

Analisis Regresi Kuliah-7

Muhammad Firlan Maulana

2024-03-05

Packages

```
library(readxl)
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(car)
```

```
## Loading required package: carData
```

```
library(randtests)
```

```
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    1.5.1
```

```
## 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()
## x dplyr::recode() masks car::recode()
## x purrr::some() masks car::some()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
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(nortest)
```

Data

```
data7 = read_xlsx("D:/firlan/Documents/College/Semester 4/Analisis Regeresi/Kuliah 7/Anreg Individu.xlsx")
head(data7)
```

```
## # A tibble: 6 x 2
##       X     Y
##   <dbl> <dbl>
## 1     2    54
## 2     5    50
## 3     7    45
## 4    10    37
## 5    14    35
## 6    19    25
```

Model Regresi Awal

```
model = lm(formula = Y ~ ., data=data7)
summary(model)
```

```
##
## Call:
## lm(formula = Y ~ ., data = data7)
##
## 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
```

Didapati bahwa model regresi awal sebagai berikut.

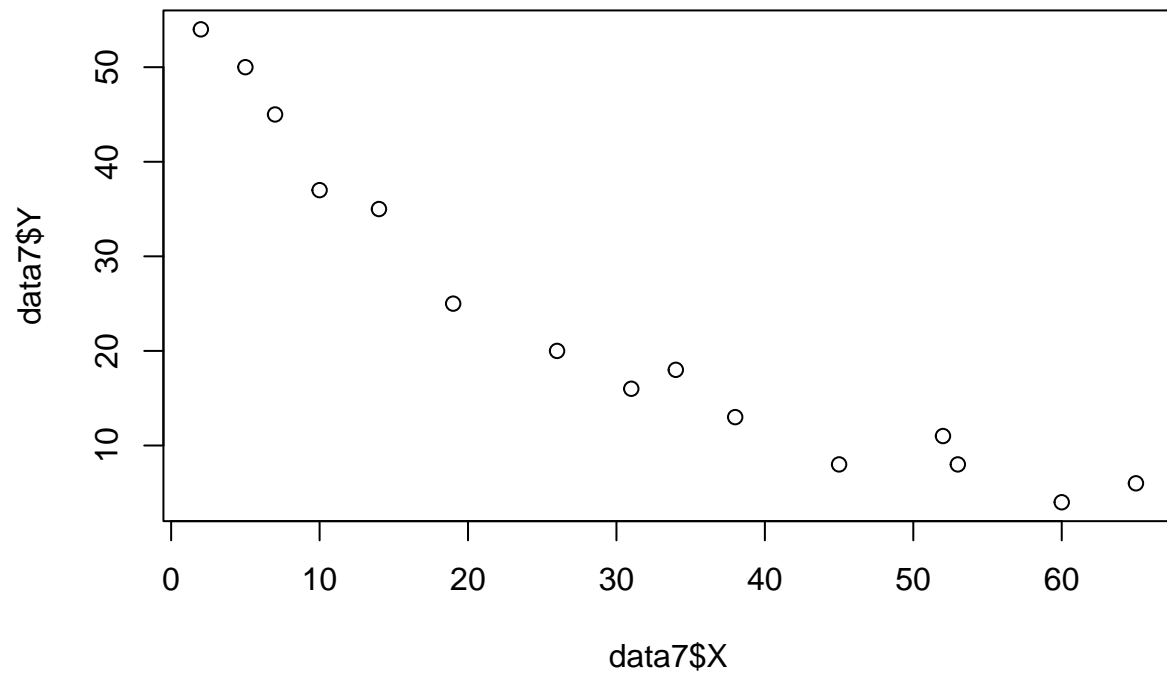
$$\hat{Y} = 46.46041 - 0.75251X + e$$

Masih belum diketahui apakah model regresi tersebut merupakan model regresi terbaik. Oleh karena itu, diperlukan eksplorasi kondisi termasuk pengujian asumsi Gauss-Marcov dan normalitas agar dapat diketahui model regresi terbaik.

Eksplorasi Kondisi

Scatter Plot Hubungan X dan Y

```
plot(x=data7$X, y=data7$Y)
```

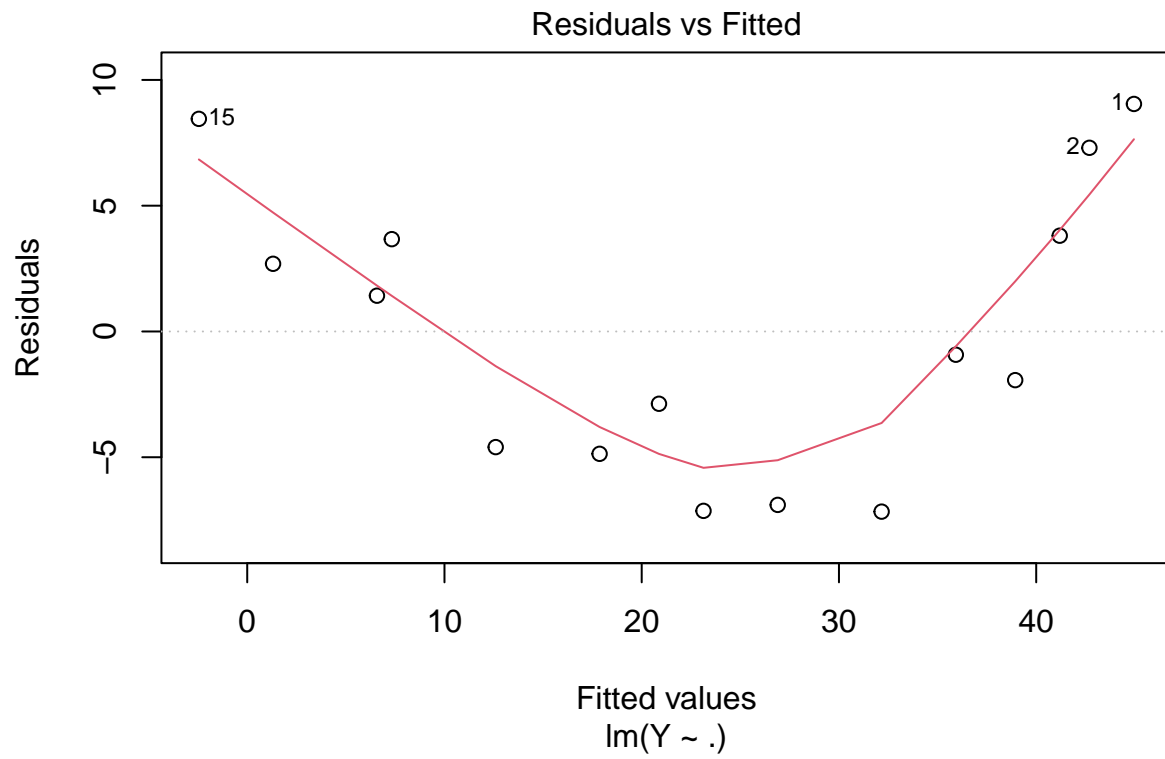


Dari scatter plot hubungan X dan Y di atas, dapat dilihat bahwa tidak terbentuk pola garis lurus namun berbentuk parabola.

