

Tugas Individu Analisis Regresi

Rakesha Putra Antique Yusuf

2024-03-05

```
library(readxl)
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.2
## Warning: package 'ggplot2' 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_conflict
s() —
## ✗ 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(ggribes)

## Warning: package 'ggribes' was built under R version 4.3.2

library(GGally)

## 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

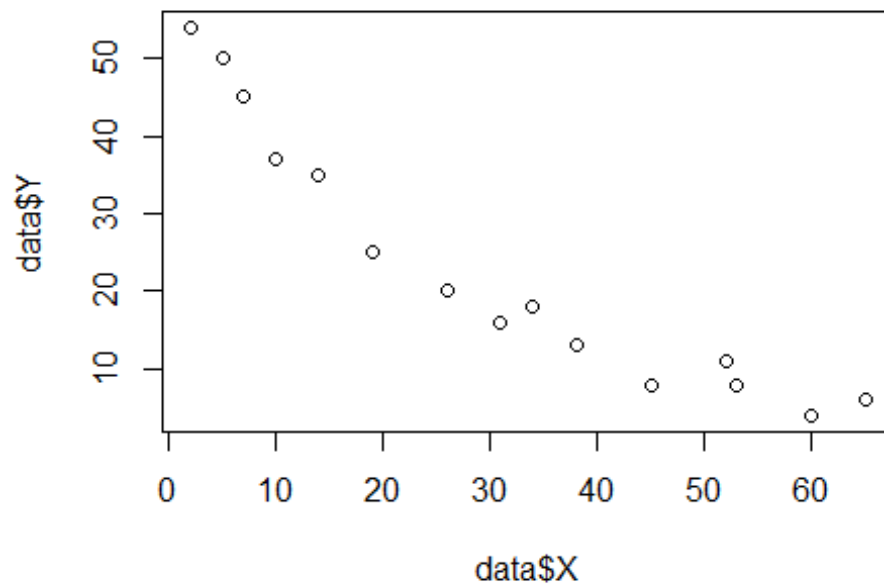
library(stats)
```

Data

```
data <- read_xlsx("C:/Users/RAKESHA/Downloads/Anreg Individu.xlsx")
```

