

Tugas Individu Analisis Regresi Pt 7

Thufaillah Ulfah J

2024-03-05

Library

```
library(readxl)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(plotly)

## Loading required package: ggplot2

##
## 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 4.3.3

## Loading required package: zoo

##
## Attaching package: 'zoo'
```

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

library(car)

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##      recode

library(randtests)
library(lmtest)
```

Data

```
data <- read_xlsx("C:\\Users\\MyBook Hype\\Downloads\\Data anreg pt7 .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

```
model1 = lm(formula = Y ~ X, data)
summary(model1)

##
## 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

model1

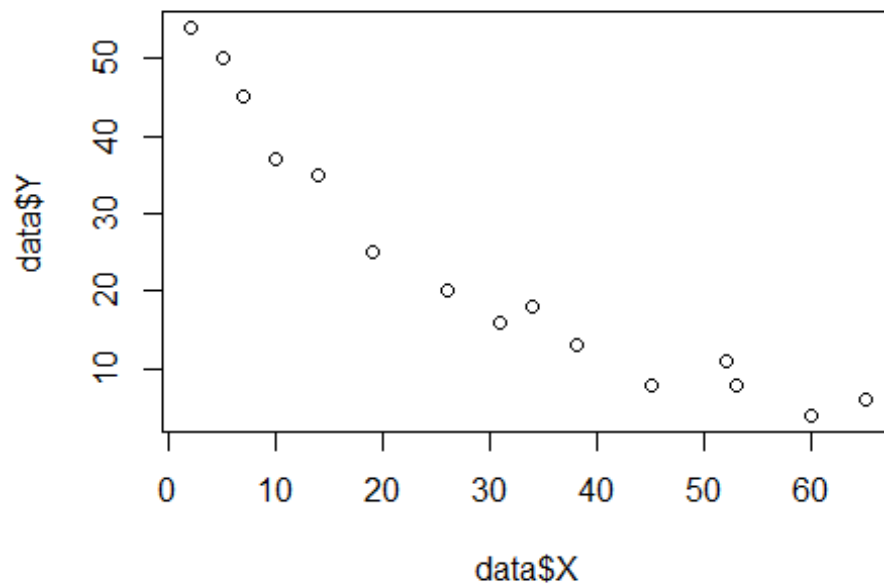
##
## Call:
## lm(formula = Y ~ X, data = data)
##
## Coefficients:
## (Intercept)          X
##    46.4604    -0.7525
```

Didapatkan Model Regresi sebagai berikut:

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

Model tersebut belum dikatakan sebagai model terbaik karena belum melalui eksplorasi data dan serangkaian uji asumsi serta normalitas untuk memperoleh model yang optimal.

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



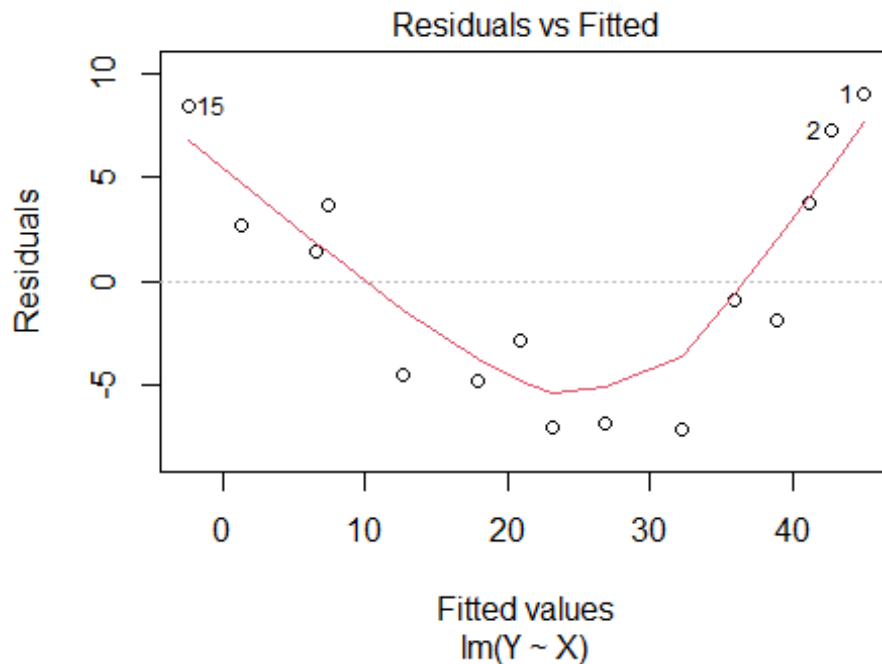
Berdasarkan plot tersebut, dapat disimpulkan bahwa hubungan antara X dan Y tidak linear.