Pemeriksaan Asumsi

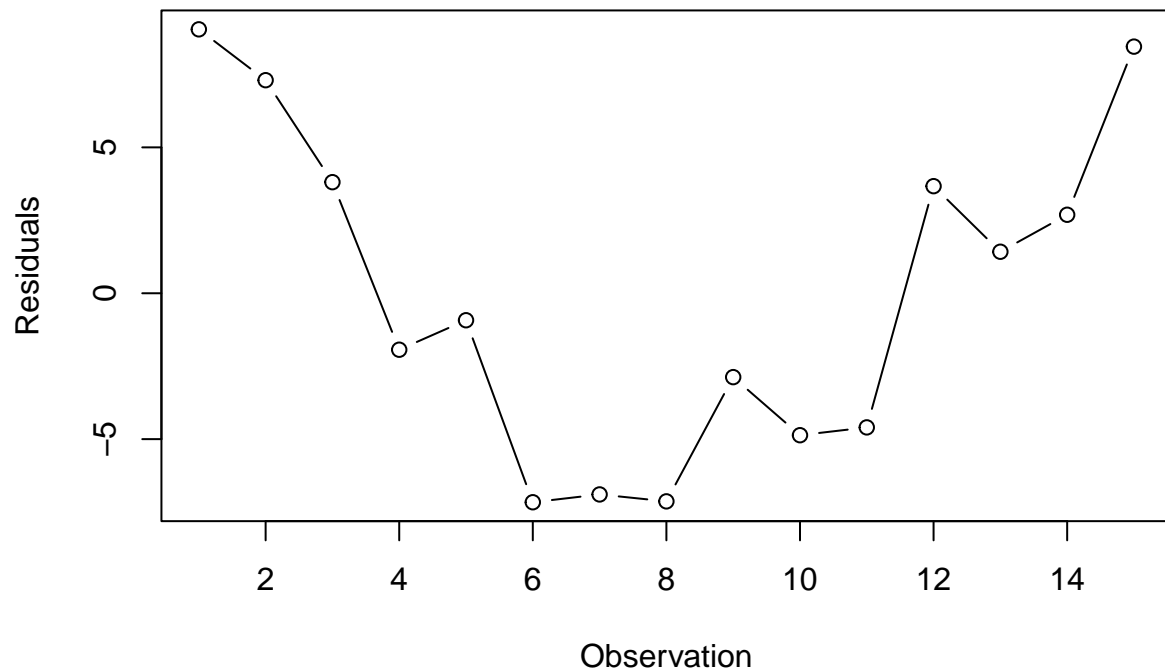
Plot sisaan VS Y duga

```
plot(model,1)
```



Plot sisaan VS Urutan

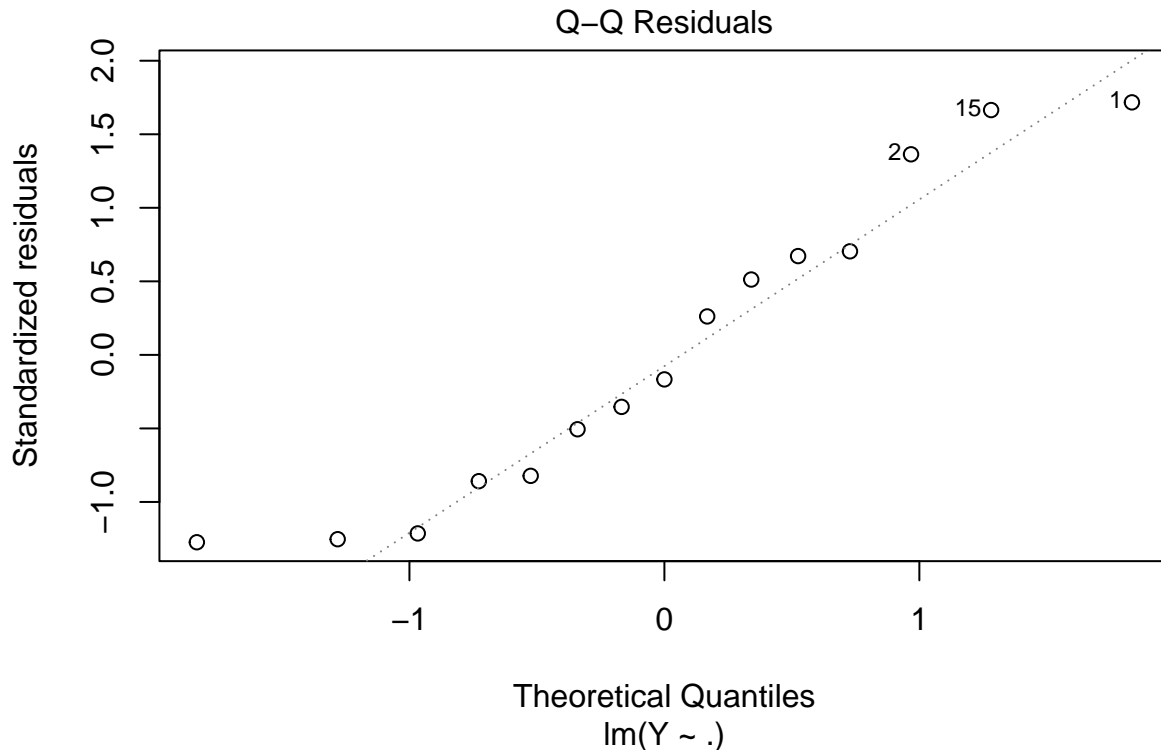
```
plot(x = 1:dim(data7)[1],
     y = model$residuals,
     type = 'b',
     ylab = "Residuals",
     xlab = "Observation")
```



Dari plot sisaan vs urutan di atas terlihat bahwa sisaan menyebar tidak berpola. Sehingga asumsi kondisi Gauss-Marcov mengenai sisaan saling bebas/tidak ada korelasi terpenuhi.

Eksplorasi Normalitas Sisaan dengan QQ-plot

```
plot(model,2)
```



Terlihat pada QQ-plot tersebut bahwa sisaan menyebar Normal sehingga memenuhi asumsi $e(i \sim N)$.

Uji Formal Kondisi Gauss-Marcov

1. Nilai harapan/rataan sisaan sama dengan nol

H_0 : Nilai harapan sisaan sama dengan 0 H_1 : Nilai harapan sisaan tidak sama dengan 0

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

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

$p\text{-value} = 1 > \alpha = 0.05$, sehingga tak tolak H_0 . Belum terdapat cukup bukti untuk menyatakan bahwa nilai harapan sisaan tidak sama dengan 0.

2. Ragam sisaan homogen

H0: Ragam sisaan homogen H1: Ragam sisaan tidak homogen

```
apakah.homogen = lm(formula = abs(model$residuals) ~ X, # y: abs residual
  data = data7)
summary(apakah.homogen)
```

```
##
## Call:
## lm(formula = abs(model$residuals) ~ X, data = data7)
##
## 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
```

```
bptest(model)
```

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

```
ncvTest(model)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.1962841, Df = 1, p = 0.65774
```

Didapati bahwa $p\text{-value} = 0.4674 > \alpha = 0.05$, sehingga tak tolak H0. Belum cukup bukti untuk menyatakan bahwa ragam sisaan tidak homogen.

3. Sisaan saling bebas/tidak ada autokorelasi

H0: Sisaan saling bebas/tidak ada autokorelasi H1: Sisaan tidak saling bebas/ada autokorelasi

```
runs.test(model$residuals)
```



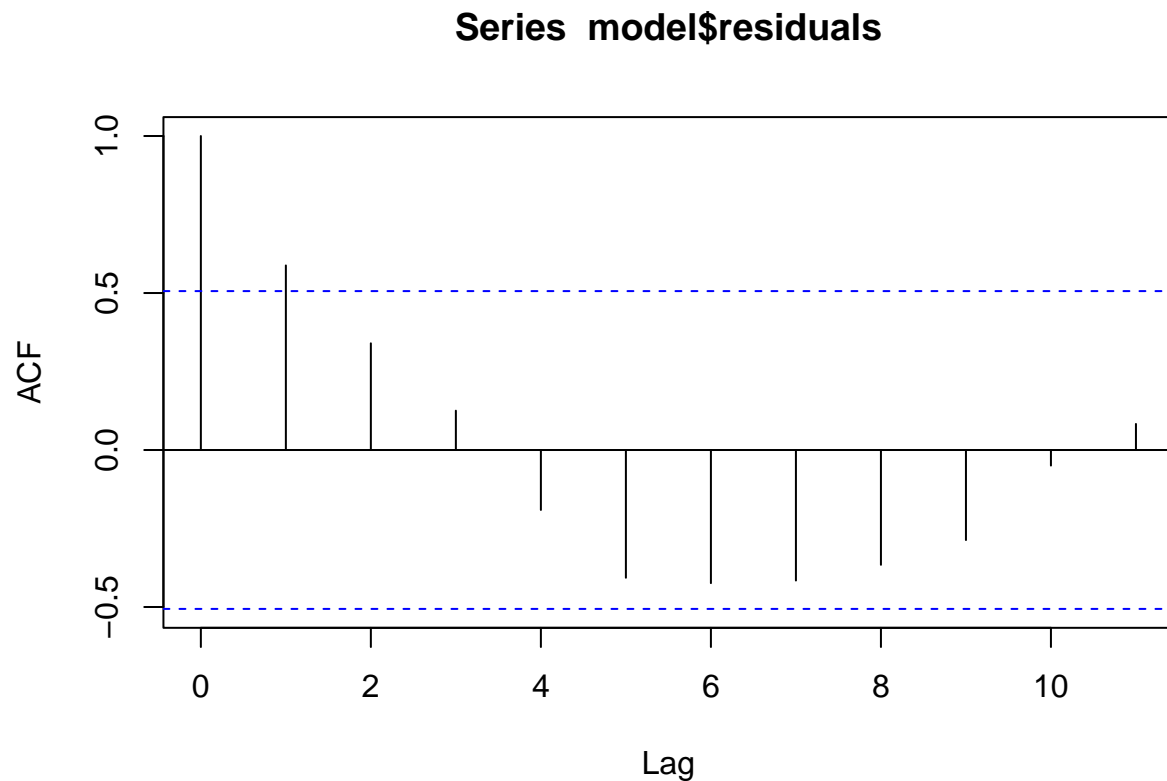
```
##
## Runs Test
##
## data: model$residuals
## statistic = -2.7817, runs = 3, n1 = 7, n2 = 7, n = 14, p-value =
## 0.005407
## alternative hypothesis: nonrandomness
```

```
dwtest(model)
```