Eksplorasi Data

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

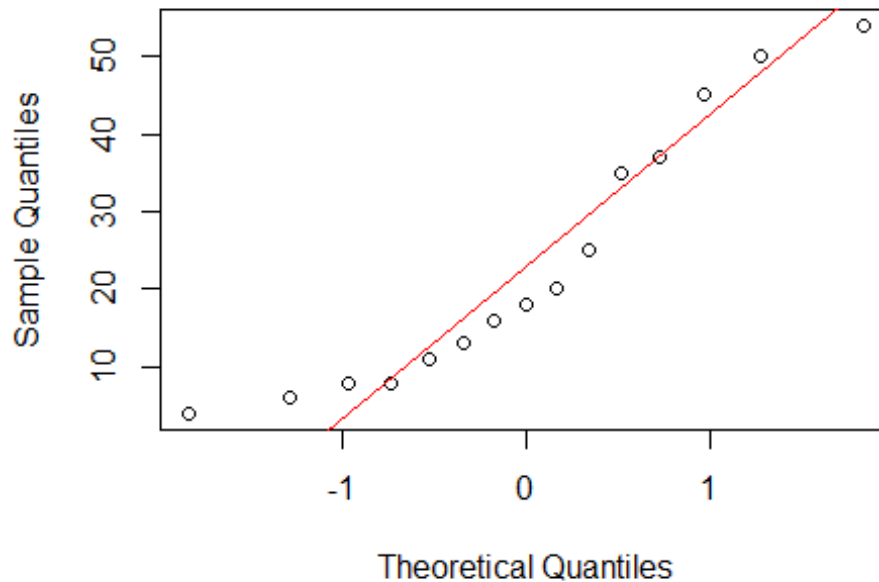


Dari scatter plot di atas menunjukkan jika Y dan X tidak berhubungan linear

Uji Normalitas

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

Normal Q-Q Plot



```
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

Declare 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

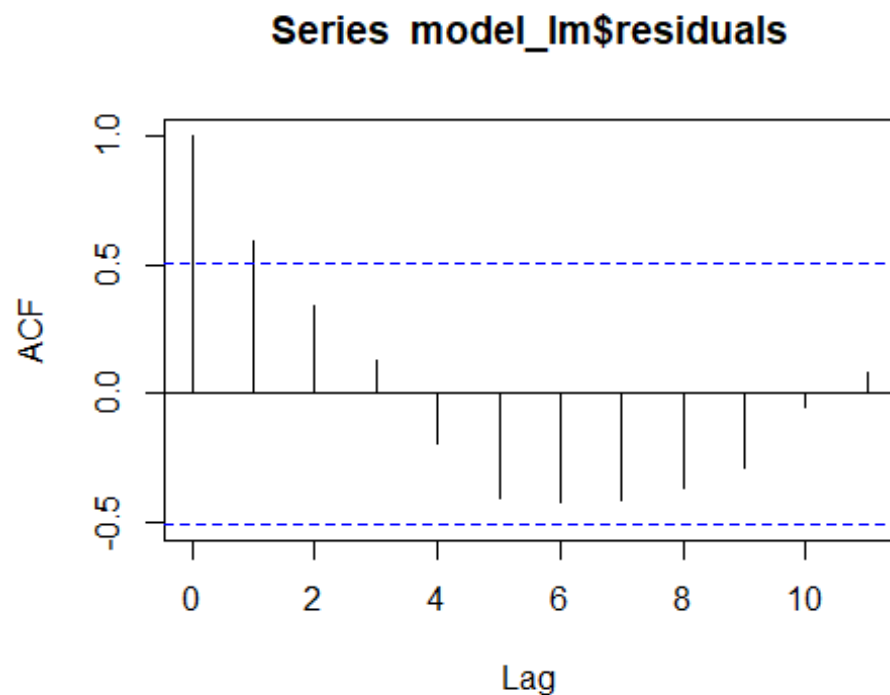
##
## Call:
## lm(formula = Y ~ X, data = data)
##
## Coefficients:
## (Intercept)          X
##    46.4604      -0.7525
```

Model Regresi:

$$\hat{Y} = 46.4604 - 0.7525X$$

Uji Autokorelasi

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



```
dwtest(model_lm)
```

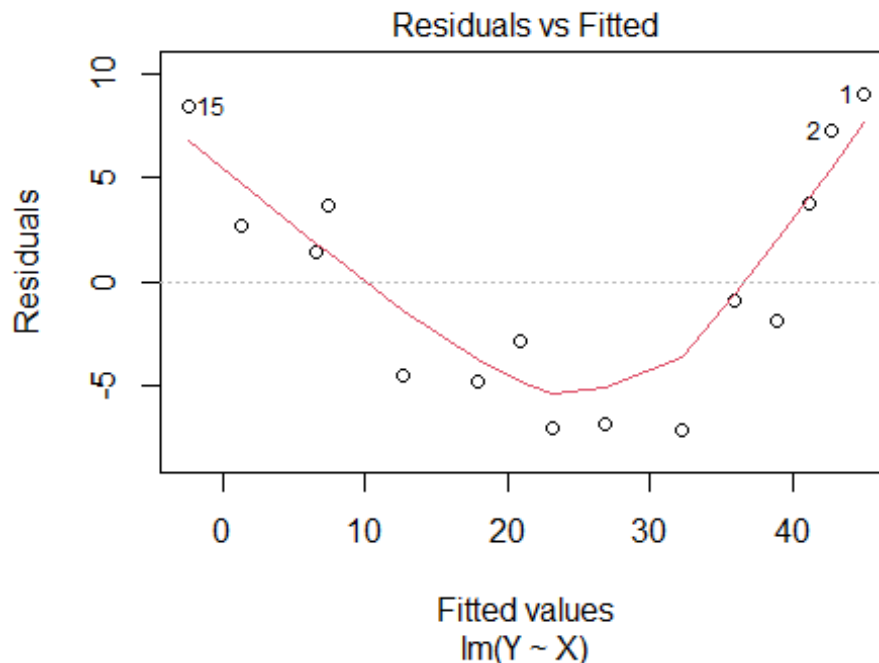
```
##
## Durbin-Watson test
##
## data: model_lm
## DW = 0.48462, p-value = 1.333e-05
## alternative hypothesis: true autocorrelation is greater than 0
```

Pada gambar ACF, nilai autokorelasi pada lag 1 adalah 0.5 dan nilai autokorelasi pada lag 2 adalah 0.4. Kedua nilai ini berada di luar batas kepercayaan 95%, yang menunjukkan bahwa autokorelasi pada lag 1 dan 2 signifikan.

Gambar tersebut menunjukkan adanya asumsi Gauss-Markov yang tidak terpenuhi, yaitu asumsi non-autokorelasi. Hal tersebut juga diperkuat dari p-test hasil Uji Durbin-Watson yang bernilai kurang dari 0,05

Uji Homoskedastisitas

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



Berdasarkan gambar di atas, terlihat bahwa varians residual konstan. Varian residual cenderung meningkat seiring dengan nilai prediksi. Hal ini menunjukkan bahwa homoskedastisitas terjadi.

Transformasi

WLS

Mencari nilai bobot:

