

Tugas Individu Analisis Regresi

Biki Nurul Afida (G1401221043)

2024-03-06

Package installation

```
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 'tibble' was built under R version 4.3.2
## Warning: package 'tidyr' was built under R version 4.3.2
## Warning: package 'readr' was built under R version 4.3.2
## Warning: package 'purrr' 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    1.5.1
## ✓ 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()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(ggribes)
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(lmtest)

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

## 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(dplyr)
library(car)

## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:dplyr':
##
##   recode
##
## The following object is masked from 'package:purrr':
##
##   some
```

```
library(nortest)
library(randtests)
library(stats)
```

Data

```
data <- read_excel("D:/anreg/Data Tugas Pekan 7.xlsx")
data

## # A tibble: 15 × 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
```

Model Regresi

```
model.reg= lm(formula = Y ~ X, data)
model.reg

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

Diperoleh model regresi berikut

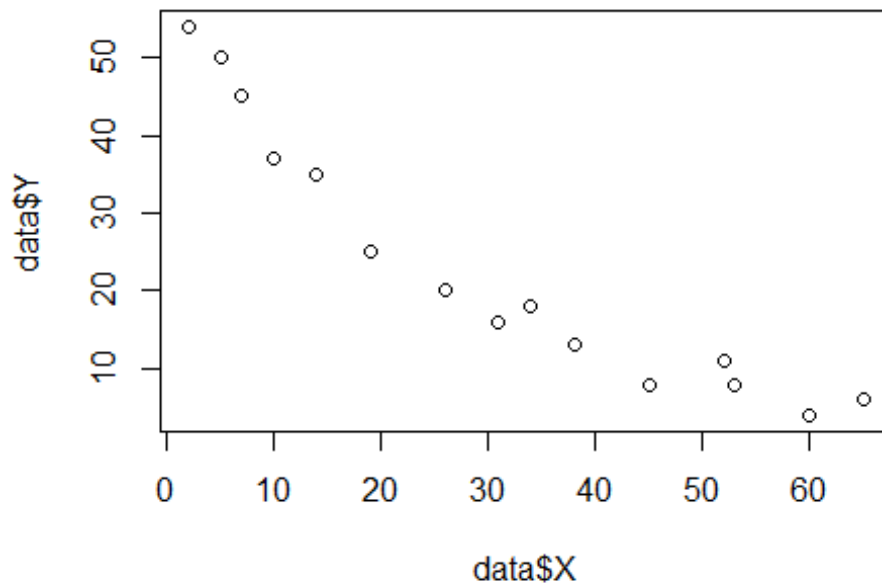
$$\hat{Y} = 46.4604 - 0.7525X + e$$

Namun, model regresi tersebut belum dapat dipastikan merupakan model yang terbaik sehingga perlu dilakukan eksplorasi data dan juga serangkaian pemeriksaan asumsi Gauss-Markov serta uji normalitas agar menghasilkan model terbaik. Asumsi tentang bentuk model adalah regresi linear sederhana -> Plot antara X dan Y.

Eksplorasi Data

Scatter Plot

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



Berdasarkan scatter plot di atas menunjukkan jika X dan Y tidak berhubungan linier.

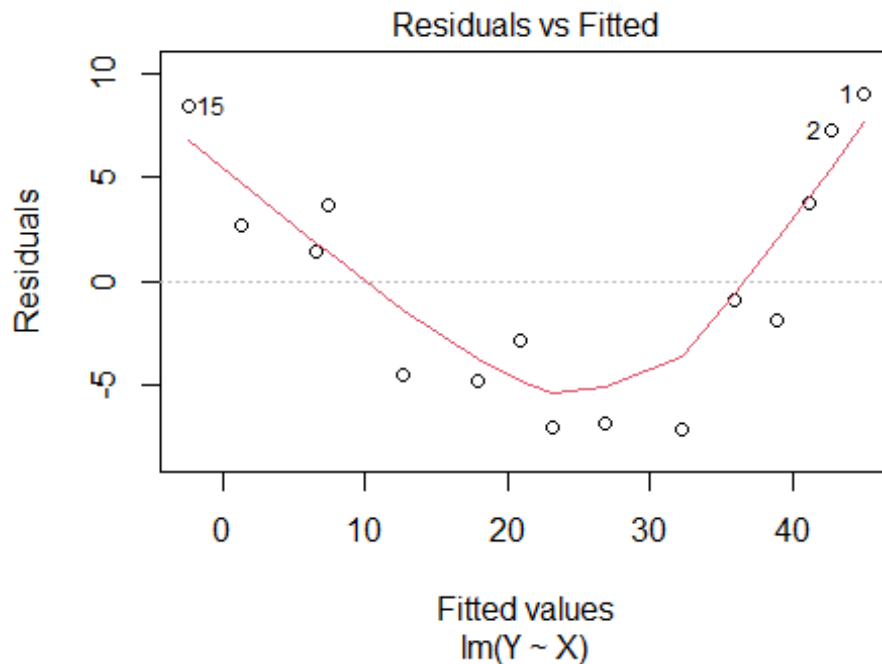
Pemeriksaan Asumsi

Eksplorasi Kondisi Gauss-Markov

Pemeriksaan $\varepsilon(i \sim N)$ dan $E[\varepsilon i] = 0$

sisaan vs yduga