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

Didapati dari Durbin-Watson bahwa $p\text{-value} = 0.00001333 < \alpha = 0.05$, maka tolak H_0 . Pada taraf nyata 5%, cukup bukti untuk menyatakan bahwa sisaan tidak saling bebas. Asumsi tidak terpenuhi.

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



Hasil uji Durbin-Watson selaras dengan eksplorasi sisaan di atas bahwa nilai autokorelasi pada lag 1 bernilai 0.5 dan lag 2 bernilai 0.4 yang mana berada di luar batas kepercayaan 95%, terdapat autokorelasi.

Uji Formal Normalitas Sisaan

H0: Sisaan menyebar Normal H1: Sisaan tidak menyebar Normal

```
ks.test(model$residuals, "pnorm", mean=mean(model$residuals), sd=sd(model$residuals))
```

```
##  
## Exact one-sample Kolmogorov-Smirnov test  
##  
## data: model$residuals  
## D = 0.12432, p-value = 0.9521  
## alternative hypothesis: two-sided
```

Didapati bahwa $p\text{-value} = 0.9521 > \alpha = 0.05$, tak tolak H0.

```
shapiro.test(model$residuals)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: model$residuals  
## W = 0.92457, p-value = 0.226
```

Didapati bahwa $p\text{-value} = 0.226 > \alpha = 0.05$, tak tolak H0.

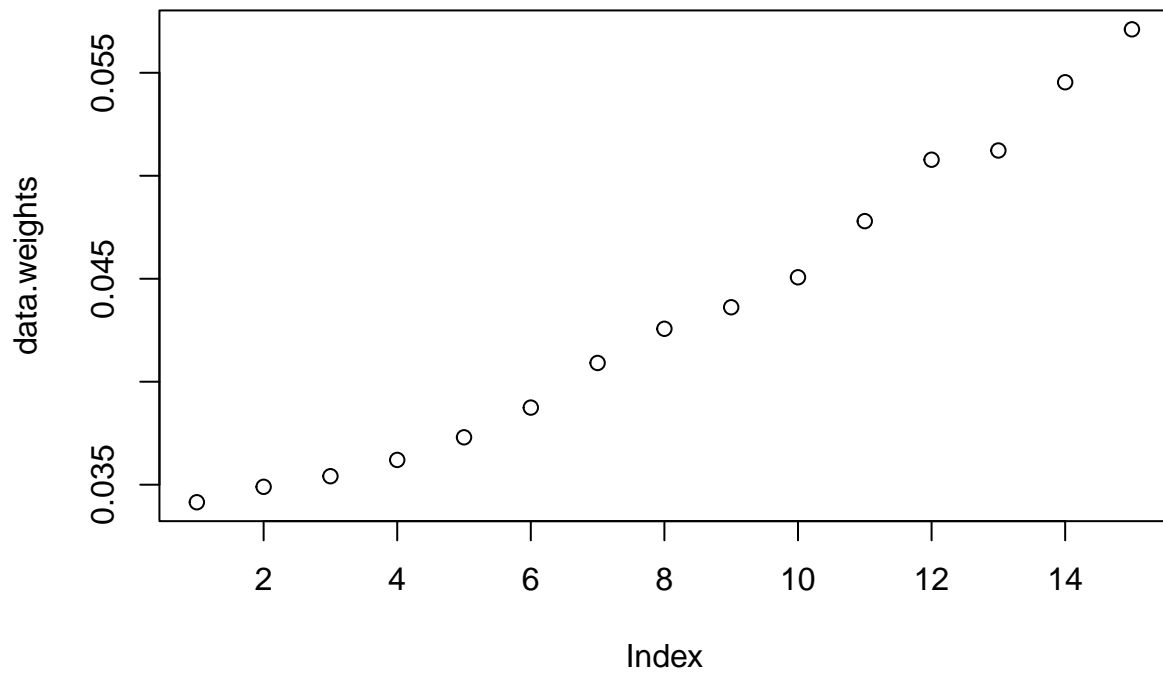
Dari uji Kolmogorov-Smirnov dan Shapiro-Wilk, keduanya didapati tak tolak H0. Pada taraf nyata 5%, belum cukup bukti untuk menyatakan sisaan tidak menyebar normal.

Weighted Least Squares

```
resid_abs <- abs(model$residuals)  
fitted_val <- model$fitted.values  
fit <- lm(resid_abs ~ fitted_val, data7)  
data.weights <- 1 / fit$fitted.values^2  
data.weights
```

