Tugas Kuliah Analisis Regresi Pertemuan 7

Nafisa Zalfa Maulida - G1401221092

2024-03-06

##Data

```
library(readxl)
```

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

```
data<-read_excel("D:/SEMESTER 4/Analisis Regresi/Pertemuan 7/Data Anreg
Kuliah Pertemuan 7.xlsx")
data
```

```
## # A tibble: 15 × 3
##      No.     X     Y
##    <dbl> <dbl> <dbl>
##  1     1     2    54
##  2     2     5    50
##  3     3     7    45
##  4     4    10    37
##  5     5    14    35
##  6     6    19    25
##  7     7    26    20
##  8     8    31    16
##  9     9    34    18
## 10    10    38    13
## 11    11    45     8
## 12    12    52    11
## 13    13    53     8
## 14    14    60     4
## 15    15    65     6
```

```
library(tidyverse)
```

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

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

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

```
## Warning: package 'forcats' 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 ──
## ✓ dplyr     1.1.4     ✓ readr     2.1.4
```

```
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ─────────────────────────────────────────
tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ 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
```

```
library(stats)
```

##Model Regresi Awal

```
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
```

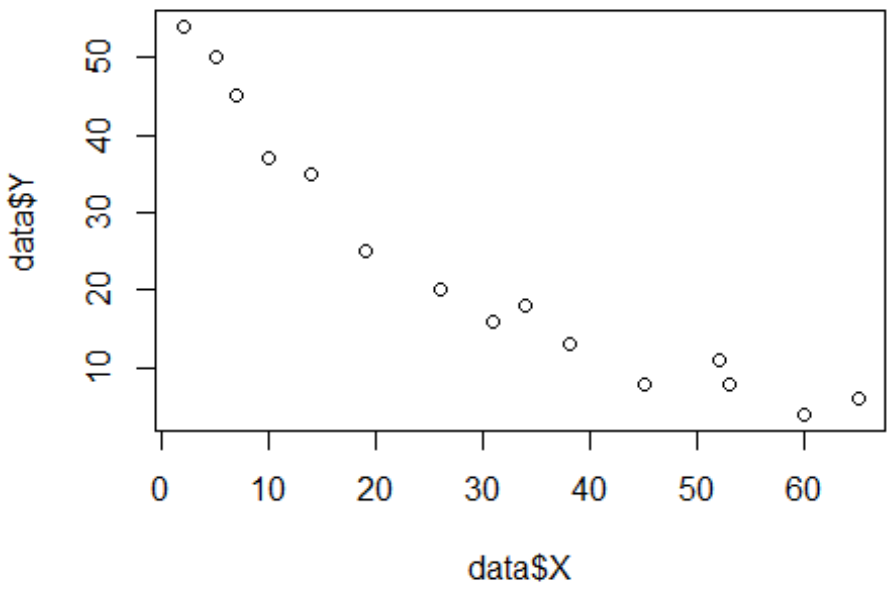
Model Regresi:

$$\hat{Y} = 46.46041 - 0.75251X + e$$

Karena belum melalui serangkaian uji asumsi, maka diperlukan eksplorasi kondisi, pengujian asumsi Gauss-Markov, dan normalitas untuk menghasilkan model terbaik.

##Eksplorasi Data #Plot Hubungan X dan Y

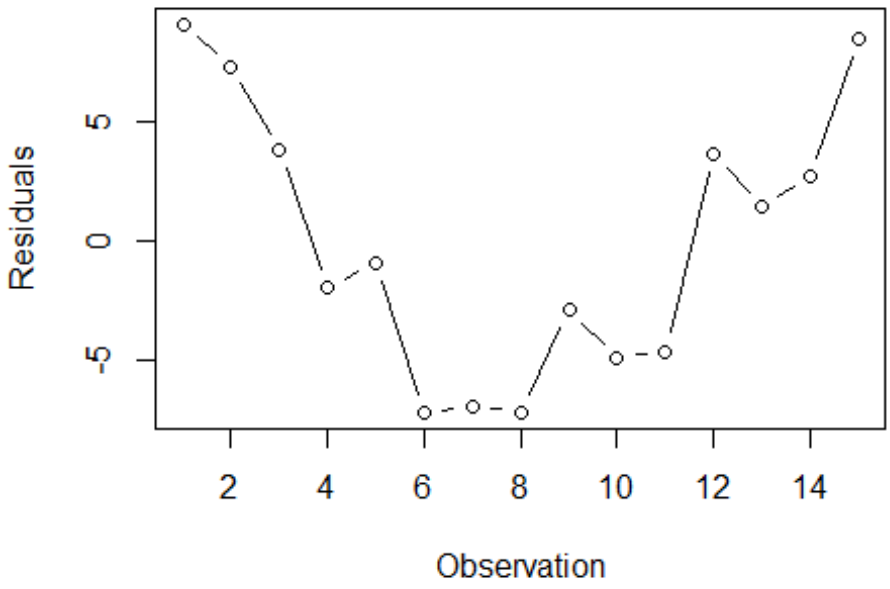
```
plot(x = data$X, y = data$Y)
```



Berdasarkan scatter plot di atas, dapat diketahui bahwa X dan Y tidak mempunyai hubungan linear karena cenderung membentuk pola parabola.