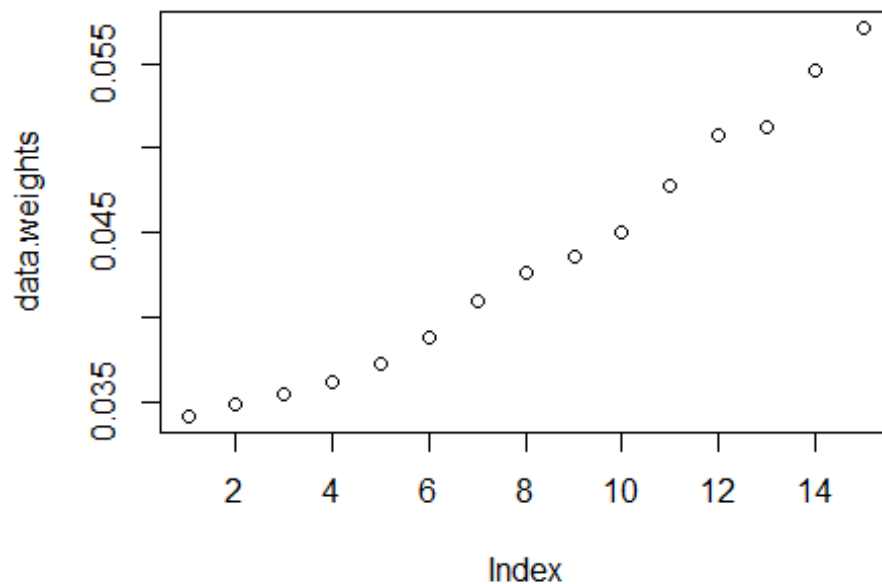
```

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 0.040910
34
##           8           9          10          11          12          13
14
## 0.04257072 0.04361593 0.04507050 0.04779711 0.05077885 0.05122749 0.054541
32
##          15
## 0.05710924

plot(data.weights)

```

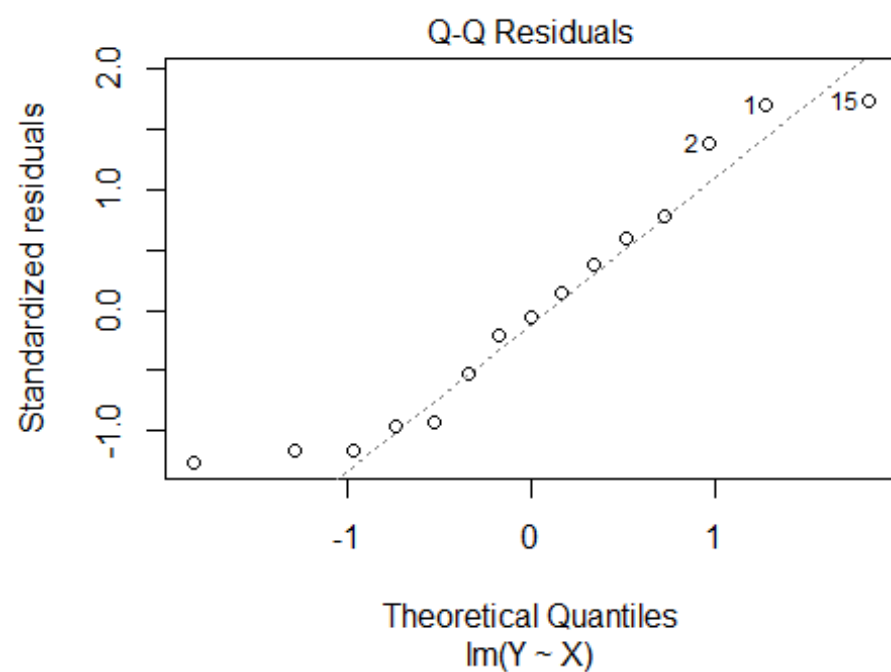
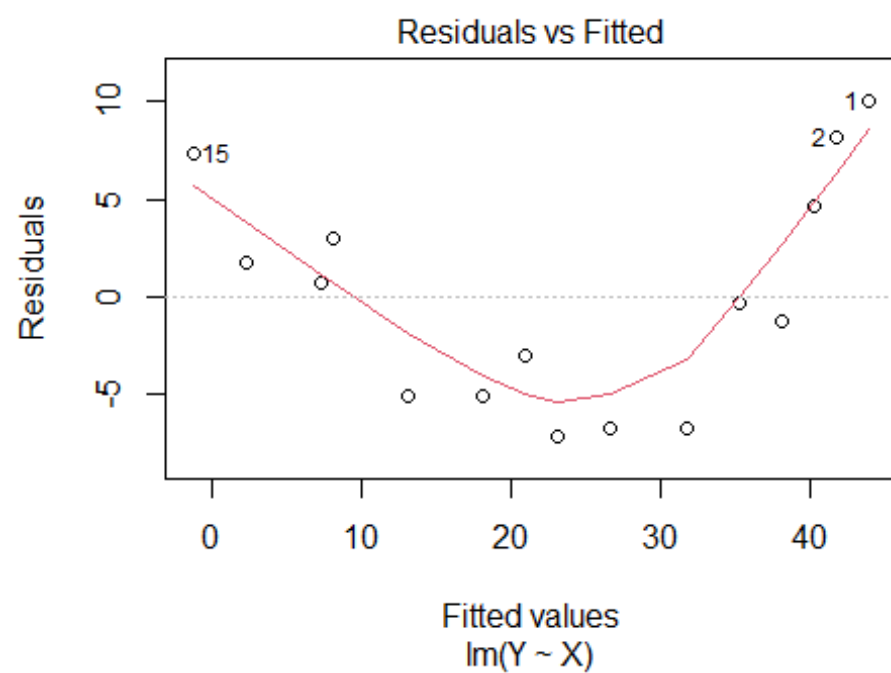