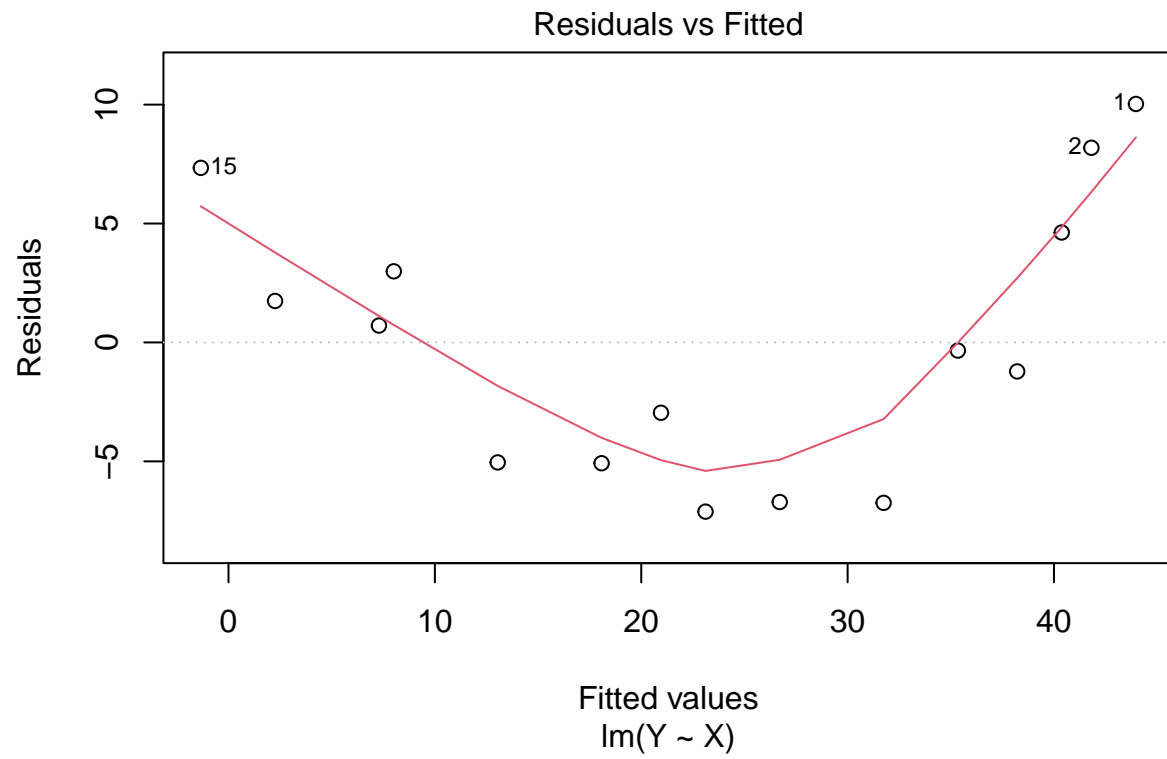
```
##          1          2          3          4          5          6          7  
## 0.03414849 0.03489798 0.03541143 0.03620311 0.03730067 0.03874425 0.04091034  
##          8          9         10         11         12         13         14  
## 0.04257072 0.04361593 0.04507050 0.04779711 0.05077885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

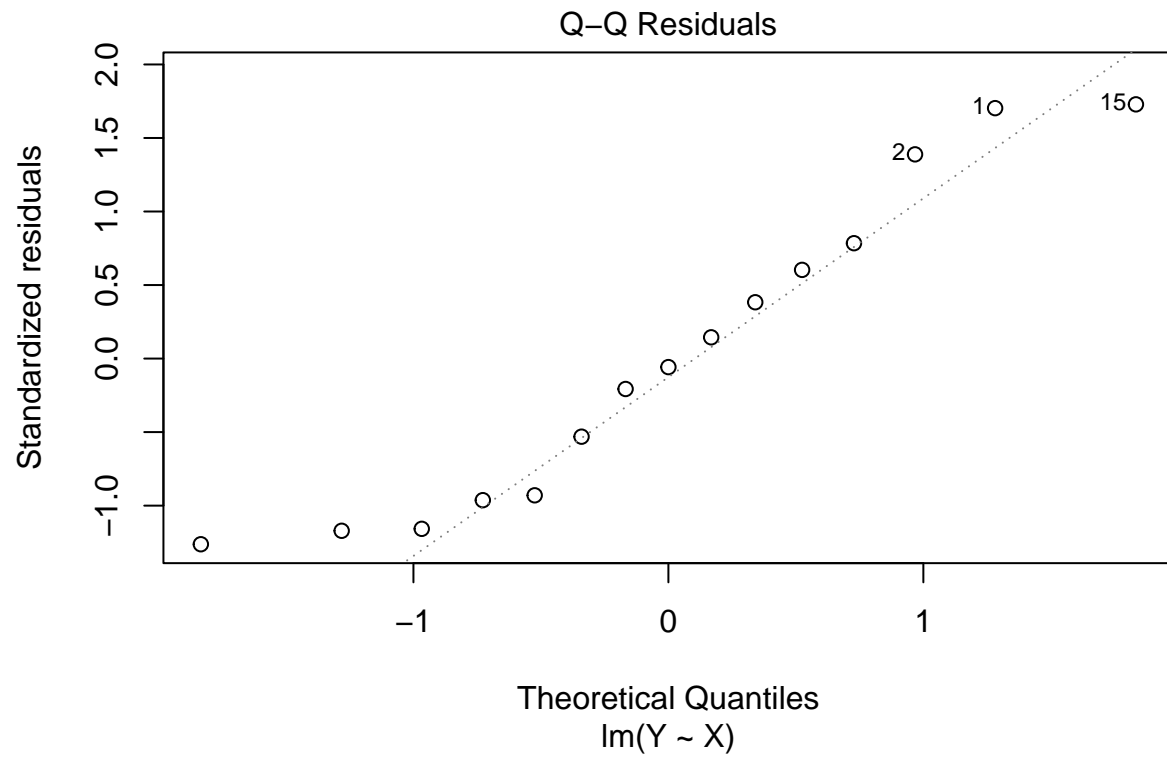
```
plot(data.weights)
```

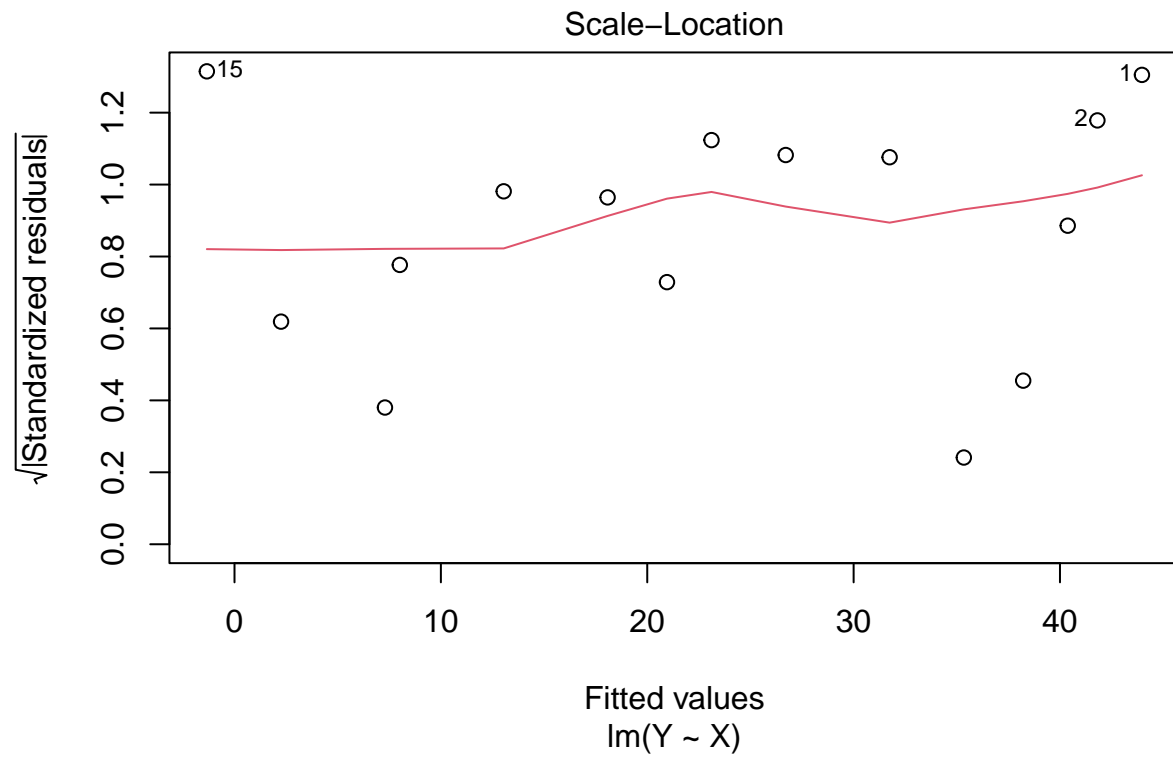


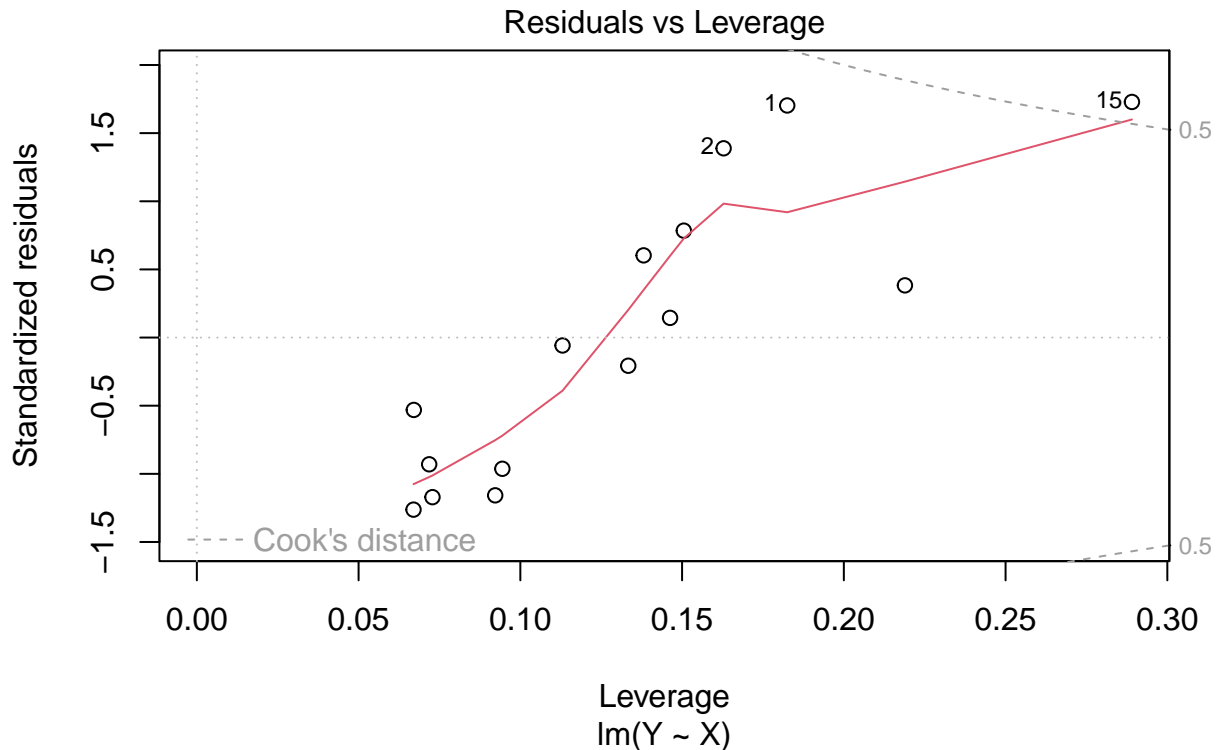
Model Regresi yang Terboboti:

```
modelweight = lm(Y~X, data = data7, weights = data.weights)
plot(modelweight)
```









```
summary(modelweight)
```

```
##
## Call:
## lm(formula = Y ~ X, data = data7, 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
```

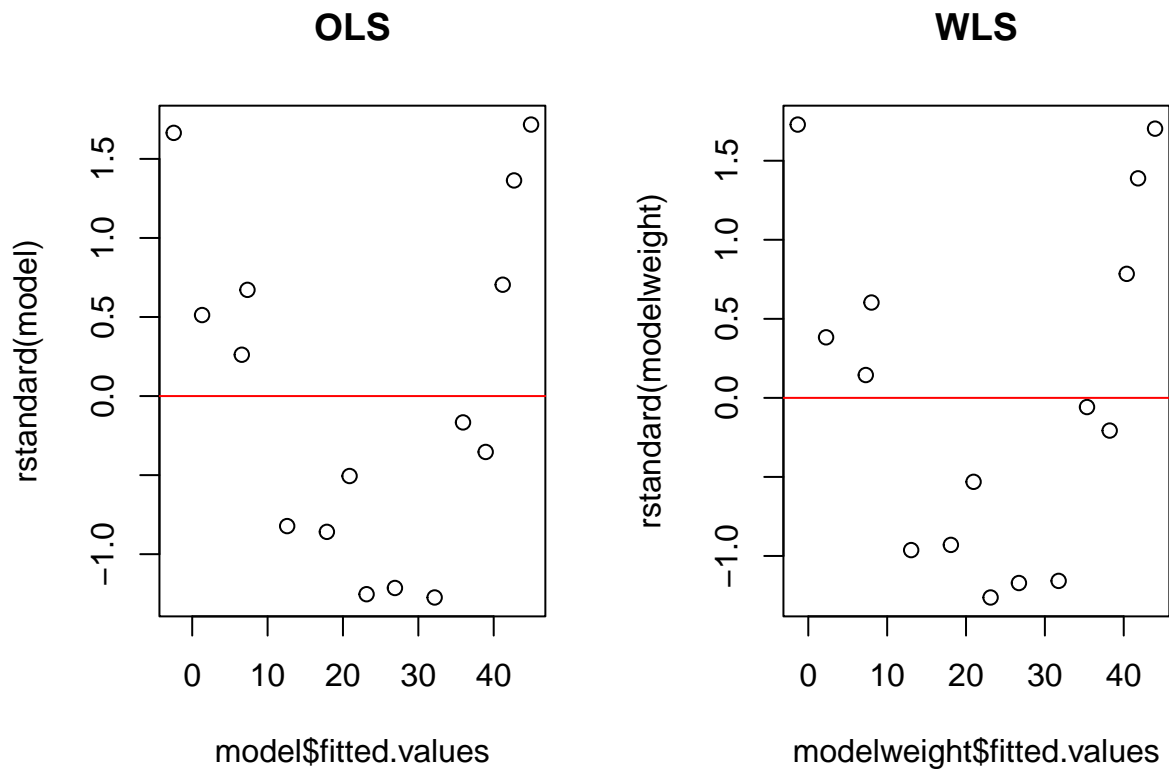
Transformasi WLS di atas menunjukkan bahwa WLS belum efektif dalam mentransformasikan model regresi. Hal tersebut dikatakan demikian karena berdasarkan eksplorasi di atas, asumsi Gauss-Marcov belum terpenuhi.

Didapati model sebagai berikut.

$$\hat{Y} = 45.41058 - 0.71925X + e$$

OLS vs WLS

```
par(mfrow=c(1,2))
plot(y = rstandard(model),
     x = model$fitted.values,
     main="OLS")
abline(h=0, col="red")
plot(y = rstandard(modelweight),
     x = modelweight$fitted.values,
     main="WLS")
abline(h=0, col="red")
```



Penyesuaian pada Data

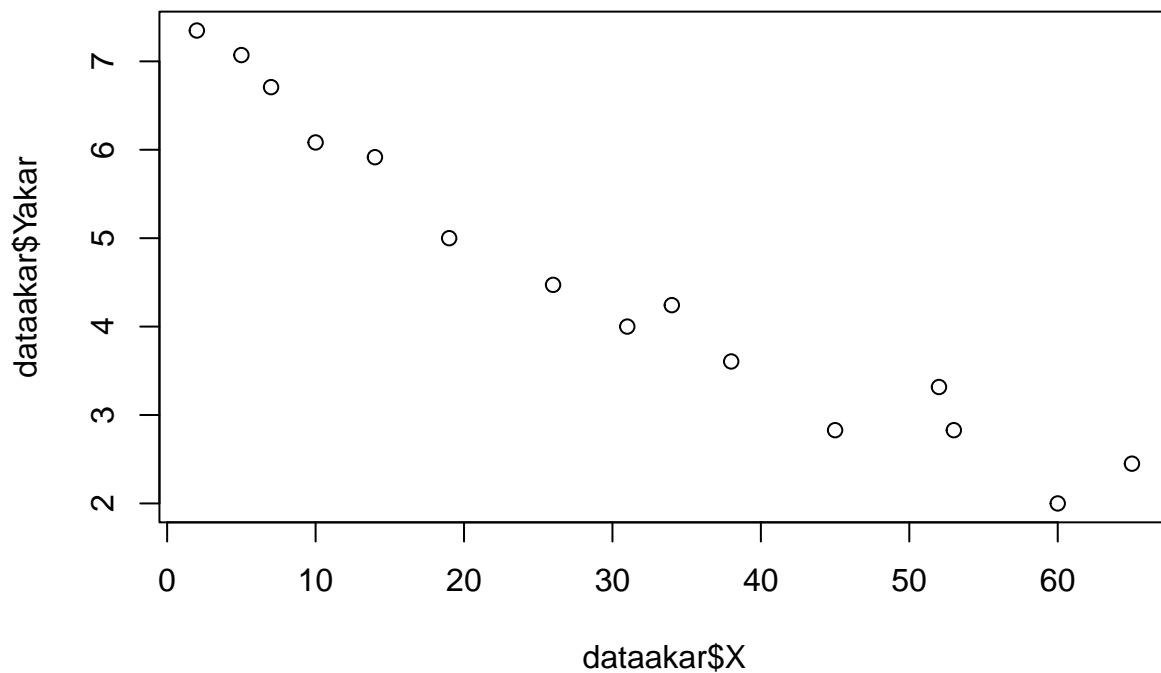
Transformasi Akar pada X, Y, atau X dan Y