```
plot(model.reg, 1)
```



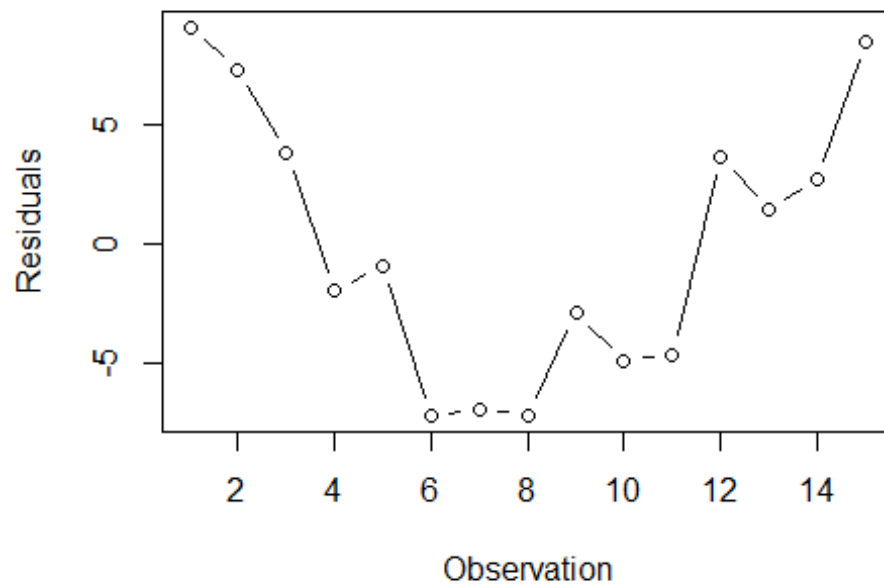
Dari plot di atas dapat dilihat bahwa bentuk plot sisaan vs yduga sisaan berada di sekitar 0, artinya nilai harapan galat sama dengan nol dan lebar pita sama untuk setiap nilai dugaan (ragam homogen). Bentuk pola adalah pola kurva artinya model tidak pas (memerlukan suku-suku lain dalam model atau transformasi terhadap Y).

Pemeriksaan $var(\epsilon_i) = \sigma^2$ (ragam homogen)

Plot Sisaan vs Urutan

Cek apakah sisaan saling bebas

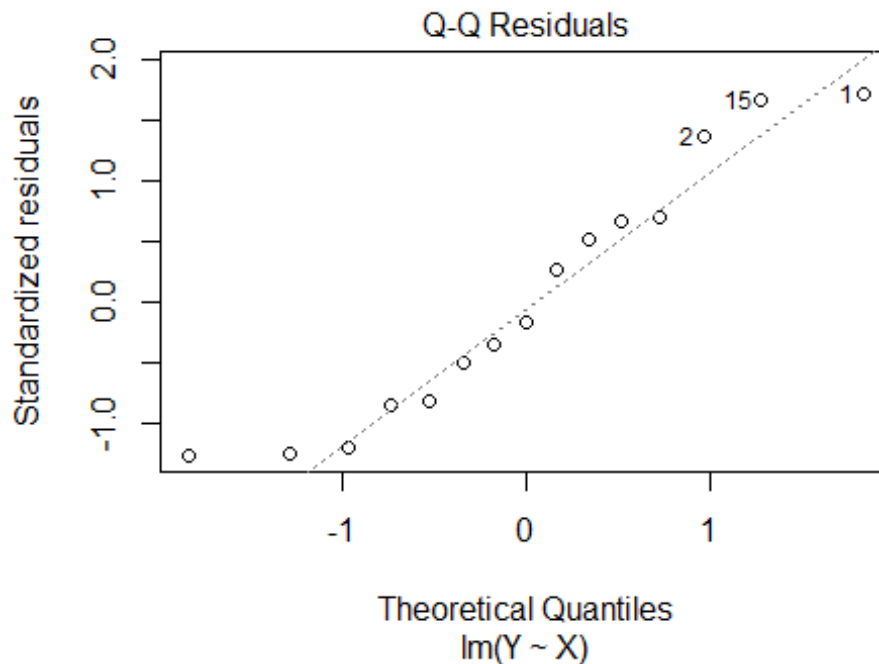
```
plot(x = 1:dim(data)[1],
     y = model.reg$residuals,
     type = 'b',
     ylab = "Residuals",
     xlab = "Observation")
```



Berdasarkan plot di atas, tebaran berpola sehingga sisaan tidak saling bebas dan model belum tepat (belum pas).

Eksplorasi Normalitas Sisaan – qq-plot