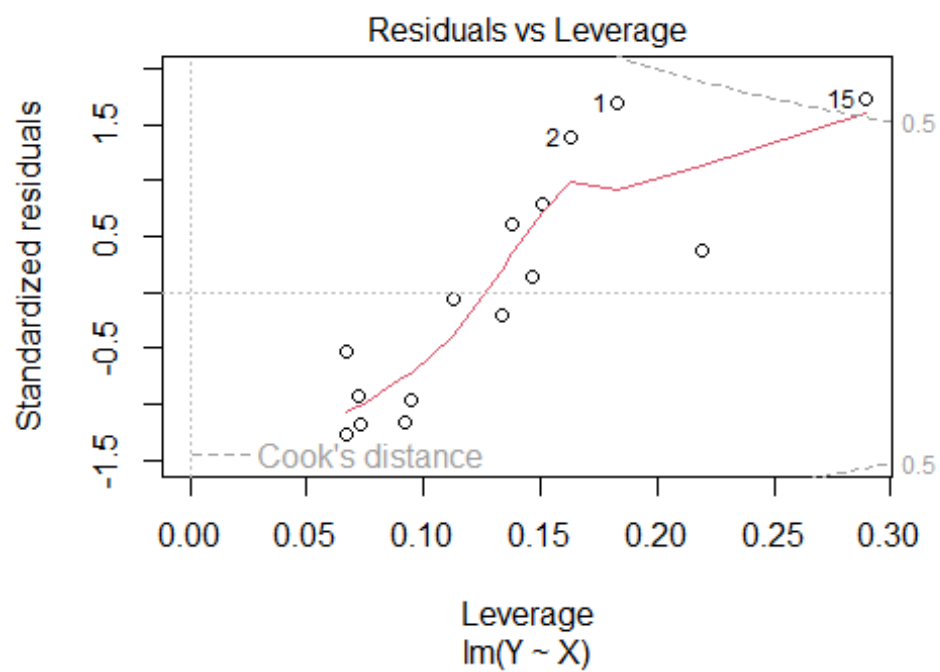
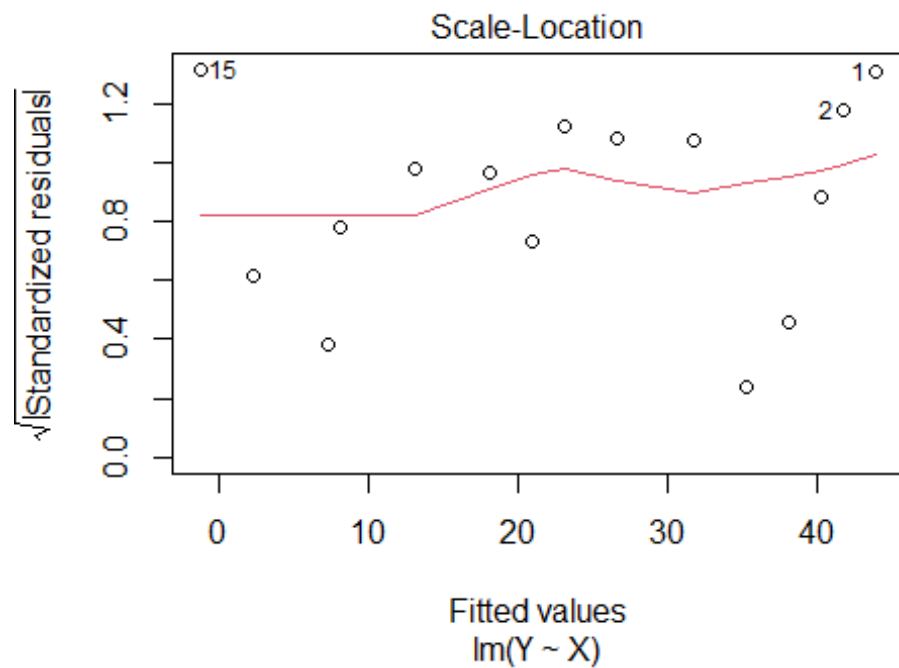
Hasil model regresi yang terboboti:

```

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

