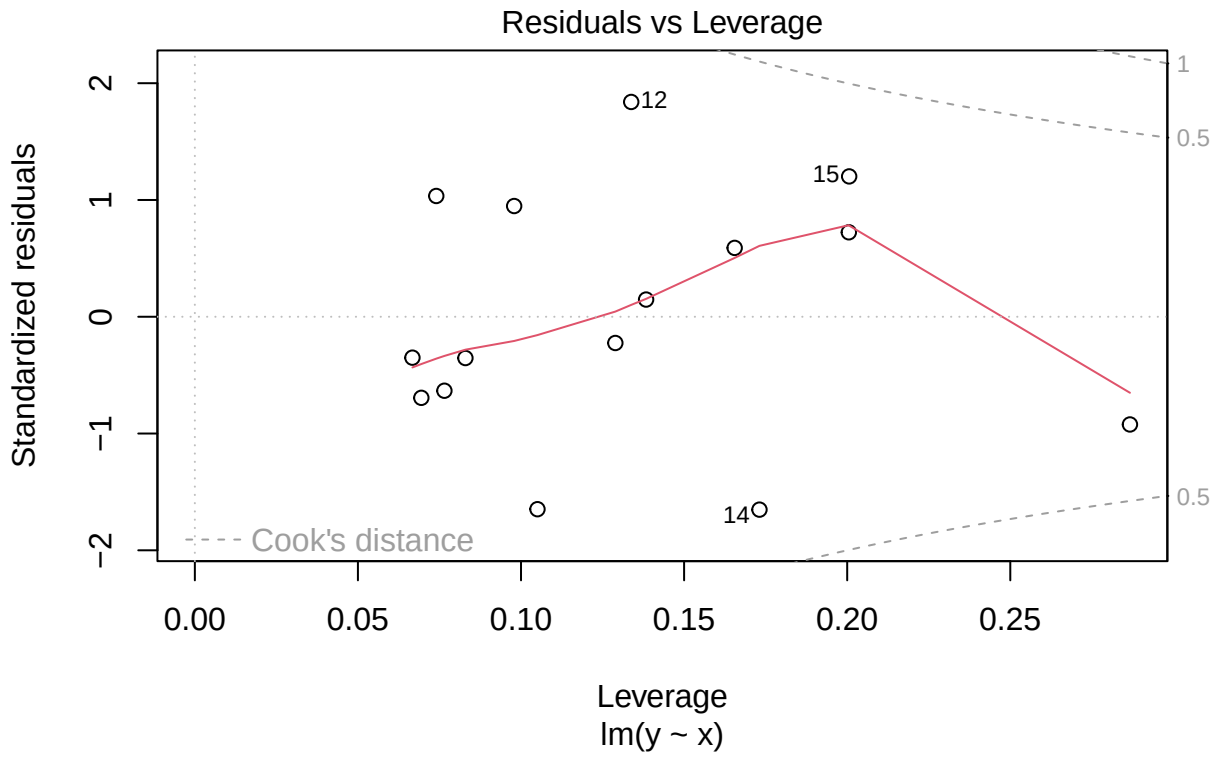



```
plot(model_sqrt)
```


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 = newdata)
##
## 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
```

#UJI AUTO KORELASI MODEL REGRESI

```r
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 :

Nilai p yang lebih besar dari 0.05 menunjukkan bahwa tidak ada bukti yang cukup untuk menolak hipotesis nol. Dalam kasus ini, hipotesis nol adalah tidak ada autokorelasi.

Dari hasil transformasi di atas dapat diambil kesimpulan jika transformasi akar Y membuat persamaan regresi menjadi lebih efektif. Model regresi setelah transformaasi:

$$Y = 8.71245 - 0.81339X + e$$
$$Y = Y^{\frac{1}{2}}$$
$$X = X^{\frac{1}{2}}$$

Dilakukan Transformasi Balik menjadi

$$Y = (8.71245 - 0.81339X^{\frac{1}{2}})^2 + e)$$

Interpretasi Model ini menunjukkan bahwa Y berbanding terbalik dengan akarX, dengan hubungan kuadratik. Semakin besar nilai akarX, semakin kecil nilai rata-rata Y, dengan kecepatan yang semakin meningkat. Puncak kurva menunjukkan nilai rata-rata Y maksimum untuk nilai X tertentu. Konstanta 8.71245 mewakili nilai Y ketika X sama dengan 0. Koefisien -0.81339 adalah koefisien regresi untuk variabel X. Nilai negatif menunjukkan hubungan invers antara Y dan akarX. Semakin besar nilai akarX, semakin kecil nilai Y.Pangkat dua pada koefisien regresi menunjukkan bahwa hubungan antara Y dan X adalah kuadratik. Artinya, perubahan Y tidak proporsional dengan perubahan X, tetapi berubah dengan kecepatan yang semakin meningkat.