```
plot(model.reg,2)
```



Dari qq-plot di atas dapat dilihat bahwa sisaan menyebar normal.

Uji Formal Kondisi Gauss-Markov

Pada uji formal asumsi ini, diharapkan nilai p-value > 0.05 dengan kesimpulan tak tolak H_0

1. Nilai harapan sisaan sama dengan nol Hipotesis H_0 : \text{Nilai harapan sisaan sama dengan nol} \\ H_1: \text{Nilai harapan tidak sama dengan nol}

```
t.test(model.reg$residuals, mu = 0, conf.level = 0.95)
```

```
##
## One Sample t-test
##
## data: model.reg$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
```

Dari uji t-test di atas didapatkan bahwa p-value > α , maka pada taraf nyata 5% tak tolak H_0 , artinya nilai harapan sisaan sama dengan nol.

2. Ragam sisaan homogen Hipotesis $H_0: \text{Ragam sisaan homogen}$ $H_1: \text{Ragam sisaan tidak homogen}$

```
bptest(model.reg)
```

```
##  
## studentized Breusch-Pagan test  
##  
## data: model.reg  
## BP = 0.52819, df = 1, p-value = 0.4674
```

Berdasarkan Breusch-Pagan test didapatkan bahwa nilai p-value $> \alpha$, maka pada taraf nyata 5% tak tolak H_0 , artinya ragam sisaan homogen.

3. Sisaan saling bebas Hipotesis $H_0: \text{Sisaan saling bebas}$ $H_1: \text{Sisaan tidak saling bebas}$

```
dwtest(model.reg)
```

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

Berdasarkan Durbin-Watson test didapatkan bahwa p-value $< \alpha$, maka pada taraf nyata 5% Tolak H_0 , artinya sisaan tidak saling bebas atau dengan kata lain tidak ada autokorelasi.

Uji Formal Normalitas Sisaan Hipotesis $H_0: \text{Sisaan menyebar normal}$ $H_1: \text{Sisaan tidak menyebar normal}$

```
sisaanmodel.reg <- resid(model.reg)  
(normmodel.reg <- lillie.test(sisaanmodel.reg))
```