```





```
summary(model_weighted)
```

```
##
```

```
## Call:
```

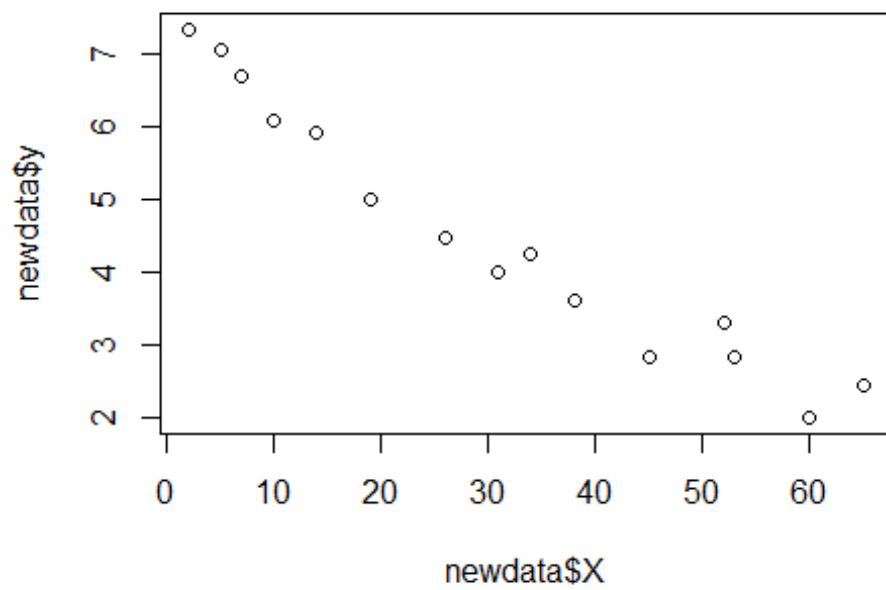
```
## lm(formula = Y ~ X, data = data, weights = data.weights)
##
## Weighted Residuals:
##      Min        1Q    Median        3Q        Max
## -1.46776 -1.09054 -0.06587  0.77203  1.85309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.41058    2.90674  15.623 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
```

Dari hasil transformasi WLS di atas dapat disimpulkan WLS belum efektif dalam mentransformasi model regresi dapat dibuktikan dari hasil eksplorasi di atas masih belum memenuhi asumsi Gauss-Markov

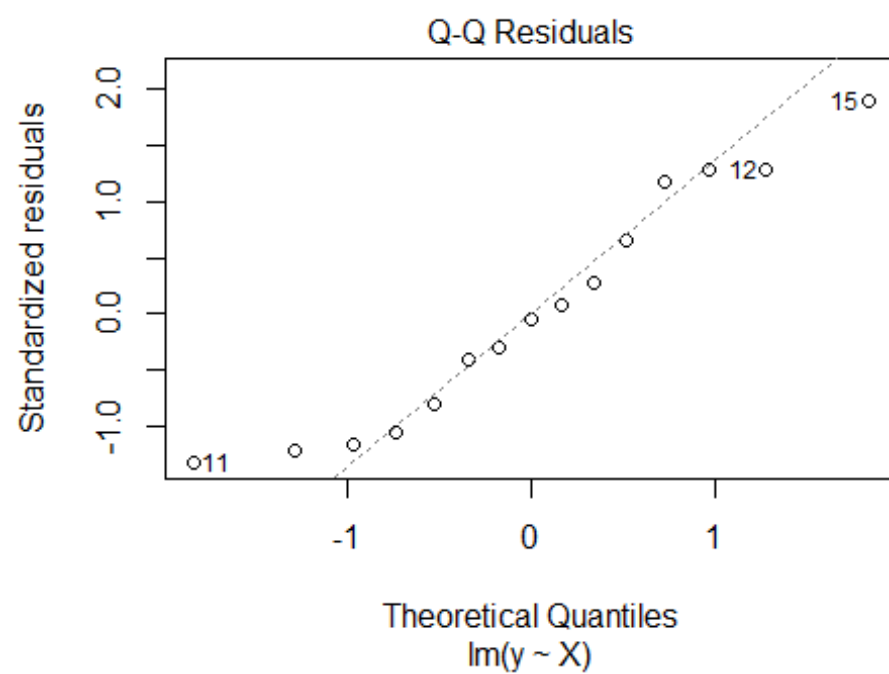
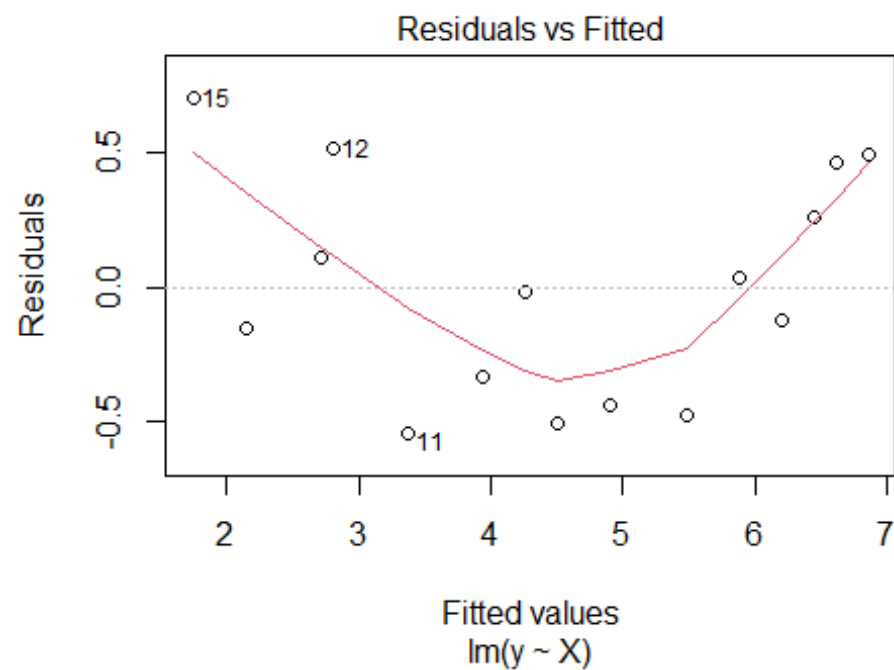
Transformasi Akar pada x, y, atau X dan y