```
dataakar = data7 %>% mutate(Yakar = sqrt(Y)) %>% mutate(Xakar = sqrt(X))
dataakar
```

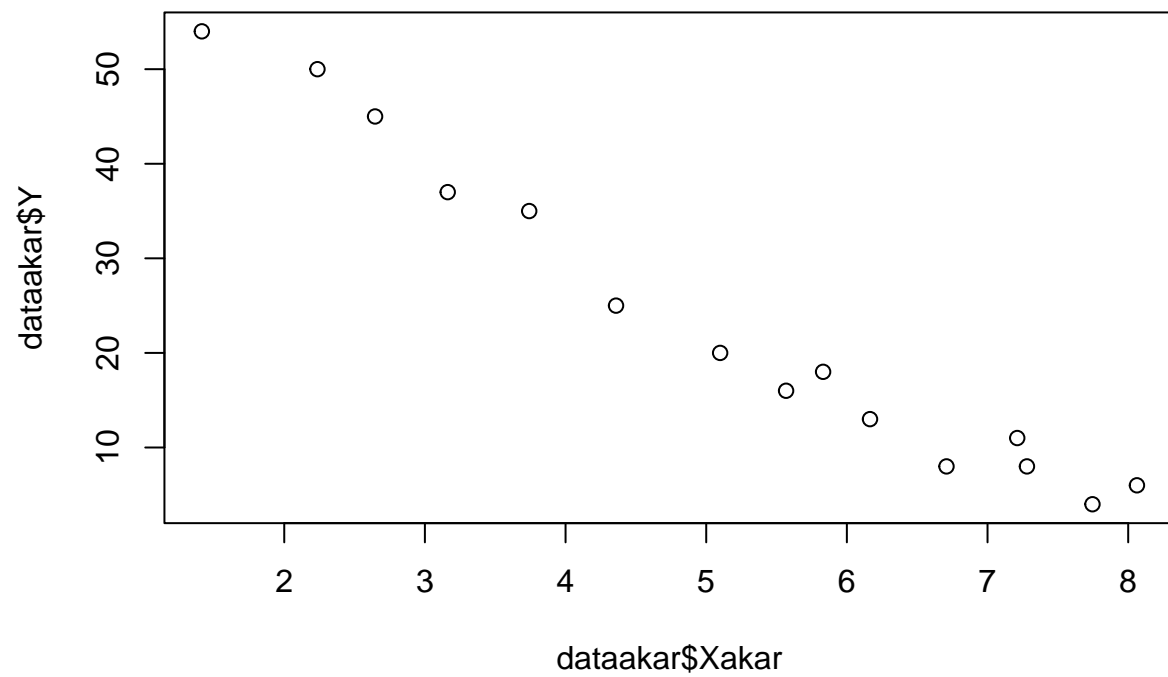


```
## # A tibble: 15 x 4
##       X     Y Yakar Xakar
##   <dbl> <dbl> <dbl> <dbl>
## 1     2    54  7.35  1.41
## 2     5    50  7.07  2.24
## 3     7    45  6.71  2.65
## 4    10    37  6.08  3.16
## 5    14    35  5.92  3.74
## 6    19    25   5     4.36
## 7    26    20  4.47  5.10
## 8    31    16   4     5.57
## 9    34    18  4.24  5.83
## 10   38    13  3.61  6.16
## 11   45     8  2.83  6.71
## 12   52    11  3.32  7.21
## 13   53     8  2.83  7.28
## 14   60     4   2     7.75
## 15   65     6  2.45  8.06
```

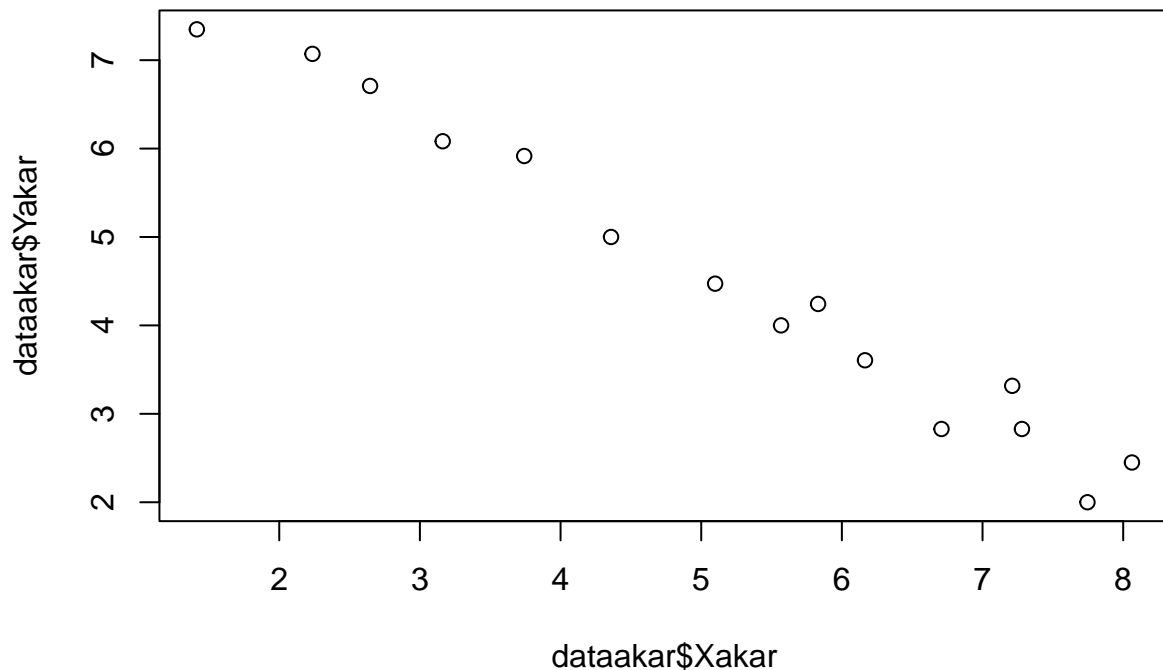
```
plot(x = dataaakar$X, y = dataaakar$Yakar)
```



```
plot(x = dataaakar$Xakar, y = dataaakar$Y)
```



```
plot(x = dataakar$Xakar, y = dataakar$Yakar)
```



```
data.sqrt = data.frame(dataakar$Xakar, dataakar$Yakar)
```

Diketahui di awal bahwa hubungan antara X dan Y cenderung membentuk pola parabola dan nilai $B1 < 0$, sehingga diperlukan transformasi data dengan mengecilkan nilai X dan/atau Y. Transformasi mengecilkan tersebut dapat dilakukan dengan membentuk X dan/atau Y menjadi akar atau pangkat setengah dari data asli.

Uji nonformal dilakukan melalui plot hubungan Xakar dengan Y, X dengan Yakar, dan Xakar dengan Yakar. Terlihat perbedaan dari masing-masing plot, sehingga dirasa perlu untuk diadakan uji lebih lanjut dalam rangka memperoleh model terbaik. Pemeriksaan asumsi dilakukan pada data dengan sisaan paling bebas.

Model Asumsi dan Pemeriksaannya

Xakar dengan Y

```
modelXakar = lm(formula = dataakar$Y ~ dataakar$Xakar)
summary(modelXakar)
```

```
##
## Call:
## lm(formula = dataakar$Y ~ dataakar$Xakar)
##
```

```
## 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 ***
## dataakar$Yakar -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
```

Didapati model sebagai berikut.

$$\hat{Y} = 63.2250 - 7.7481X^{\frac{1}{2}} + e$$

Dilakukan Durbin-Watson test

```
dwtest(modelXakar)
```

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

Diperoleh nilai p-value = 0.01422 < alpha = 0.05, sehingga tolak H0. Pada taraf 5% terdapat cukup bukti untuk menyatakan bahwa sisaan tidak saling bebas. Hal ini mengakibatkan asumsi tidak terpenuhi sehingga model tersebut bukanlah model terbaik.

X dengan Yakar