```
##  
## Lilliefors (Kolmogorov-Smirnov) normality test  
##  
## data: sisaanmodel.reg  
## D = 0.12432, p-value = 0.7701
```

Berdasarkan Lilliefors (Kolmogorov-Smirnov) normality test didapatkan bahwa p-value $> \alpha$, maka pada taraf nyata 5% tak tolak H_0 , artinya sisaan menyebar normal.

Berdasarkan serangkaian pemeriksaan asumsi dan uji normalitas, didapatkan bahwa sisaan tidak saling bebas dan hal tersebut melanggar asumsi yang seharusnya sisaan saling bebas. Oleh karena itu perlu dilakukan transformasi agar asumsi tersebut dapat terpenuhi.

Penanganan sisaan tidak saling bebas

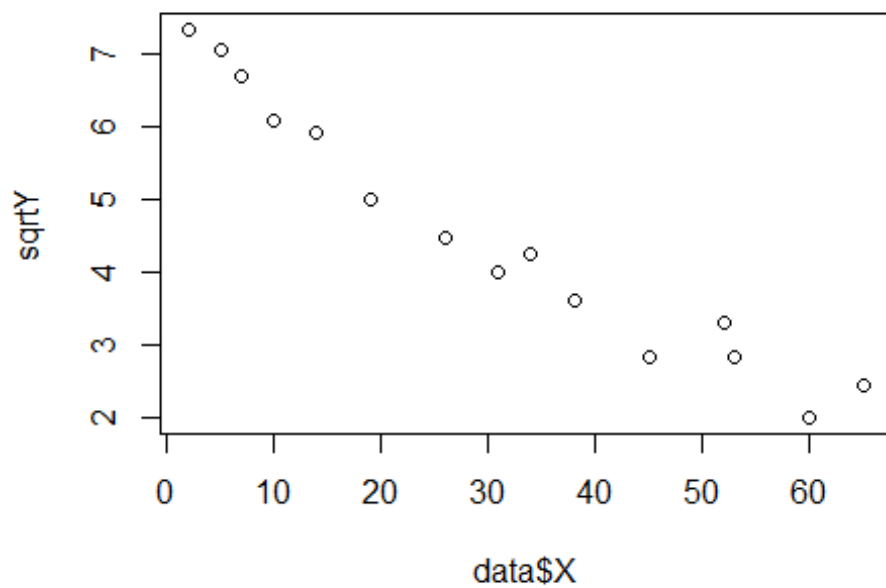
data transformasi