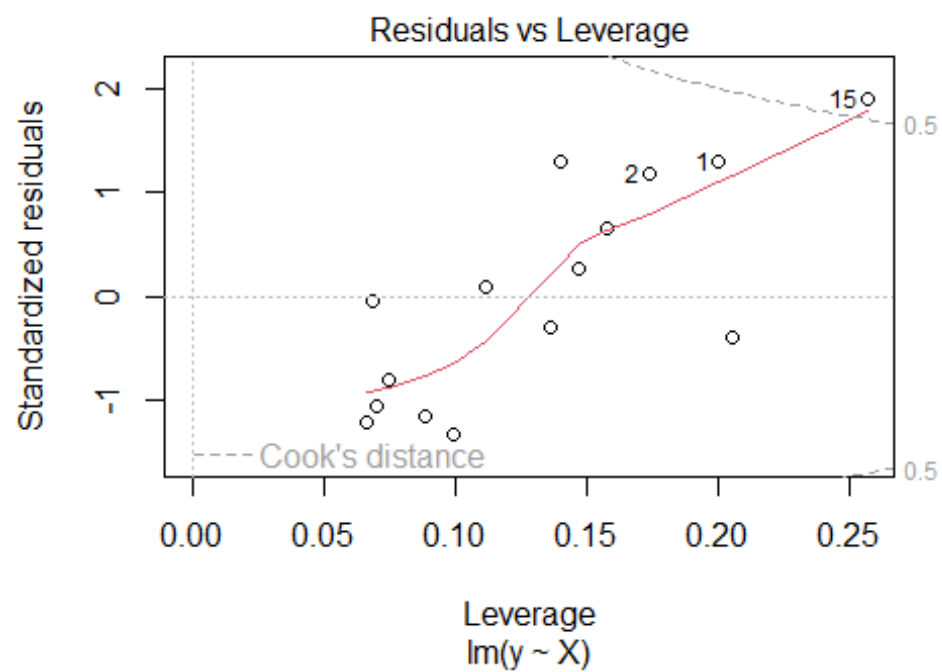
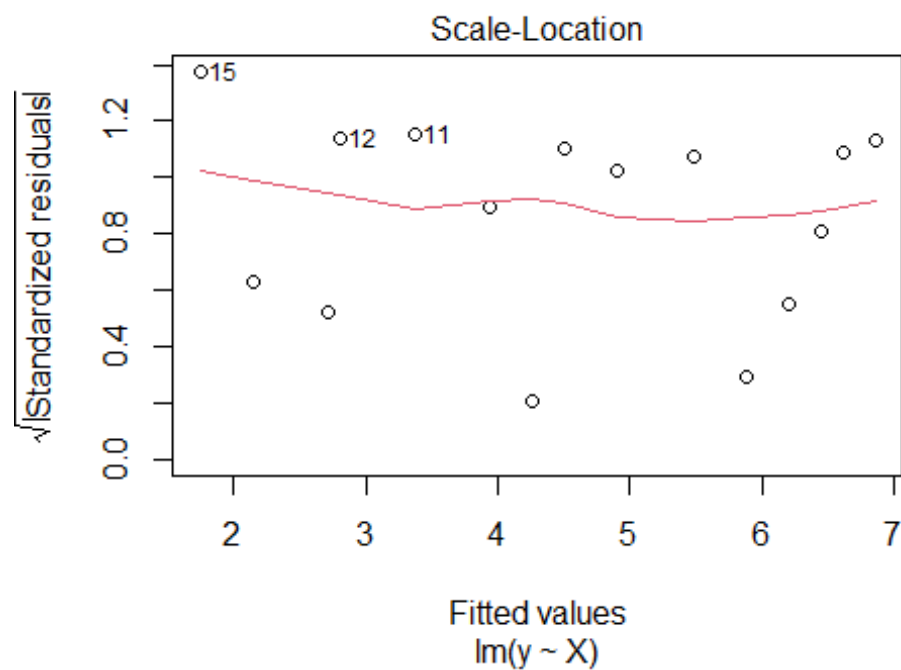
```
newdata <- data %>%
  mutate(y = sqrt(Y)) %>%
  mutate(x = sqrt(X))

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