```
modelYakar = lm(formula = dataakar$Yakar ~ dataakar$X)
summary(modelYakar)
```

```
##
## Call:
## lm(formula = dataakar$Yakar ~ dataakar$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 ***
## dataakar$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
```

Didapati model sebagai berikut.

$$\hat{Y}_{\frac{1}{2}} = 7.015455 - 0.081045X + e$$

Dilakukan Durbin-Watson test

```
dwtest(modelYakar)
```

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

Dari uji Durbin-Watson didapati bahwa $p\text{-value} = 0.02493 < \alpha = 0.05$, tolak H_0 . Pada taraf 5% terdapat cukup bukti menyatakan bahwa sisaan tidak saling bebas. Hal ini mengakibatkan asumsi tidak terpenuhi sehingga model tersebut bukanlah model terbaik.

Xakar dengan Yakar

```
modelXYakar = lm(formula = dataakar$Yakar ~ dataakar$Xakar)
summary(modelXYakar)
```

```
##
## Call:
## lm(formula = dataakar$Yakar ~ dataakar$Xakar)
##
## 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 ***
## dataakar$Xakar -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
```

Didapati model sebagai berikut.

$$\hat{Y}_{\frac{1}{2}} = 8.71245 - 0.81339X_{\frac{1}{2}} + e$$

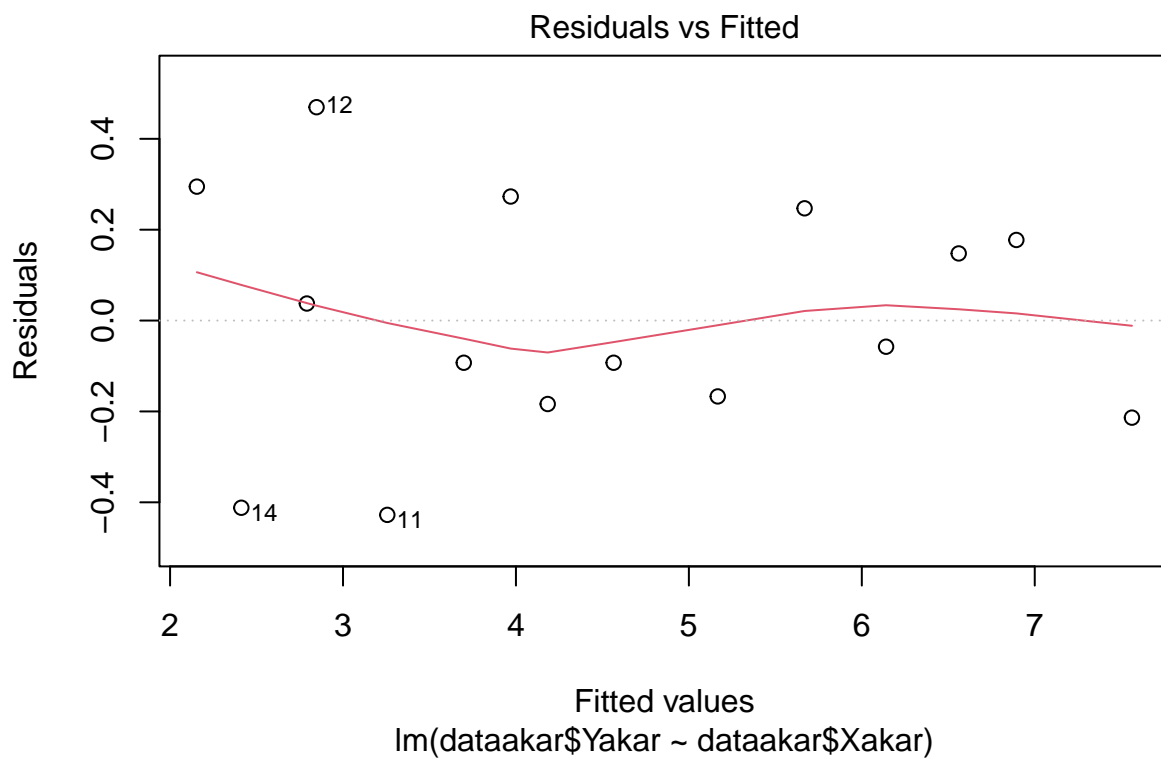
Dilakukan Durbin-Watson test

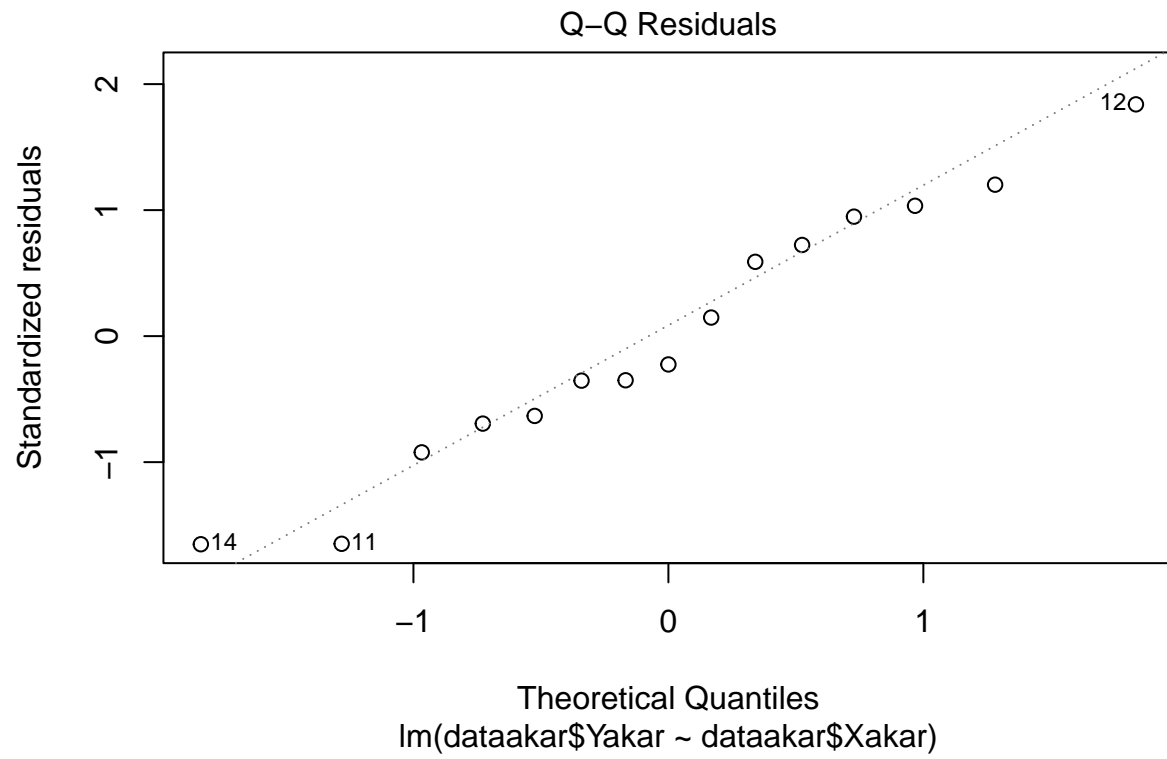
```
dwtest(modelXYakar)
```

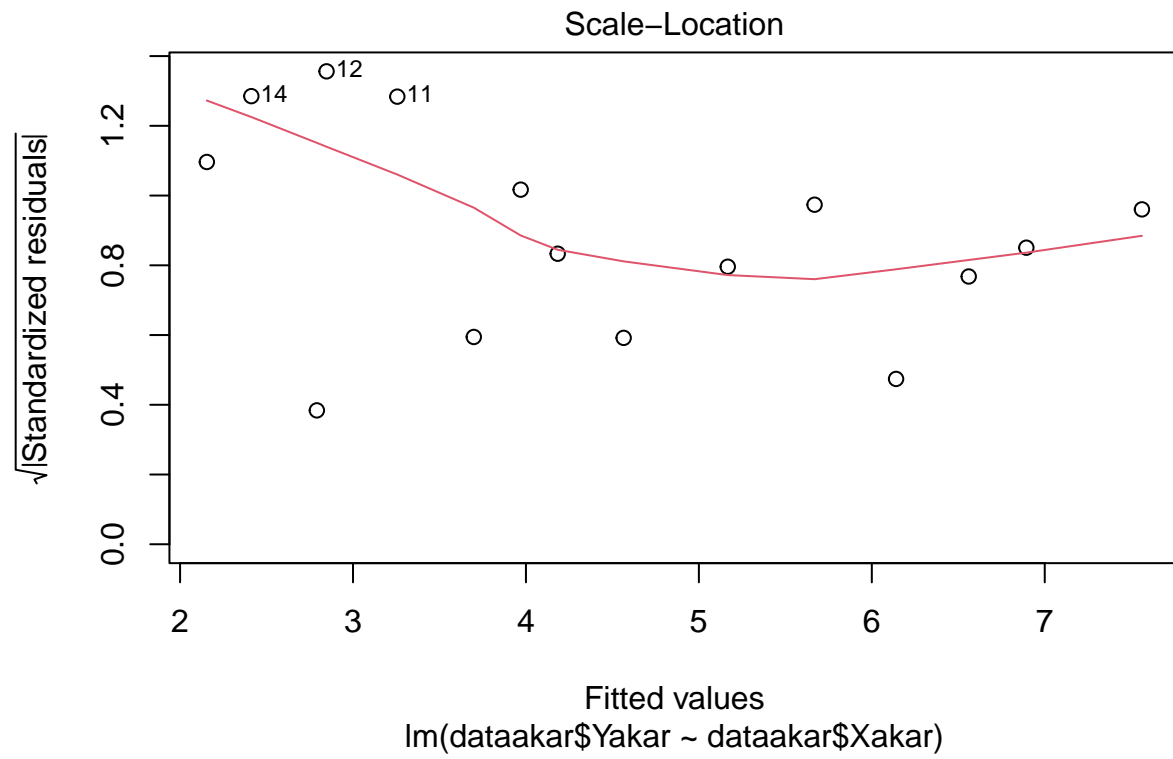
```
##  
## Durbin-Watson test  
##  
## data: modelXYakar  
## DW = 2.6803, p-value = 0.8629  
## alternative hypothesis: true autocorrelation is greater than 0
```

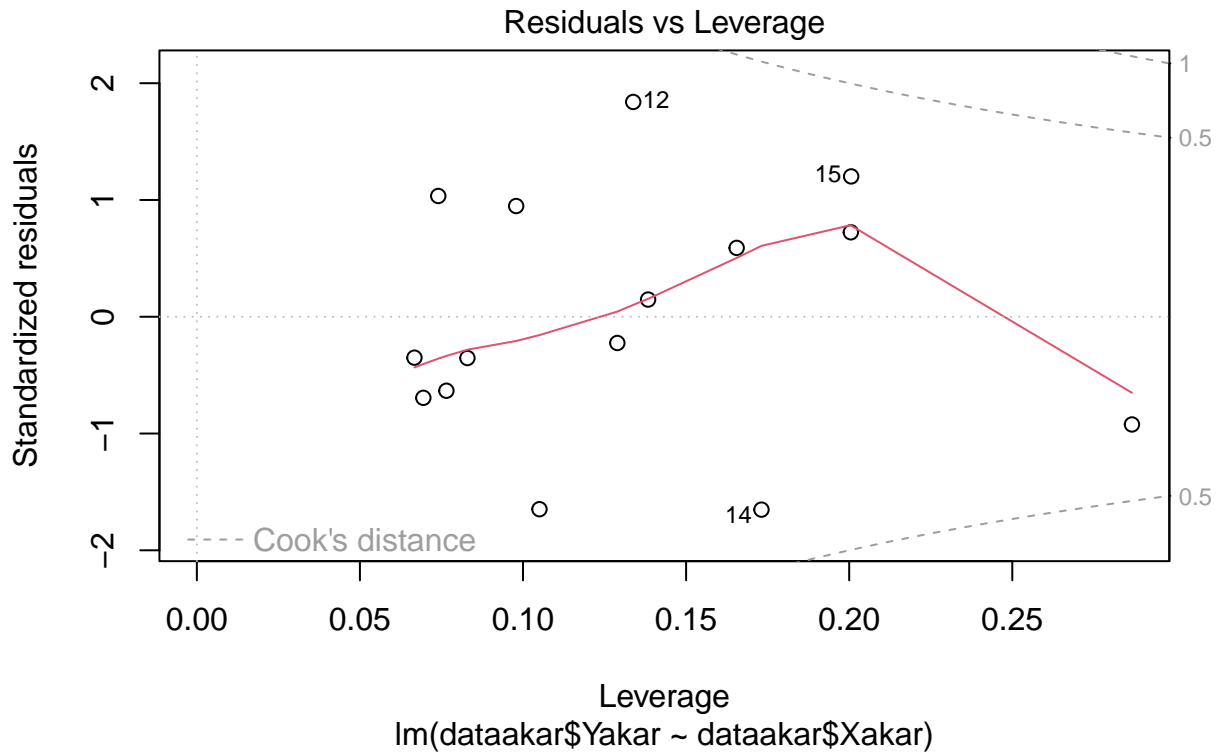
Dari Durbin-Watson test didapati bahwa $p\text{-value} = 0.8629 > \alpha = 0.05$, tak tolak H_0 . Pada taraf nyata 5%, belum cukup bukti untuk menyatakan bahwa sisaan tidak saling bebas. Berdasarkan uji autokorelasi tersebut (Durbin-Watson test) diperoleh hasil bahwa sisaan saling bebas. Namun masih perlu diperiksa dengan uji asumsi lain untuk memastikan bahwa model tersebut merupakan model terbaik.