##Plot Sisaan vs Urutan

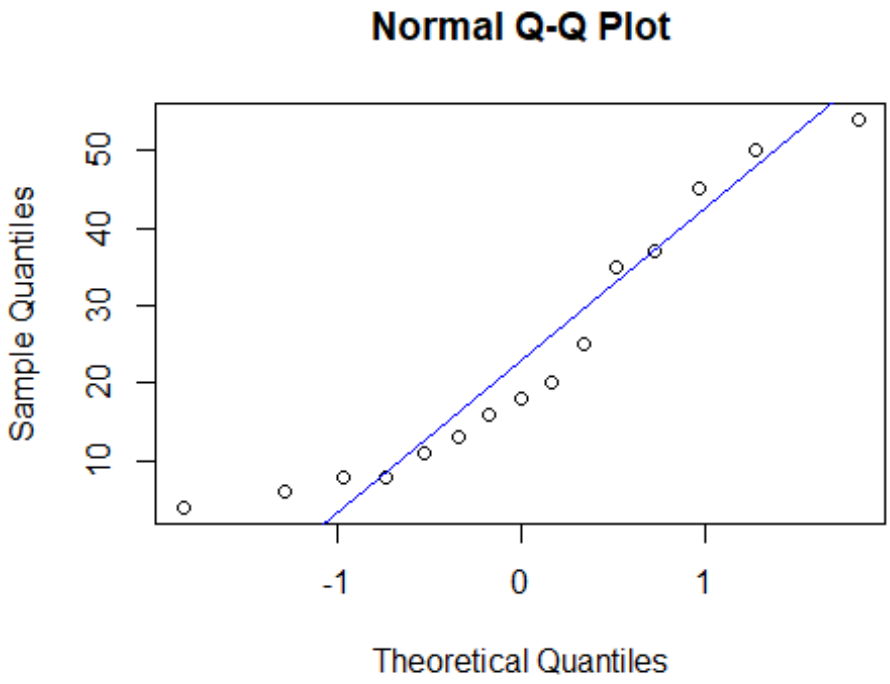
```
plot(x = 1:dim(data)[1],
y = model_lm$residuals,
type = 'b',
ylab = "Residuals",
xlab = "Observation")
```



Sebaran tersebut membentuk pola kurva menandakan sisaan tidak saling bebas.

##Uji Normalitas

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

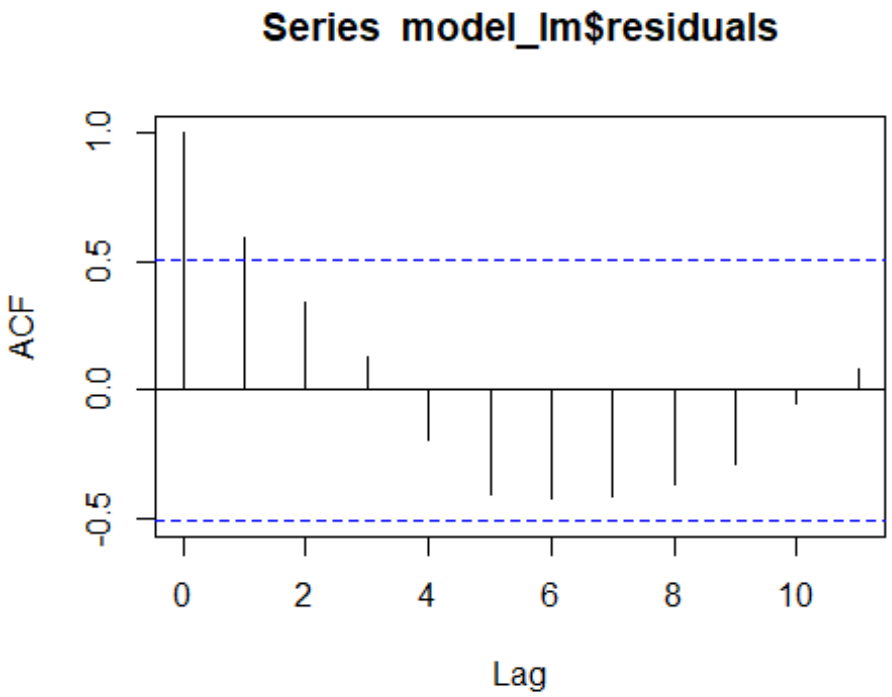



```
shapiro.test(data$Y)

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

QQ Plot cenderung menunjukkan bahwa data yang digunakan menyebar normal. Hal tersebut juga didukung dengan hasil Shapiro Test yang besarnya lebih dari 0.05, yaitu 0.89636.

##Uji Autokorelasi

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



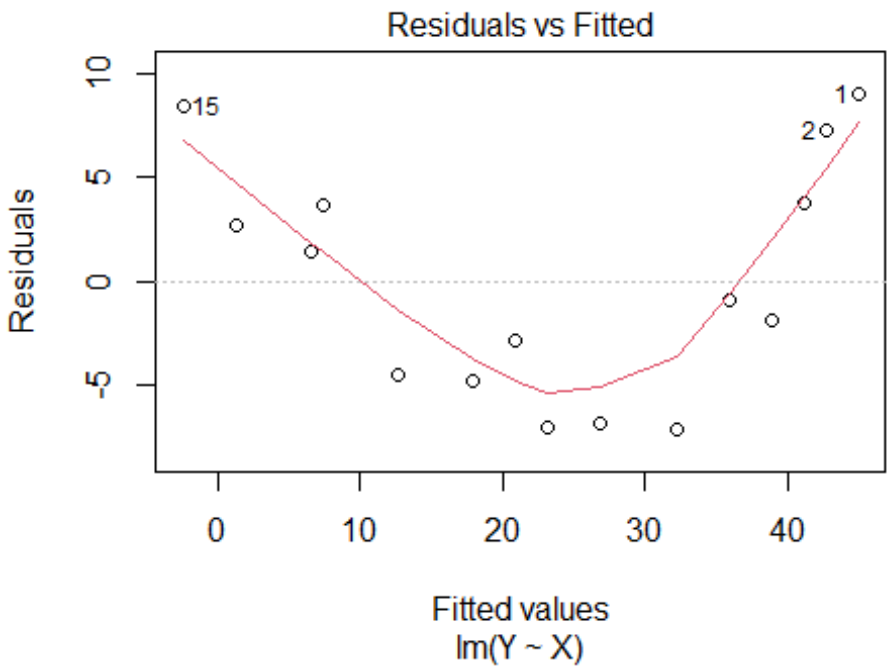