```
plot(model_sqrtx)
```





```
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
```

Hasil transformasi pertama dengan mengakar kan variable dependen menghasilkan model regresi:

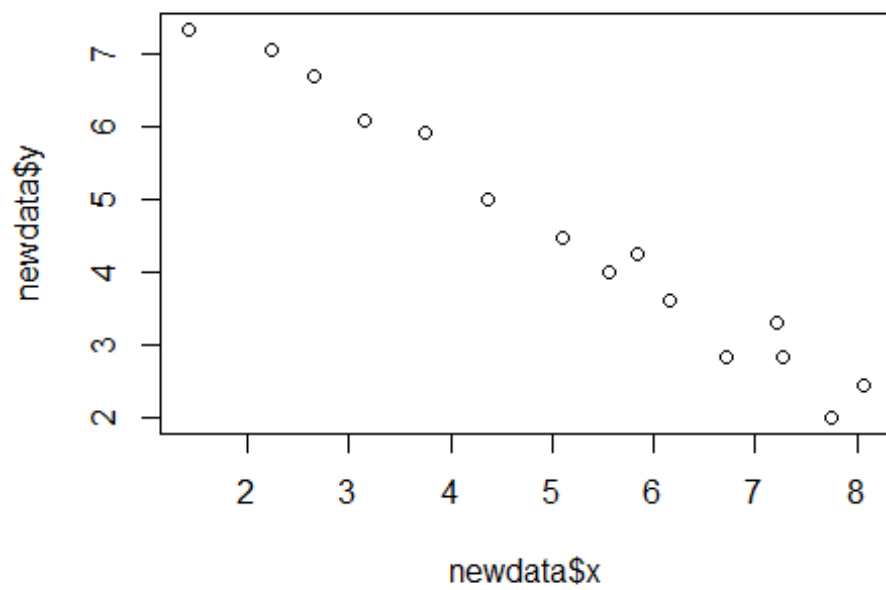
$$\hat{Y} = 7.015455 - 0.081045X$$

Uji Autokorelasi model regresi transformasi

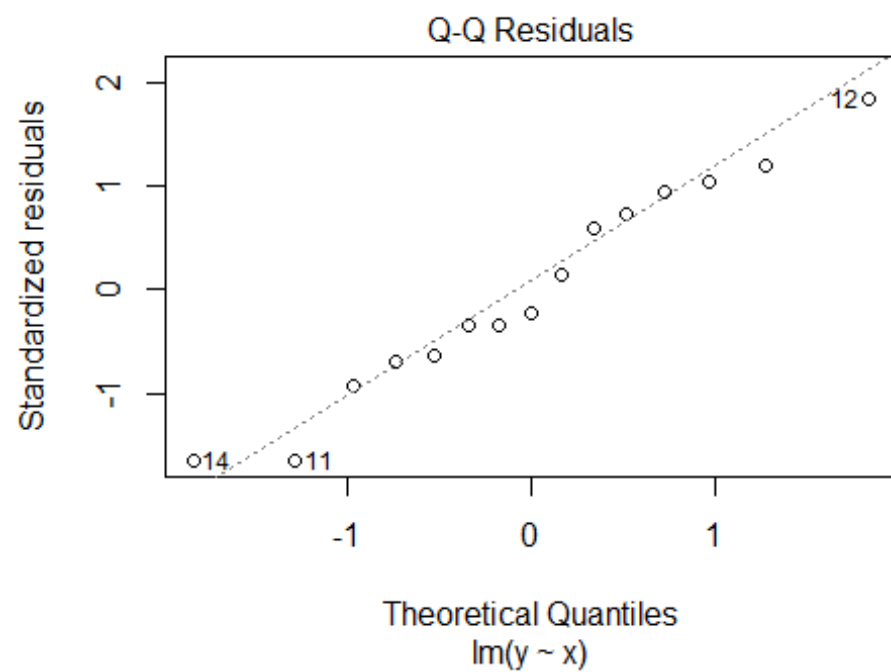
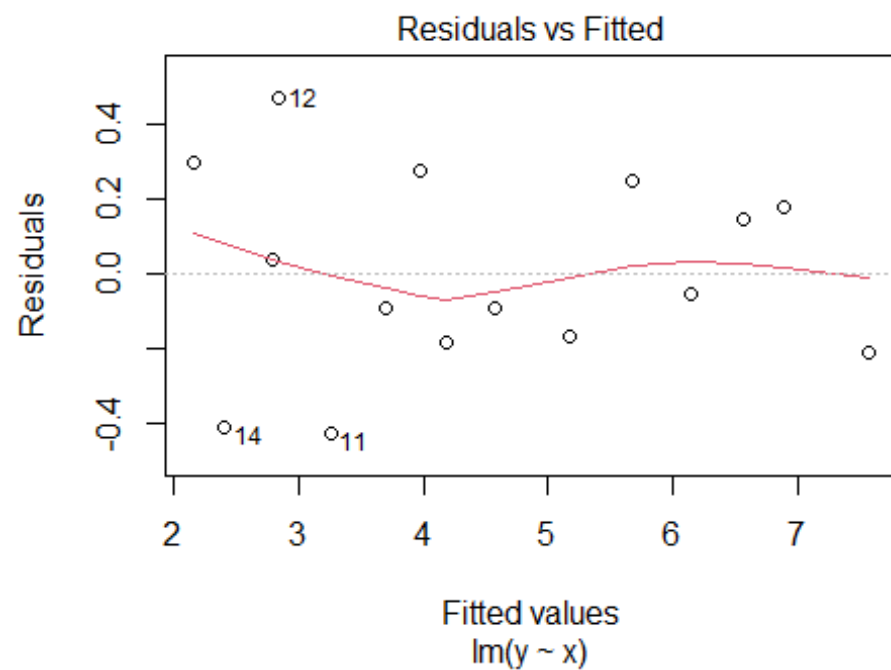
```
dwtest(model_sqrtx)
##
## Durbin-Watson test
##
## data: model_sqrtx
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0
```

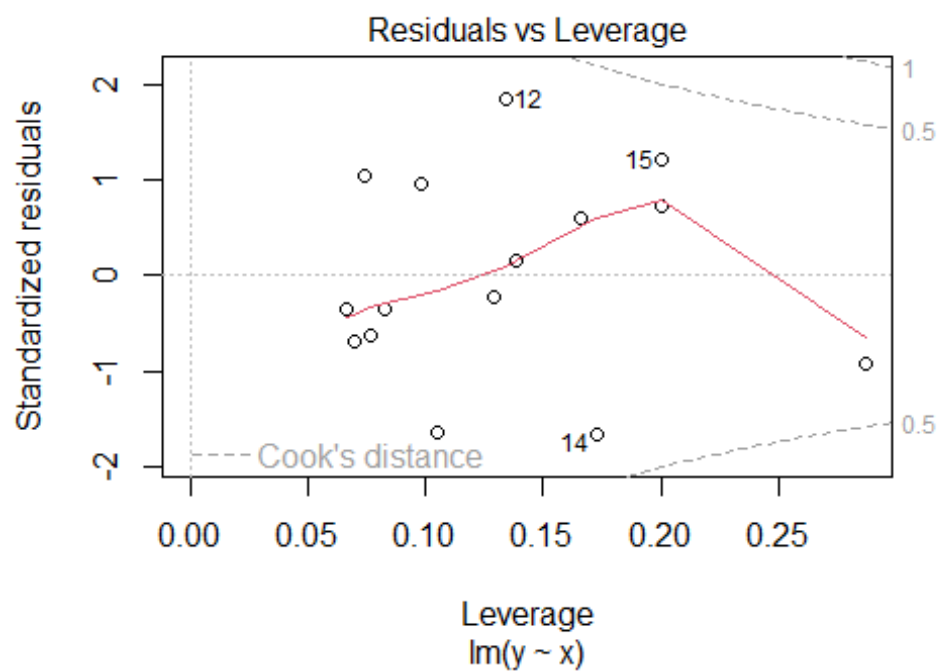
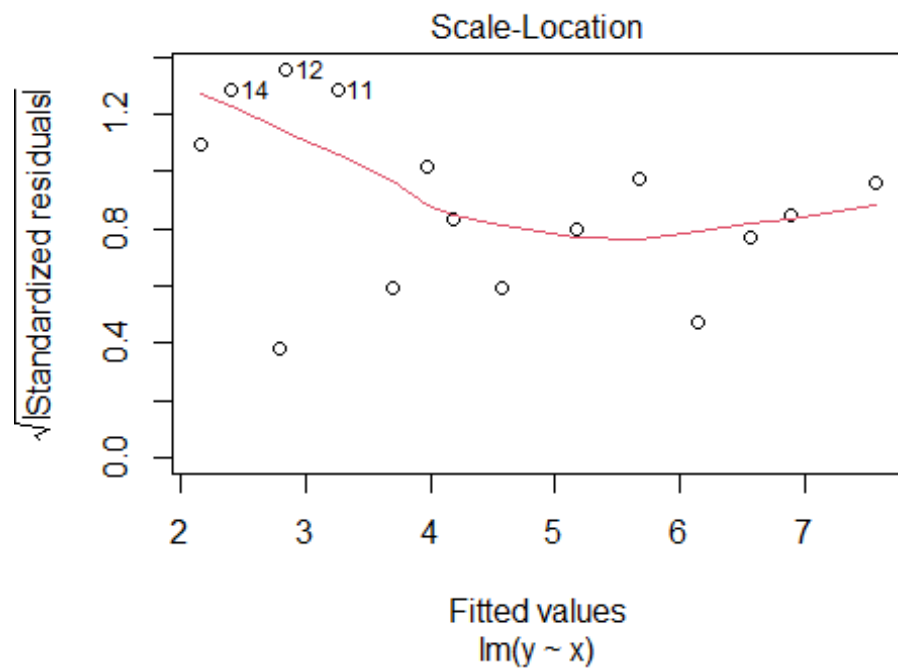
Dengan nilai DW yang rendah dan p-value yang signifikan, hasil tes Durbin-Watson ini menunjukkan adanya autokorelasi positif. uji Durbin-Watson di atas terbukti masih adanya autokorelasi yang dibuktikan p-value yang kurang dari 0,05

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



```
plot(model_sqrt)
```





```
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
```

Hasil transformasi pertama dengan mengakar kan variable dependen dan juga independennya menghasilkan model regresi:

$$\hat{Y} = 8.71245 - 0.81339X$$

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
```

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 transformasi:

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

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

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

Dilakukan Transformasi balik menjadi:

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

Interpretasi Model ini menunjukkan bahwa Y berbanding terbalik dengan \sqrt{X} , dengan hubungan kuadratik. Semakin besar nilai \sqrt{X} , 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 \sqrt{X} . Semakin besar nilai \sqrt{X} , 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.