```
plot(modelXYakar)
```









1. Harapan sisaan sama dengan nol

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

```
##
## One Sample t-test
##
## data: modelXYakar$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
```

p-value = 1 > alpha = 0.05, tak tolak H_0 . Pada taraf nyata 5%, belum terdapat cukup bukti untuk menyatakan bahwa nilai harapan sisaan tidak sama dengan 0.

2. Ragam sisaan homogen

```
ncvTest(modelXYakar)
```

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

p-value = 0.14161 > alpha = 0.05, tak tolak H0. Pada taraf nyata 5%, belum terdapat cukup bukti untuk menyatakan bahwa ragam sisaan tidak homogen.

3. Sisaan saling bebas

```
sisaan.modelXYakar = resid(modelXYakar)
(norm.modelXYakar = lillie.test(sisaan.modelXYakar))
```

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

p-value = 0.817 > alpha = 0.05, tak tolak H0. Pada taraf nyata 5%, belum terdapat cukup bukti untuk menyatakan bahwa sisaan tidak saling bebas.

Karena ketiga asumsi di atas terpenuhi, maka asumsi Gauss-Marcov terpenuhi.

Kesimpulan dan Transformasi Balik

Setelah melalui berbagai transformasi, didapati bahwa model terbaik dipenuhi ketika variabel X dan Y keduanya ditransformasi ke dalam bentuk akar atau pangkat 1/2. Dikatakan demikian karena semua asumsi dalam analisis regresi linear sederhana dapat terpenuhi. Model terbaik untuk data ini adalah

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

Perlu dilakukan transformasi balik pada model tersebut agar model tersebut dapat digunakan untuk menjelaskan peubah respons sebelum transformasi. Transformasi balik dilakukan dengan proses matematika yaitu dengan melakukan pemangkatan 2 pada model tersebut dengan mengutamakan pengubahan peubah respons terlebih dahulu.

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

Interpretasi: model tersebut menunjukkan hubungan terbalik antara \hat{Y} dengan $X^{\frac{1}{2}}$ sebagai hubungan kuadratik. Nilai $X^{\frac{1}{2}}$ yang semakin besar akan mengakibatkan semakin kecilnya nilai rata-rata \hat{Y} . Jika 0 berada dalam selang amatan dan $X^{\frac{1}{2}}$ bernilai 0, akan mengakibatkan nilai rata-rata \hat{Y} sebesar 8.71245. Kenaikan satu satuan $X^{\frac{1}{2}}$ akan menurunkan nilai rata-rata \hat{Y} sebesar 0.81339.