```
dwtest(model_lm)

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

Nilai autokorelasi pada lag 1 dan lag 2 berada di luar batas kepercayaan 95%, yaitu pada lag 1 = 0,5 dan pada lag 2 = 0.4. Hal tersebut menunjukkan bahwa autokorelasi pada lag 1 dan 2 adalah signifikan.

Oleh karena itu, asumsi Gauss-Markov tidak terpenuhi (asumsi non-autokorelasi). Hal tersebut pun diperkuat dengan p-test pada uji Durbin-Watson bernilai kurang dari 0.05.

##Uji Homoskedastisitas

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



Grafik tersebut menunjukkan bahwa varians residual konstan. Varian residual cenderung meningkat seiring dengan nilai prediksi. Hal tersebut akan mengindikasi bahwa homoskedastisitas terjadi.

##Transformasi

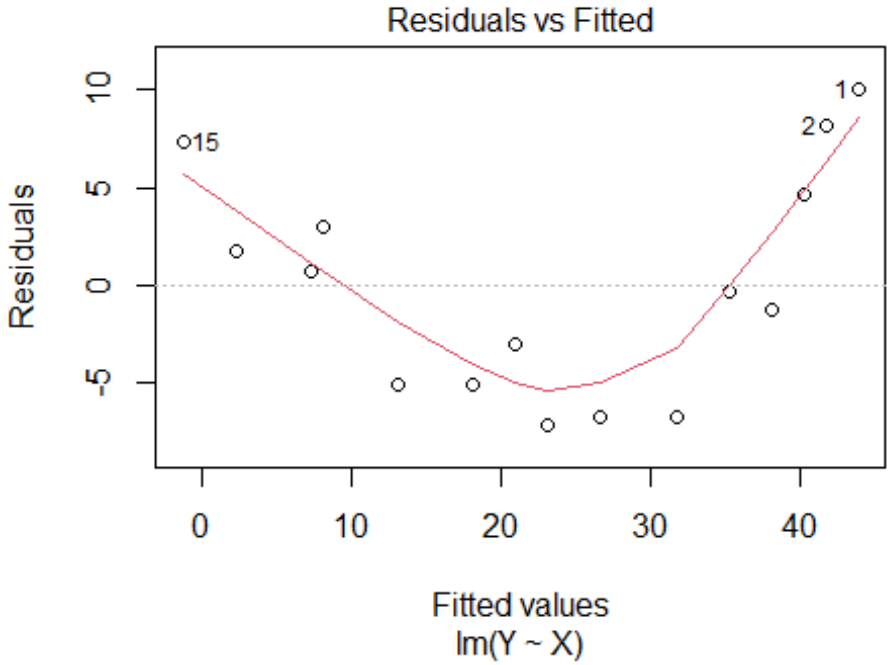
##WLS

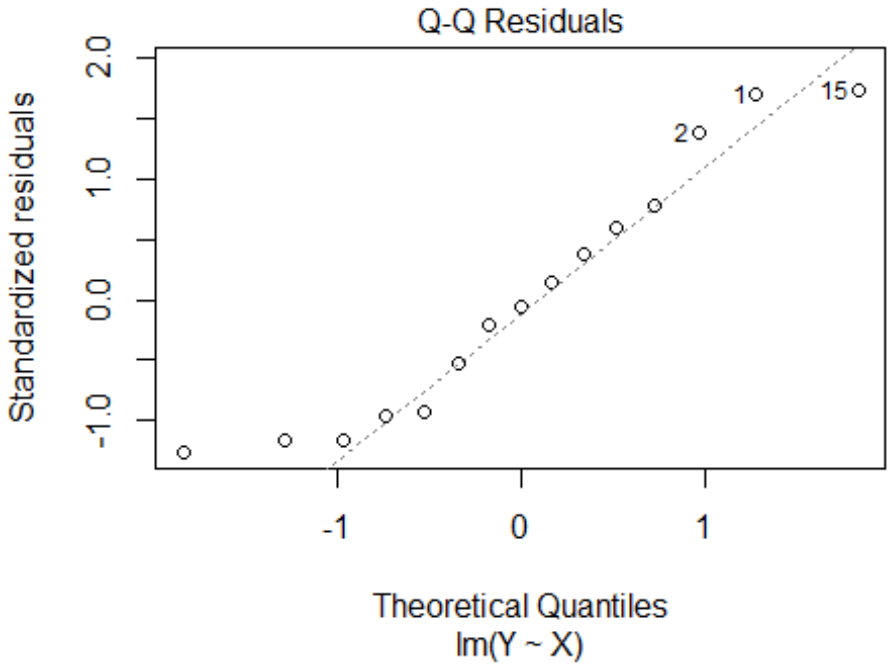
```
resid_abs <- abs(model_lm$residuals)
fitted_val <- model_lm$fitted.values
fit <- lm(resid_abs ~ fitted_val, data)
data.weights <- 1 / fit$fitted.values^2
data.weights
```