```
sqrtY <- sqrt(data$Y)
sqrtX <- sqrt(data$X)
data2 <- data.frame(data$X, data$Y, sqrtY, sqrtX)
data2
```

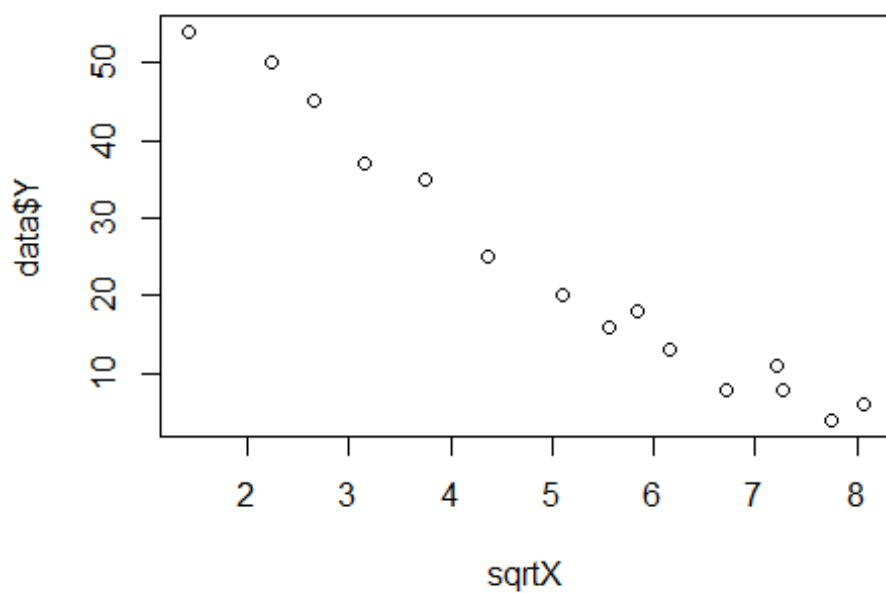
##	data.X	data.Y	sqrtY	sqrtX
## 1	2	54	7.348469	1.414214
## 2	5	50	7.071068	2.236068
## 3	7	45	6.708204	2.645751
## 4	10	37	6.082763	3.162278
## 5	14	35	5.916080	3.741657
## 6	19	25	5.000000	4.358899
## 7	26	20	4.472136	5.099020
## 8	31	16	4.000000	5.567764
## 9	34	18	4.242641	5.830952
## 10	38	13	3.605551	6.164414
## 11	45	8	2.828427	6.708204
## 12	52	11	3.316625	7.211103
## 13	53	8	2.828427	7.280110
## 14	60	4	2.000000	7.745967
## 15	65	6	2.449490	8.062258

Transformasi Data

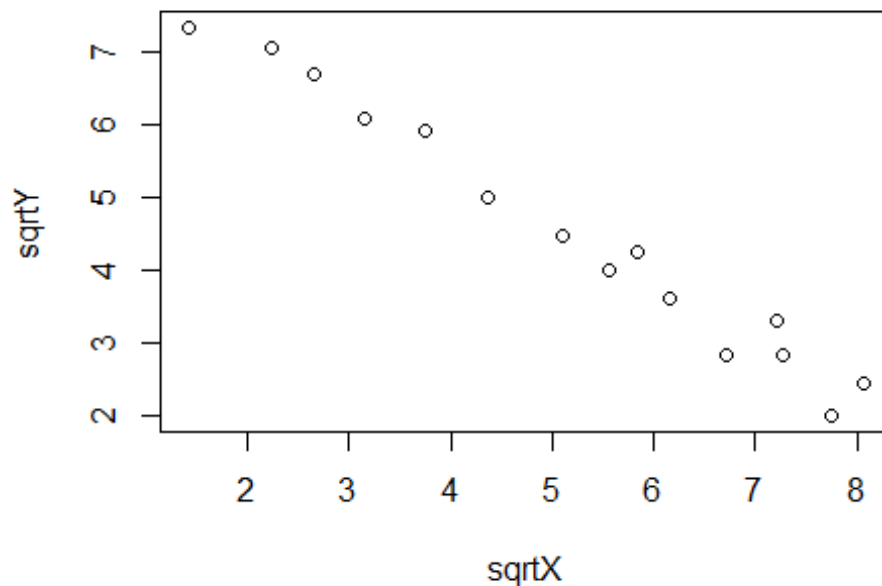
```
plot(x = data$X, y = sqrtY)
```



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



```
plot(x = sqrtX,y = sqrtY)
```



```
data.baru <- data.frame(sqrtX, sqrtY)
```

```
data.baru
```

```
##      sqrtX    sqrtY
## 1  1.414214  7.348469
## 2  2.236068  7.071068
## 3  2.645751  6.708204
## 4  3.162278  6.082763
## 5  3.741657  5.916080
## 6  4.358899  5.000000
## 7  5.099020  4.472136
## 8  5.567764  4.000000
## 9  5.830952  4.242641
## 10 6.164414  3.605551
## 11 6.708204  2.828427
## 12 7.211103  3.316625
## 13 7.280110  2.828427
## 14 7.745967  2.000000
## 15 8.062258  2.449490
```

Berdasarkan plot di atas X dan Y memiliki hubungan yang cenderung membentuk sebuah parabola dan nilai $\beta_1 < 0$, sehingga data dapat ditransformasi dengan memperkecil nilai X dan nilai Y dengan membentuknya menjadi akar dari data aslinya. Ada perbedaan dari hasil plot hubungan \sqrt{X} dengan Y, X dengan \sqrt{Y} , dan juga \sqrt{X} dengan \sqrt{Y} . Oleh karena itu, perlu ditelusuri lebih dalam agar menghasilkan model terbaik berdasarkan

pemeriksaan asumsi pada data dengan sisaan saling bebas atau dengan kata lain data memiliki autokorelasi.

Model dan Pemeriksaan Asumsi sqrtX dengan Y