Pemeriksaan Asumsi

Eksplorasi Kondisi Gauss-Markov

Plot sisaan vs Y duga

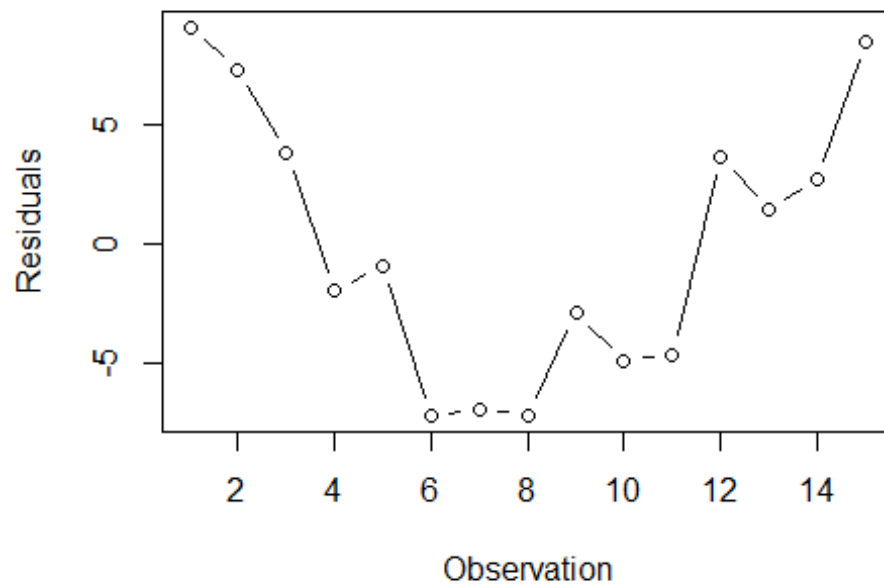
```
plot(model1,1)
```



Berdasarkan plot sisaan vs Y, terlihat sisaan berada di sekitar nilai 0, menunjukkan nilai harapan galat adalah nol. Namun, lebar pita untuk setiap nilai dugaan perlu di uji lebih lanjut untuk mengetahui ragamnya homogen atau tidak. Bentuk pola yang didapat yaitu pola kurva, artinya model kurva tidak pas sehingga perlu suku lain dalam model atau transformasi terhadap Y

Plot sisaan vs urutan

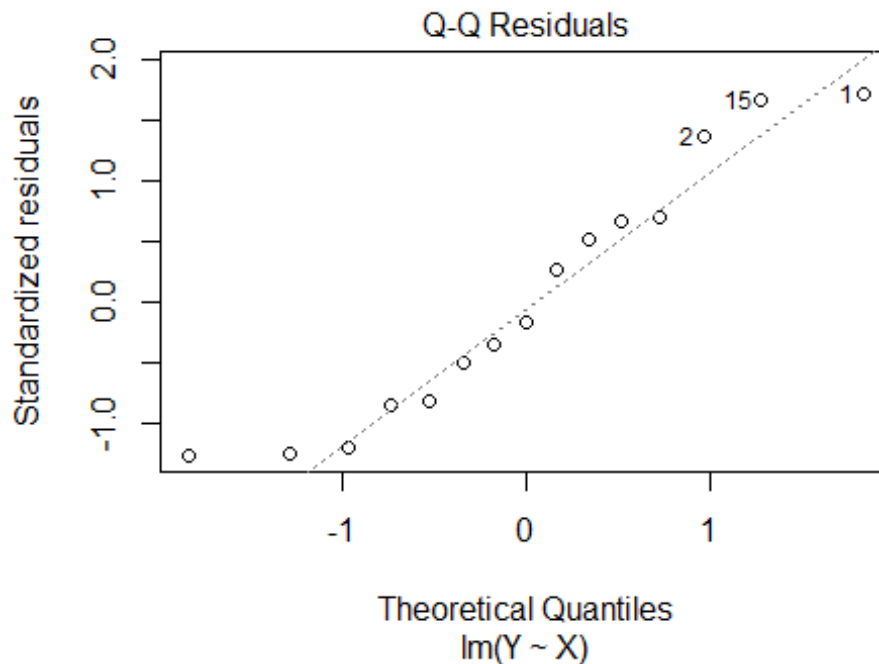
```
plot(x = 1:dim(data)[1],
     y = model1$residuals,
     type = 'b',
     ylab = "Residuals",
     xlab = "Observation")
```