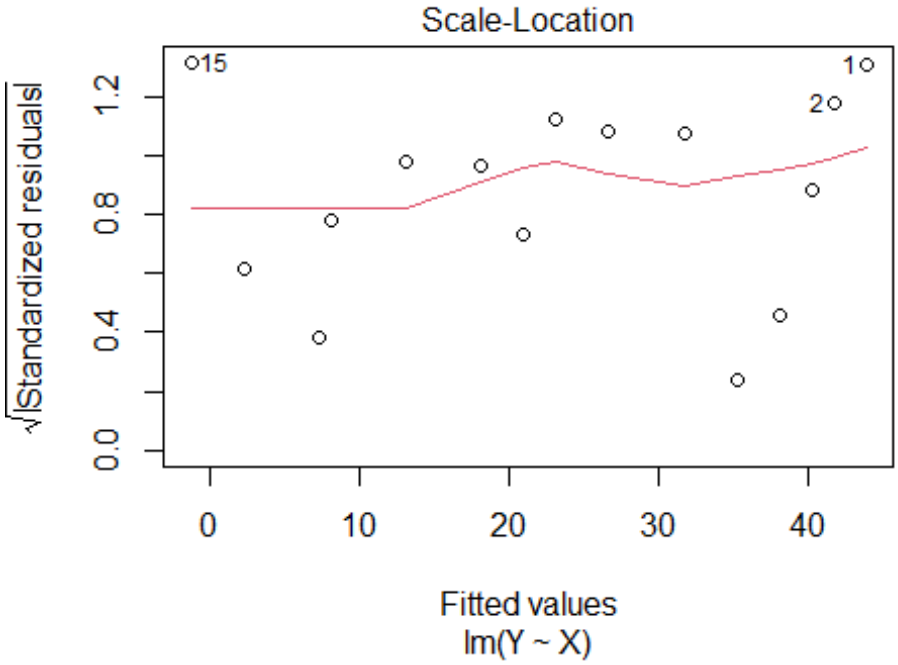
```
##          1          2          3          4          5          6
7
## 0.03414849 0.03489798 0.03541143 0.03620311 0.03730067 0.03874425
0.04091034
##          8          9         10         11         12         13
14
## 0.04257072 0.04361593 0.04507050 0.04779711 0.05077885 0.05122749
0.05454132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

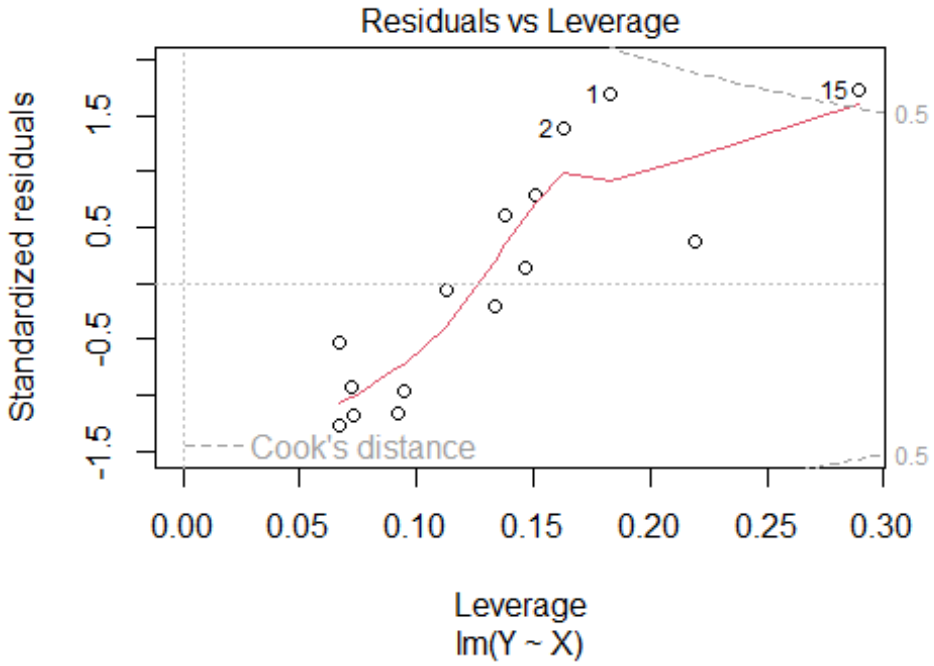
##Hasil model regresi yang terboboti:

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









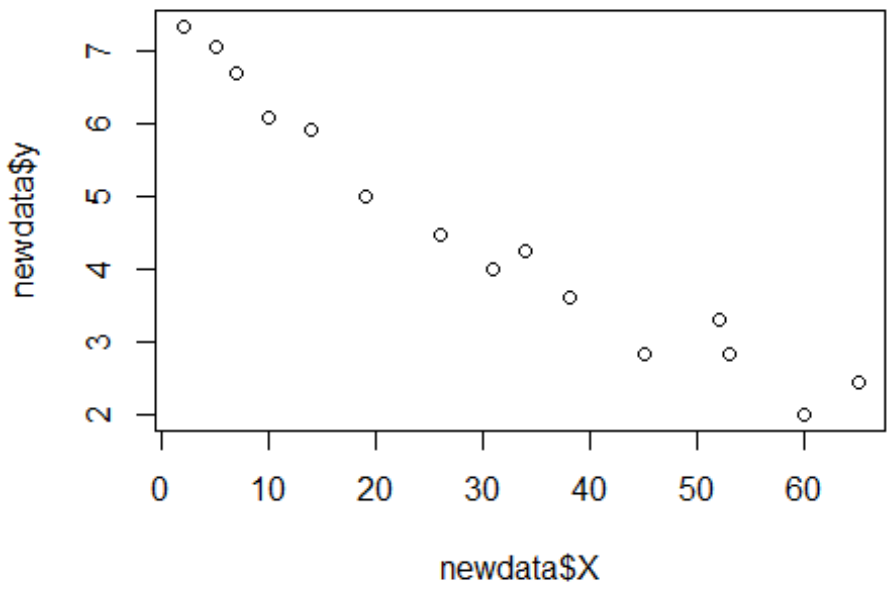