```
model.reg1 = lm(formula = data$Y ~ sqrtX)
summary(model.reg1)

##
## Call:
## lm(formula = data$Y ~ sqrtX)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.4518 -2.8559  0.7657  2.0035  5.2422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.2250     2.2712   27.84 5.67e-13 ***
## sqrtX        -7.7481     0.4097  -18.91 7.68e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.262 on 13 degrees of freedom
## Multiple R-squared:  0.9649, Adjusted R-squared:  0.9622
## F-statistic: 357.7 on 1 and 13 DF, p-value: 7.684e-11
```

Diperoleh model regresi berikut

$$\hat{Y} = 63.2250 - 7.7481X + e$$

```
dwtest(model.reg1)

##
## Durbin-Watson test
##
## data: model.reg1
## DW = 1.1236, p-value = 0.01422
## alternative hypothesis: true autocorrelation is greater than 0
```

Berdasarkan Durbin-Watson test didapatkan $p\text{-value} < \alpha$, maka tolak H_0 pada taraf nyata 5% artinya sisaan tidak saling bebas sehingga asumsi tidak terpenuhi, bukan model terbaik.

X dengan sqrtY

```
model.reg2 = lm(formula = sqrtY ~ data$X)
summary(model.reg2)

##
## Call:
## lm(formula = sqrtY ~ data$X)
```

```
##
## 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 ***
## data$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
```

Diperoleh model regresi berikut

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

```
dwtest(model.reg2)
```

```
##
## Durbin-Watson test
##
## data:  model.reg2
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0
```

Berdasarkan Durbin-Watson test didapatkan $p\text{-value} < \alpha$, maka tolak H_0 pada taraf nyata 5% artinya sisaan tidak saling bebas sehingga asumsi tidak terpenuhi, bukan model terbaik.

sqrtX dengan sqrtY