Berdasarkan plot sisaan vs urutan, terlihat tebarannya berpola, sehingga sisaan tidak saling bebas dan model tidak pas

Eksplorasi Normalitas Sisaan dengan QQ-plot

```
plot(model1,2)
```



Uji Formal Kondisi Gauss-Markov

1. Nilai harapan sisaan sama dengan nol

$H_0 : \text{Sisaan menyebar normal} \quad H_1 : \text{Sisaan tidak menyebar normal}$

```
t.test(model1$residuals,mu = 0,conf.level = 0.95)

##
## One Sample t-test
##
## data: model1$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
```

Diketahui bahwa $p\text{-value} > \alpha$, sehingga tak tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa nilai harapan sisaan sama dengan nol

2. Ragam sisaan homogen

$H_0 : \text{Ragam sisaan homogen} \quad H_1 : \text{Ragam sisaan tidak homogen}$

```
bptest(model1)

##
## studentized Breusch-Pagan test
##
## data: model1
## BP = 0.52819, df = 1, p-value = 0.4674
```

Diketahui bahwa $p\text{-value} > \alpha$, sehingga tak tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa ragam sisaan homogen

3. Sisaan saling bebas

$H_0 : \text{Sisaan saling bebas} \quad H_1 : \text{Sisaan tidak saling bebas}$

```
dwtest(model1)

##
## Durbin-Watson test
##
## data: model1
## DW = 0.48462, p-value = 1.333e-05
## alternative hypothesis: true autocorrelation is greater than 0
```

Diketahui bahwa $p\text{-value} < \alpha$, sehingga tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa sisaan tidak saling bebas

Uji Formal Normalitas Sisaan

$H_0 : \text{Sisaan menyebar normal} \quad H_1 : \text{Sisaan tidak menyebar normal}$

```
shapiro.test(model1$residuals)

##
## Shapiro-Wilk normality test
##
## data: model1$residuals
## W = 0.92457, p-value = 0.226
```

Berdasarkan Shapiro-Wilk normality test diketahui bahwa $p\text{-value} > \alpha$, sehingga tak tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa sisaan menyebar normal.

Berdasarkan Serangkaian Uji Formal Kondisi Gauss-Markov dan Uji Formal Normalitas Sisaan didapatkan pelanggaran asumsi Gauss-Markov yaitu, tidak adanya autokorelasi. Karena didapatkan $p\text{-value} < \alpha$ artinya sisaan tidak saling bebas yang seharusnya sisaan saling bebas.

Transformasi Data

Data Transformasi

```
Y = sqrt(data$Y)
X = sqrt(data$X)
data2 <- data_frame(X,Y)

## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## [i] Please use `tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

data2

## # A tibble: 15 × 2
##       X     Y
##   <dbl> <dbl>
## 1  1.41  7.35
## 2  2.24  7.07
## 3  2.65  6.71
## 4  3.16  6.08
## 5  3.74  5.92
## 6  4.36  5
## 7  5.10  4.47
## 8  5.57  4
## 9  5.83  4.24
## 10 6.16  3.61
## 11 6.71  2.83
## 12 7.21  3.32
## 13 7.28  2.83
## 14 7.75  2
## 15 8.06  2.45
```

Model Regresi

```
model2 = lm(formula = Y ~ X, data2)
summary(model2)

##
## Call:
## lm(formula = Y ~ X, data = data2)
##
## 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

model2

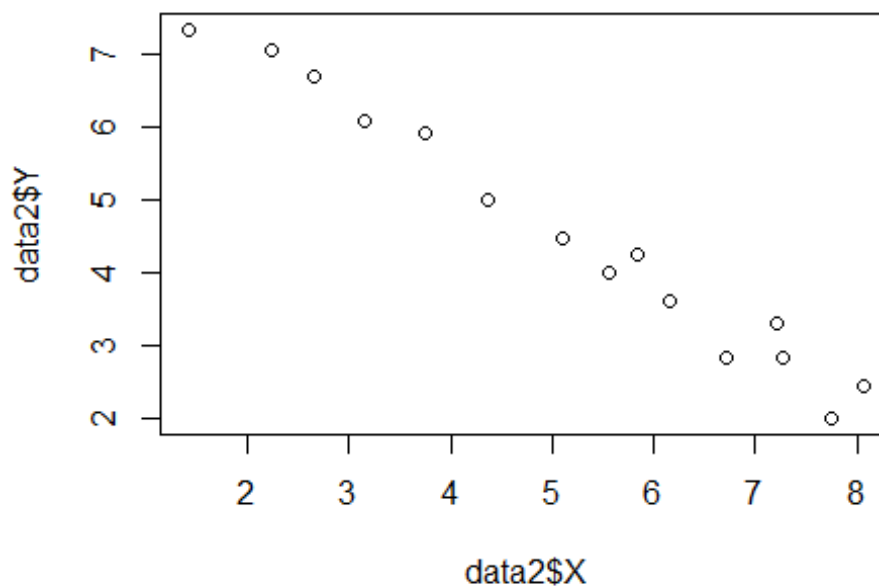
##
## Call:
## lm(formula = Y ~ X, data = data2)
##
## Coefficients:
## (Intercept)          X
##      8.7125      -0.8134
```