```
summary(model_weighted)

##
## 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
```

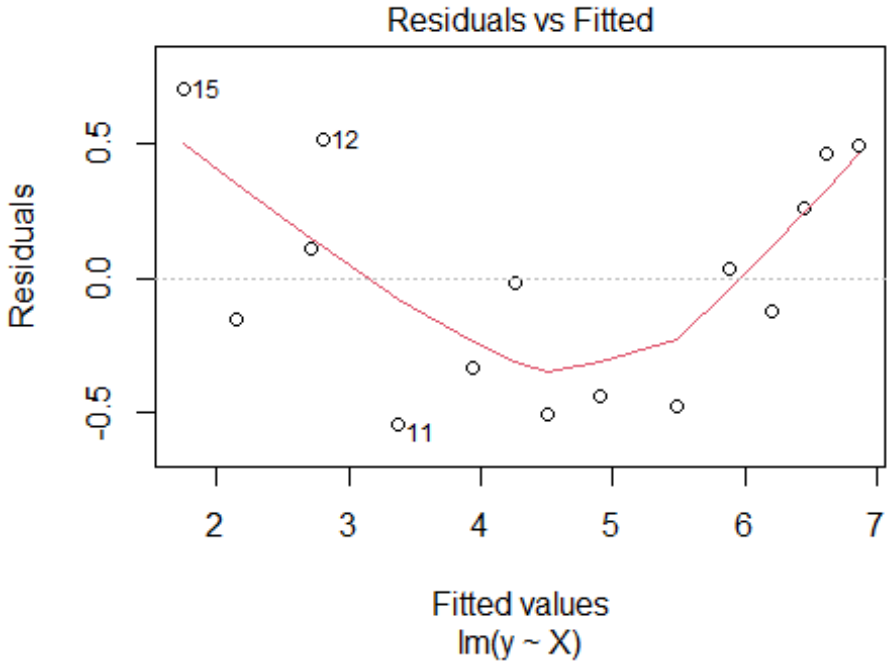
Berdasarkan hasil transformasi WLS, dapat diketahui bahwa WLS belum cukup efektif untuk mentransformasi model regresi. Hal itu dapat dibuktikan dari hasil eksplorasi yang masih belum memenuhi asumsi Gauss-Markov.

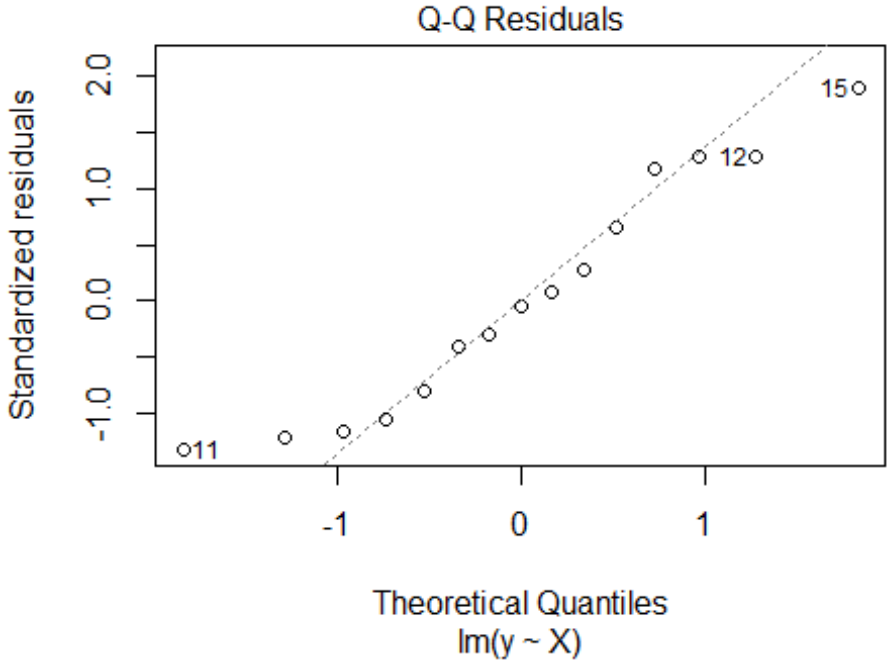
##Transformasi Akar: pada x,y atau X dan Y

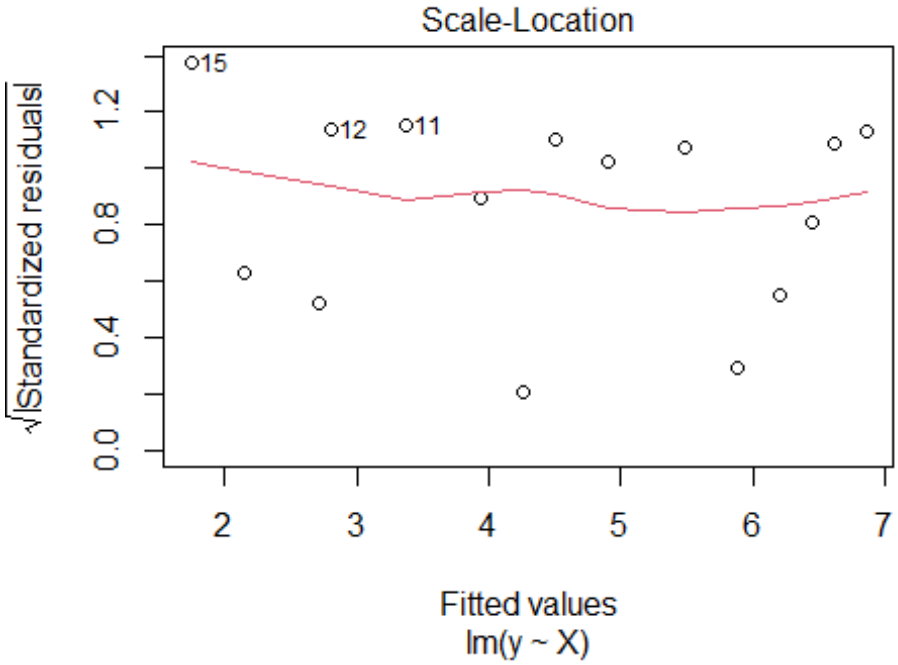
```
newdata <- data %>%
  mutate(y = sqrt(Y)) %>%
  mutate(x = sqrt(X))

