

Tugas Analisis Regresi Pertemuan 7

Qonita Husnia Rahmah

2024-03-05

Library

```
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 'dplyr' 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(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(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

```
library(readxl)
DataRegresi <- read_xlsx("C:\\Users\\ASUS\\Documents\\Nita\\SEMESTER
4\\Analisis Regresi\\Kuliah\\Data Tugas Pertemuan 7.xlsx")
DataRegresi
```

```
## # 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
```

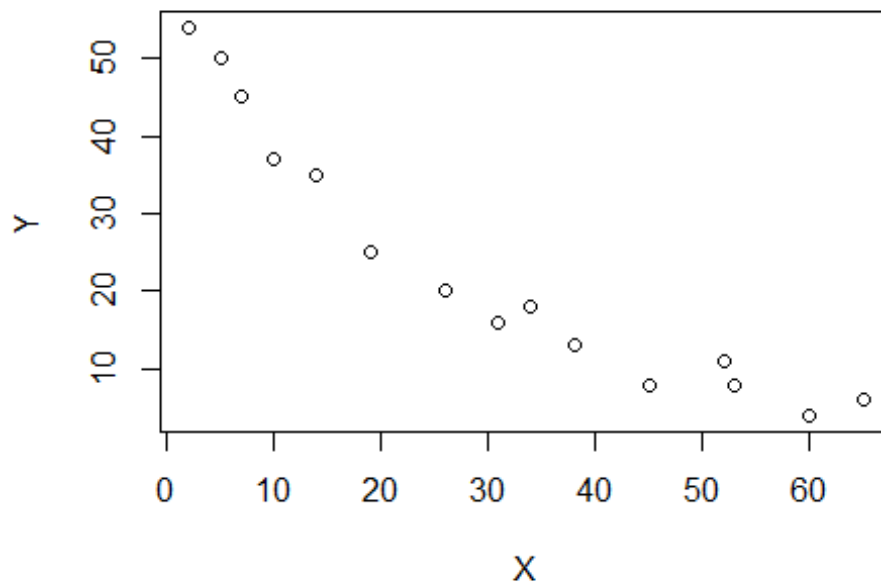
```
Y<-DataRegresi$Y
X<-DataRegresi$X
```

```
data<-data.frame(cbind(X, Y))
head(data)
```

```
##      X  Y
## 1  2 54
## 2  5 50
## 3  7 45
## 4 10 37
## 5 14 35
## 6 19 25
```

Scatter Plot Antara X dan Y

```
plot(X,Y)
```



Dari scatter plot antara X dan Y, dapat dilihat bahwa X dan Y memiliki hubungan yang tidak linear cenderung membentuk parabola.

Pemodel Regresi Linear