Didapatkan Model Regresi sebagai berikut:

$$\hat{Y} = 8.7125 - 0.8134X$$

#Eksplorasi ## Plot Hubungan X dan Y Asumsi bentuk model regresi linear sederhana dengan plot antara X dan Y

```
plot(x=data2$X,y=data2$Y)
```

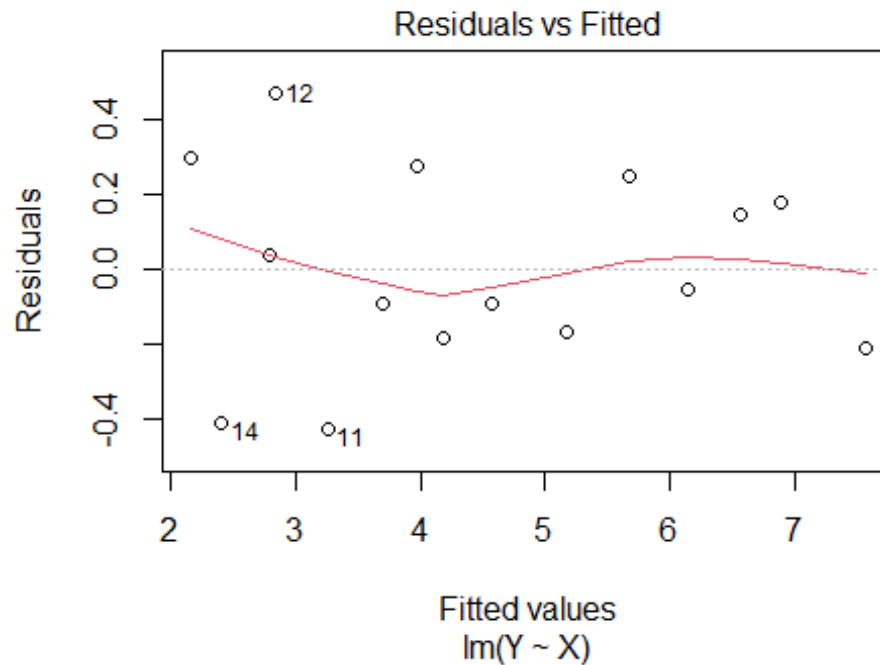


Pemeriksaan Asumsi

Eksplorasi Kondisi Gauss-Markov

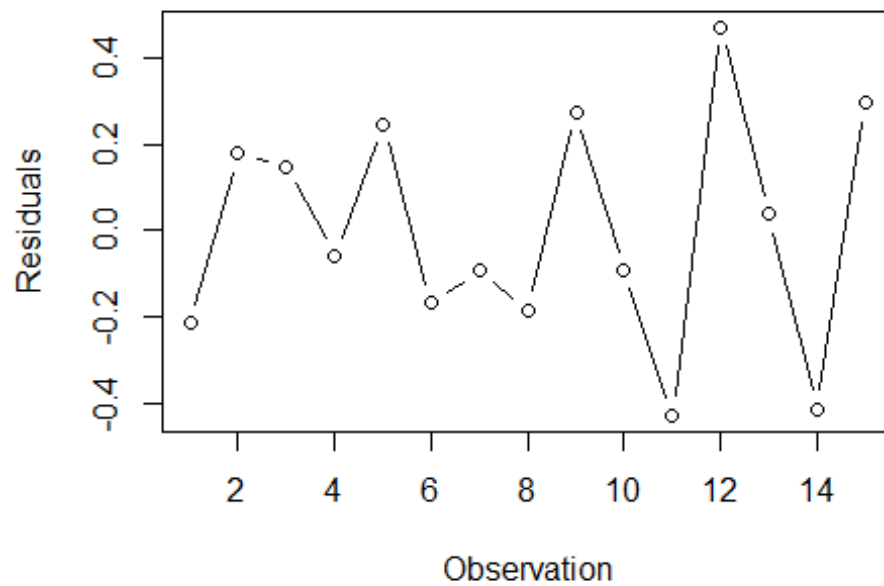
Plot sisaan vs Y duga

```
plot(model2,1)
```