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

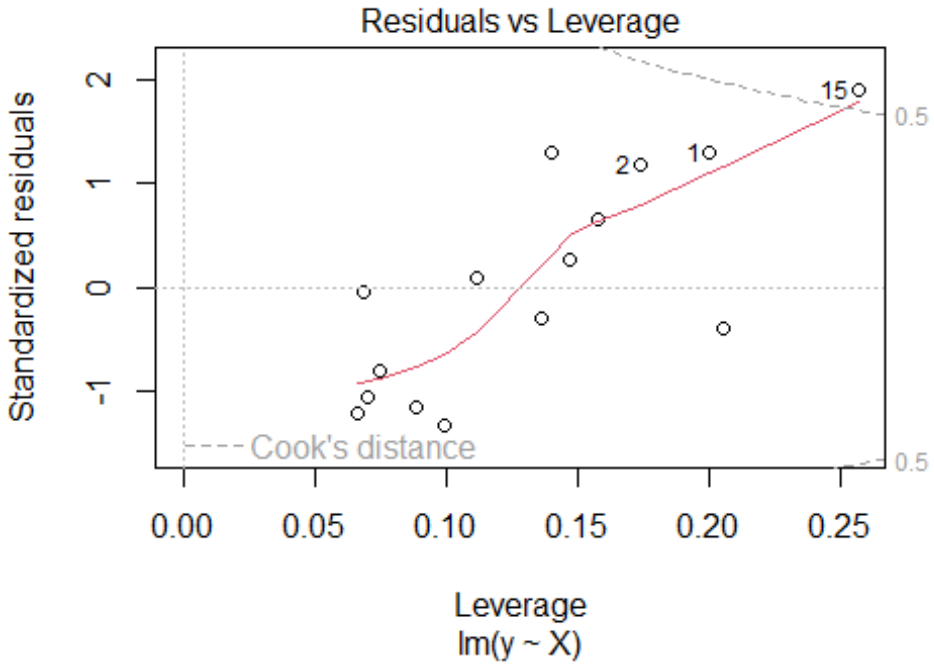



```
plot(model_sqrtx)
```


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


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






```
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
```