```
model.reg3 = lm(formula = sqrtY ~ sqrtX)
summary(model.reg3)
```

```
##
## Call:
## lm(formula = sqrtY ~ sqrtX)
##
## 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 ***
## sqrtX        -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
```

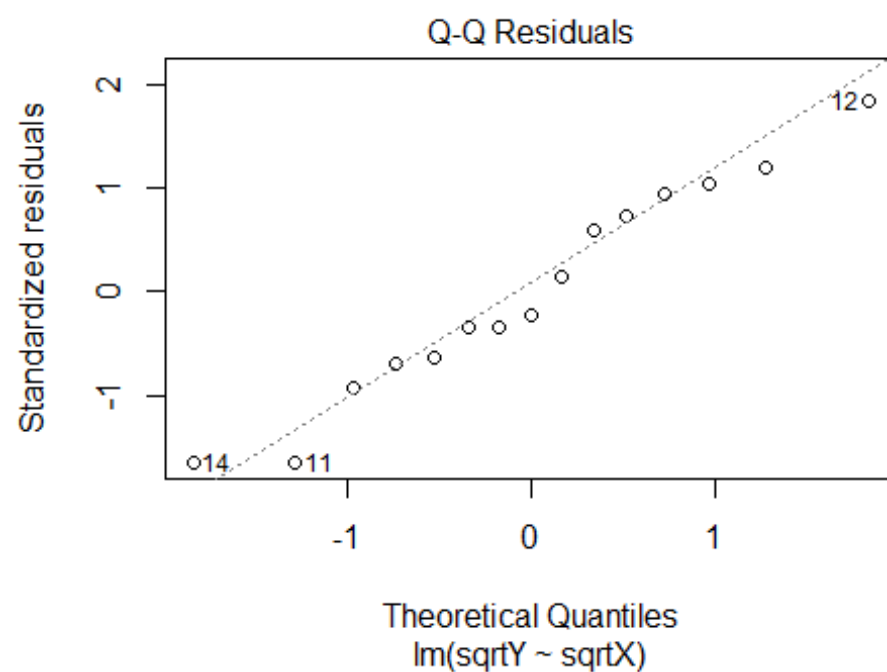
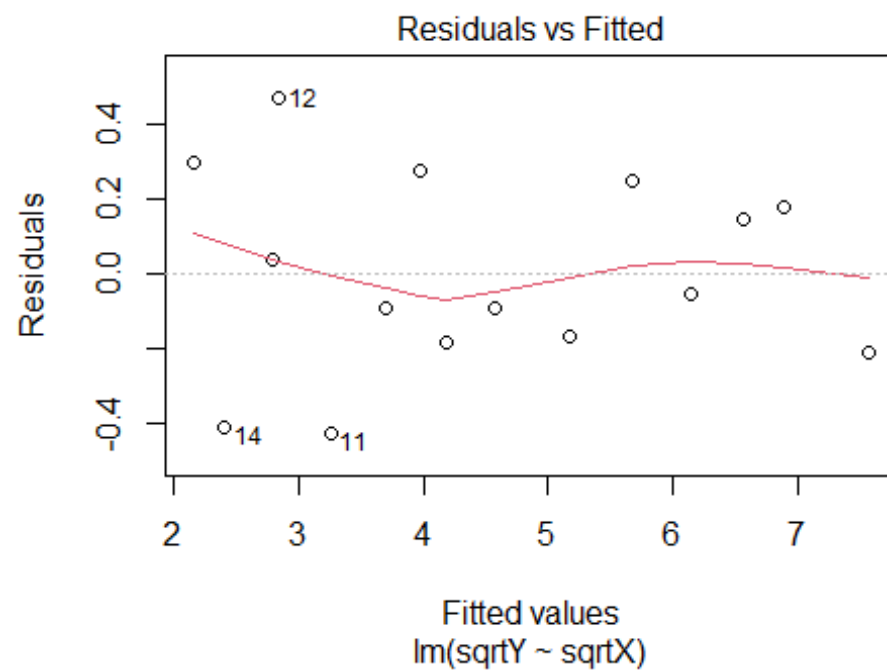
Diperoleh model regresi berikut

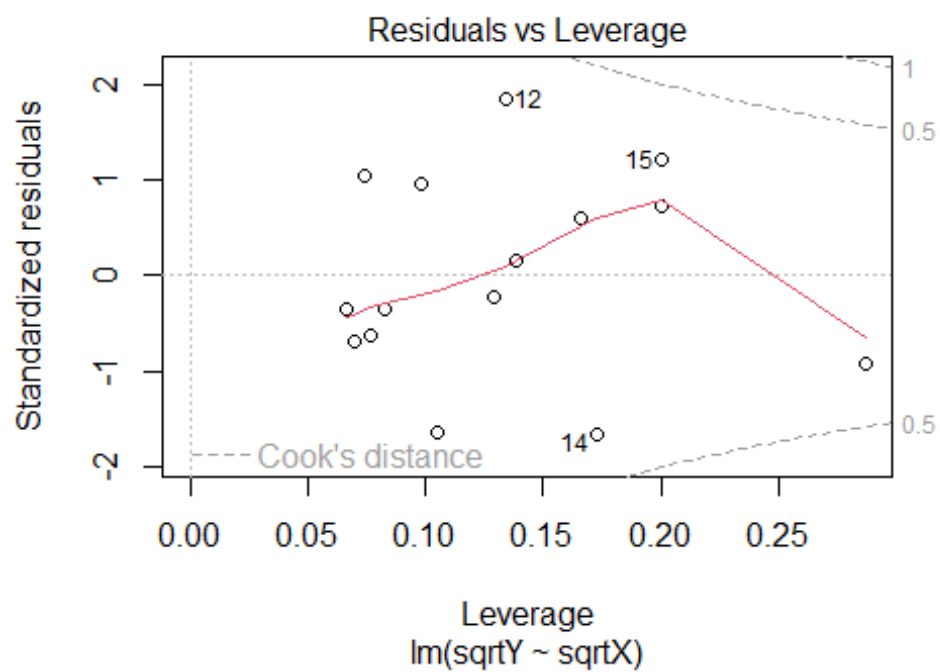
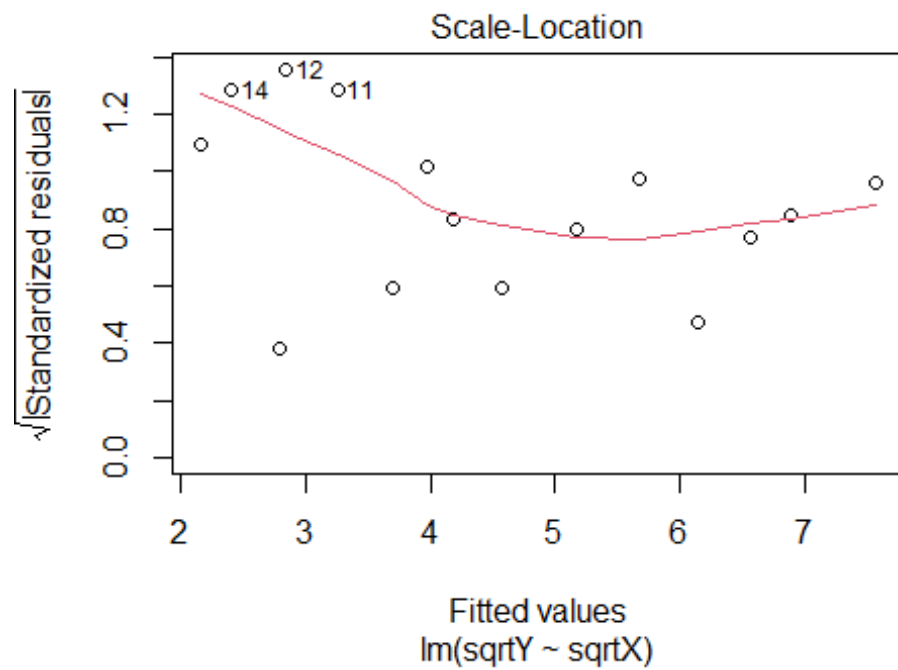
$$\hat{Y} = 8.71245 - 0.81339X + e$$