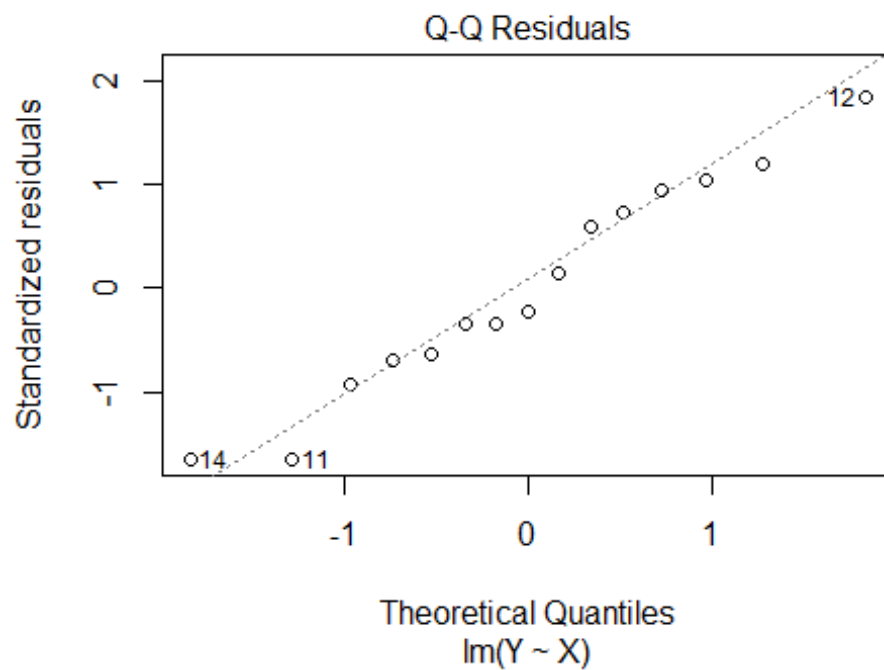
Plot sisaan vs urutan

```
plot(x = 1:dim(data2)[1],  
     y = model2$residuals,  
     type = 'b',  
     ylab = "Residuals",  
     xlab = "Observation")
```



Eksplorasi Normalitas Sisaan dengan QQ-plot

```
plot(model2,2)
```



Uji Formal Kondisi Gauss-Markov

1. Nilai harapan sisaan sama dengan nol

\$\$ H_0 : \text{Nilai harapan sisaan sama dengan 0} \\ H_1 : \text{Nilai harapan tidak sama dengan 0} \$\$

```
t.test(model2$residuals,mu = 0,conf.level = 0.95)

##
##  One Sample t-test
##
## data:  model2$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
```

Diketahui bahwa p-value > alpha, sehingga tak tolak H_0 . Oleh karena itu,dapat disimpulkan bahwa nilai harapan sisaan sama dengan nol

2.Ragam sisaan homogen

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

```
ncvTest(model2)

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

Diketahui bahwa p-value > alpha, sehingga tak tolak H_0 . Oleh karena itu,dapat disimpulkan bahwa ragam sisaan homogen

3. Sisaan saling bebas

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

```
dwtest(model2)

##
##  Durbin-Watson test
##
## data:  model2
## DW = 2.6803, p-value = 0.8629
## alternative hypothesis: true autocorrelation is greater than 0
```

Diketahui bahwa $p\text{-value} > \alpha$, sehingga tak tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa sisaan saling bebas

Uji Formal Normalitas Sisaan

$H_0 : \text{Sisaan menyebar normal} \quad H_1 : \text{Sisaan tidak menyebar normal}$

```
shapiro.test(model2$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  model2$residuals
## W = 0.96504, p-value = 0.7791
```

Berdasarkan Shapiro-Wilk normality test diketahui bahwa $p\text{-value} > \alpha$, sehingga tak tolak H_0 . Oleh karena itu, dapat disimpulkan bahwa sisaan menyebar normal.

Berdasarkan transformasi yang telah dilakukan didapat model regresi yang lebih baik dan efektif disertai dengan semua asumsi Gauss Markov dan Normalitas sudah terpenuhi dalam analisis regresi linear sederhana. Sehingga model regresi terbaik dari data ini adalah:

$$\hat{Y} = 8.7125 - 0.8134X$$