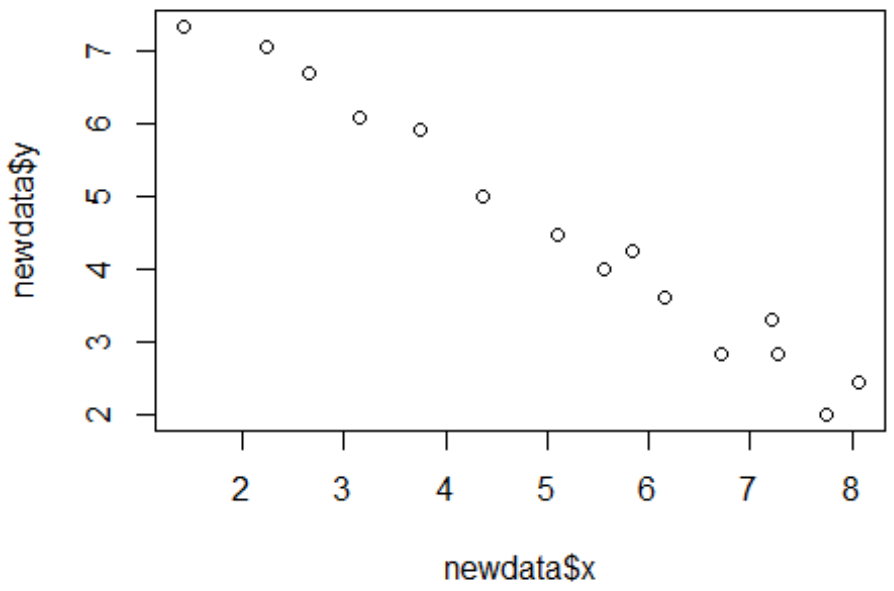
##Uji Autokorelasi Model Regresi Transformasi

```
dwtest(model_sqrtx)
```

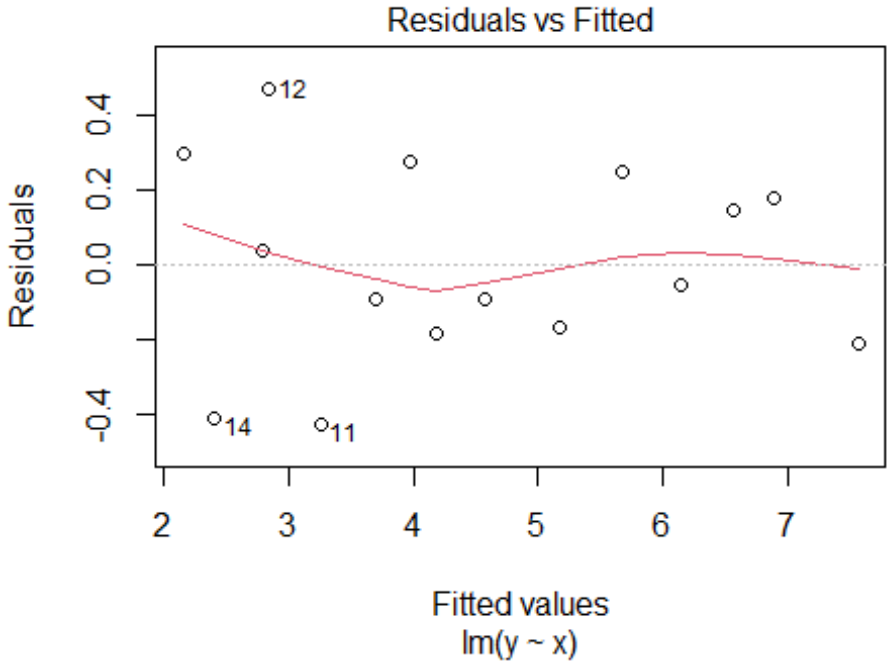
```
##
##  Durbin-Watson test
##
## data:  model_sqrtx
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0
```

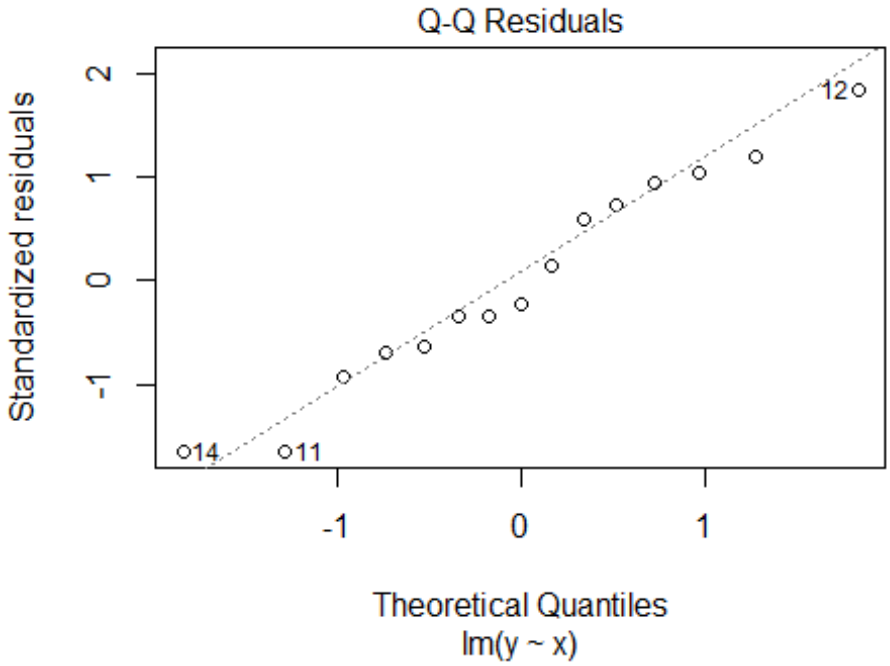
Nilai DW yang rendah dan p-value yang signifikan menunjukkan ada autokorelasi positif pada Durbin Watson. Selain itu, dibuktikan dengan p-value yang bernilai kurang dari 0.05.