```
dwtest(model.reg3)
##
## Durbin-Watson test
##
## data:  model.reg3
## DW = 2.6803, p-value = 0.8629
## alternative hypothesis: true autocorrelation is greater than 0
```

Berdasarkan Durbin-Watson test didapatkan p-value $> \alpha$, maka tak tolak H_0 pada taraf nyata 5% artinya sisaan saling bebas sehingga asumsi terpenuhi. Namun, perlu dilakukan pemeriksaan asumsi yang lain untuk memastikan bahwa model tersebut adalah model terbaik.

```
plot(model.reg3)
```





```
t.test(model.reg3$residuals,mu = 0,conf.level = 0.95)
```

```
##
```

```
## One Sample t-test
```



```
##
## data: model.reg3$residuals
## t = 2.0334e-16, df = 14, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -0.1463783 0.1463783
## sample estimates:
## mean of x
## 1.387779e-17

ncvTest(model.reg3)

## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 2.160411, Df = 1, p = 0.14161

sisaanmodel.reg3 <- resid(model.reg3)
(normmodel.reg3 <- lillie.test(sisaanmodel.reg3))

##
## Lilliefors (Kolmogorov-Smirnov) normality test
##
## data: sisaanmodel.reg3
## D = 0.11948, p-value = 0.817
```

Berdasarkan semua uji yang telah dilakukan $p\text{-value} > \alpha$ maka semua asumsi sudah terpenuhi pada taraf nyata 5%.

Kesimpulan dan Transformasi Balik

Oleh karena itu, didapatkan model terbaik dari hasil transformasi X dan Y menjadi \sqrt{X} dan \sqrt{Y} sehingga memenuhi asumsi dalam analisis regresi sederhana. Sehingga didapatkan model regresi berikut $\hat{Y}^{\frac{1}{2}} = 8.71245 - 0.81339X^{\frac{1}{2}} + e$

Untuk mengembalikan model agar dapat menjelaskan peubah respons sebelum transformasi maka perlu dilakukan transformasi kembali. Langkah transformasi yaitu dengan mengkuadratkan kedua ruas sehingga didapatkan model terbaik sebagai berikut

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