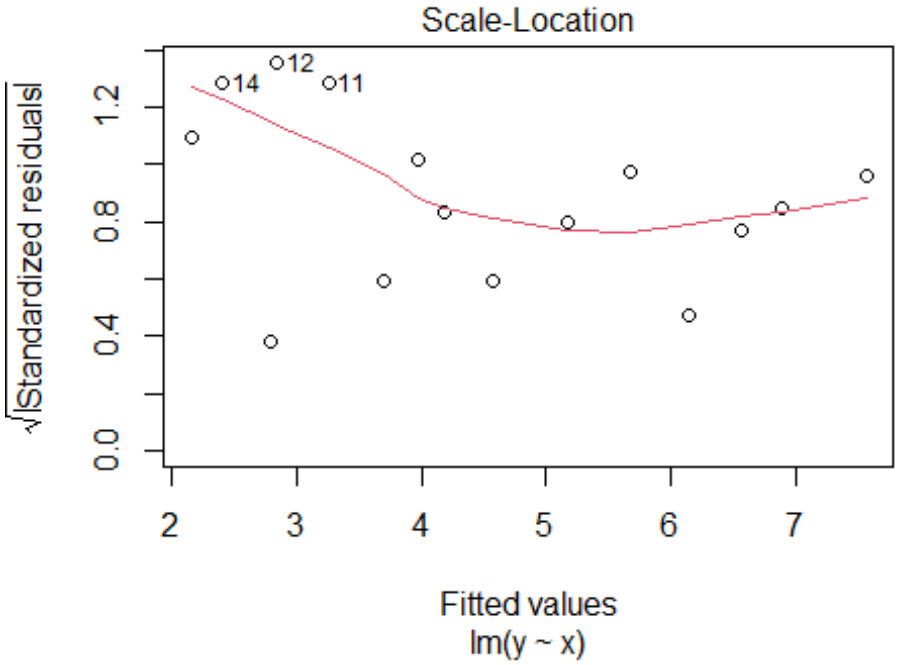
```
model_sqrt <- lm(y ~ x, data = newdata)
plot(x = newdata$x, y = newdata$y)
```

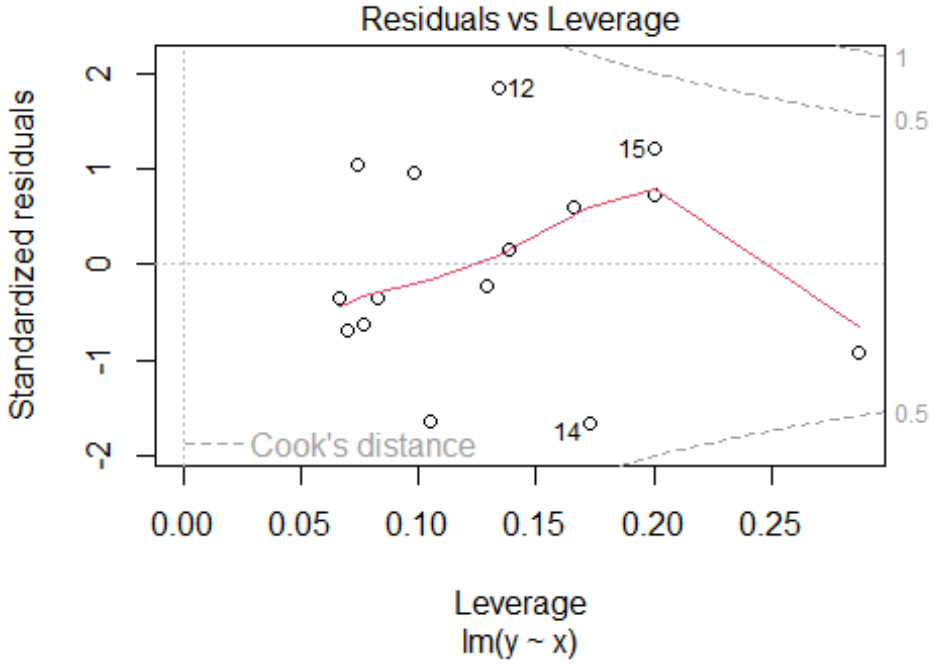



```
plot(model_sqrt)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
-0.4
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)






```
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 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
```

P-value lebih besar dari 0.05, yaitu 0.8629 menunjukkan bahwa tidak ada cukup bukti untuk menolak H0. Dimana H0 adalah tidak ada autokorelasi.

Dari hasil transformasi, dapat disimpulkan jika transformasi akar Y membuat persamaan regresi jadi lebih efektif dengan model regresi menjadi:

$$Y^* = 8.71245 - 0.81339X^* + e$$

$$Y^* = \sqrt{Y}$$

$$X^* = \sqrt{X}$$

#Dilakukan Transformasi Balik Menjadi:

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

#Interpretasi Model tersebut mengindikasi bahwa adanya hubungan berbanding terbalik (kuadrat negatif) antara Y dengan X. Saat X meningkat, Y akan cenderung turun dengan kecepatan yang semakin cepat. Nilai konstanta 8.71245 mewakili nilai Y ketika X=0. Koefisien regresi untuk variabel X adalah -0.81339. Semakin besar nilai absolut koefisien, semakin besar pengaruh X terhadap Y.