```
model <- lm(formula = Y ~ X, data = data)
summary(model)
```

```
##
## 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

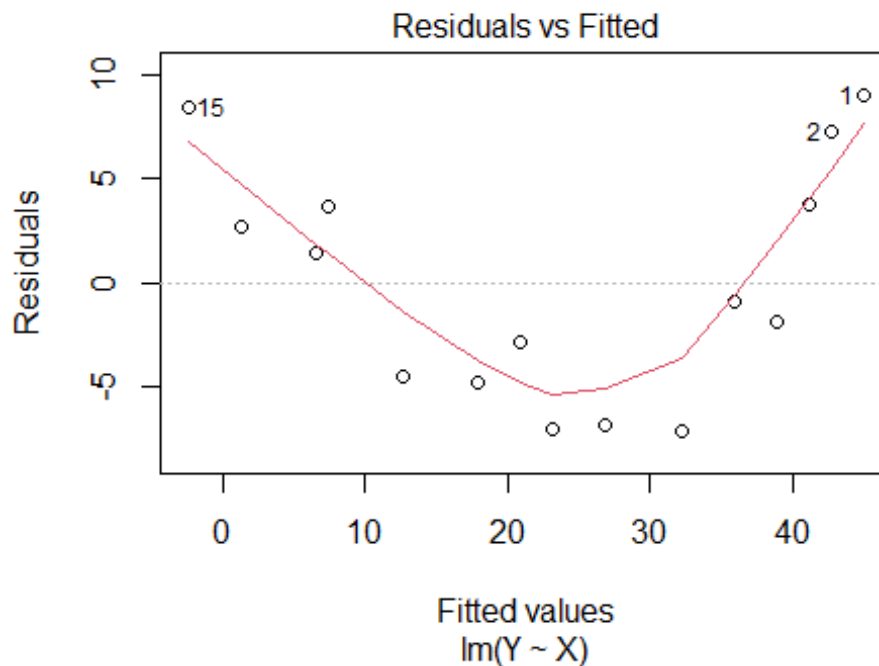
model
```

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

Eksplorasi Kondisi Gauss-Markov, Pemeriksaan dengan Grafik

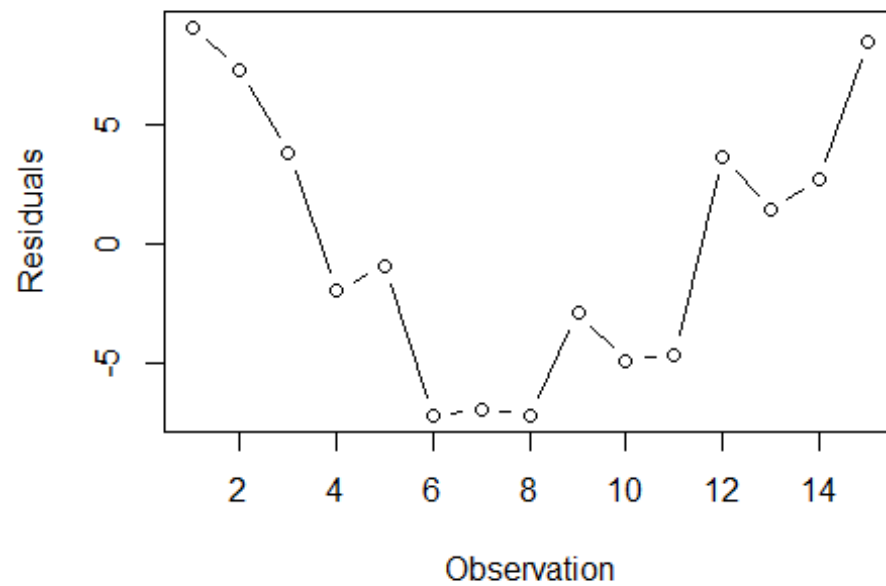
Plot Sisaan vs Y duga

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



Plot Sisaan vs Urutan

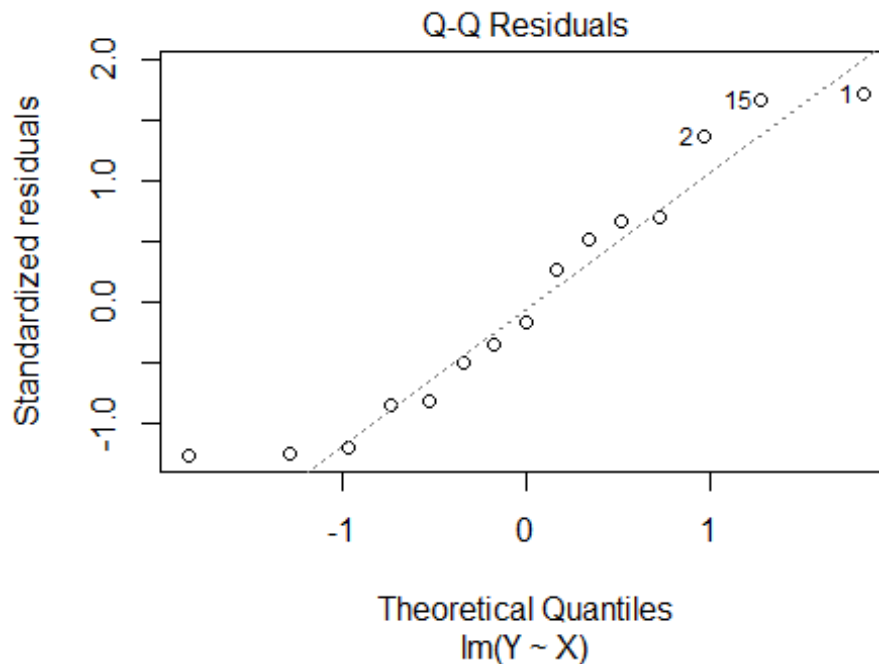
```
plot(x = 1:dim(data)[1],
     y = model$residuals,
     type = 'b',
     ylab = "Residuals",
     xlab = "Observation")
```



Sebaran membentuk pola kurva, artinya Sisaan tidak saling bebas

Normalitas Sisaan dengan QQ-Plot

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



Pengujian Asumsi

Uji Asumsi Gauss-Markov

1. Nilai Harapan Sisaan sama dengan nol

H_0 : \text{Nilai harapan sama dengan 0} \\ H_1 : \text{Nilai harapan 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
```

Dalam pengujian asumsi Gauss-Markov yang pertama, didapat bahwa $p\text{-value} > 0.05$, maka tak tolak H_0 , atau nilai harapan sisaan sama dengan nol

2. Ragam Sisaan Homogen (Homogenitas)

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

```
kehomogenan = lm(formula = abs(model$residuals) ~ X,
data = data)
summary(kehomogenan)

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

Didapat bahwa p-value > 0.05, maka taktolak H_0 , atau ragam sisaan homogen untk setiap nilai x.

3. Sisaan Saling Bebas/Tidak Ada Autokorelasi

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

```
library (randtests)
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
```

Didapat bahwa $p\text{-value} < 0.05$, maka tolak H_0 , atau sisaan tidak saling bebas (terdapat autokorelasi)

Uji Normalitas Sisaan

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

```
shapiro.test(model$residuals)

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

Dari Uji Shapiro-Wilk, didapat hasil bahwa $p\text{-value} > \alpha$, maka tak tolak H_0 , atau sisaan menyebar normal dengan $\alpha 0.05$.

Kesimpulan

Dari ketiga asumsi Gauss-Markov yang ada, didapat hasil bahwa terdapat asumsi yang tidak terpenuhi yaitu pelanggaran asumsi tidak ada autokorelasi dengan melihat Durbin Watson Test yang telah dilakukan.

Penanganan Kondisi Tak Standar

Transformasi Data

```
new_x <- sqrt(data$X)
new_y <- sqrt(data$Y)
transformasi <- data.frame(new_x, new_y)
transformasi

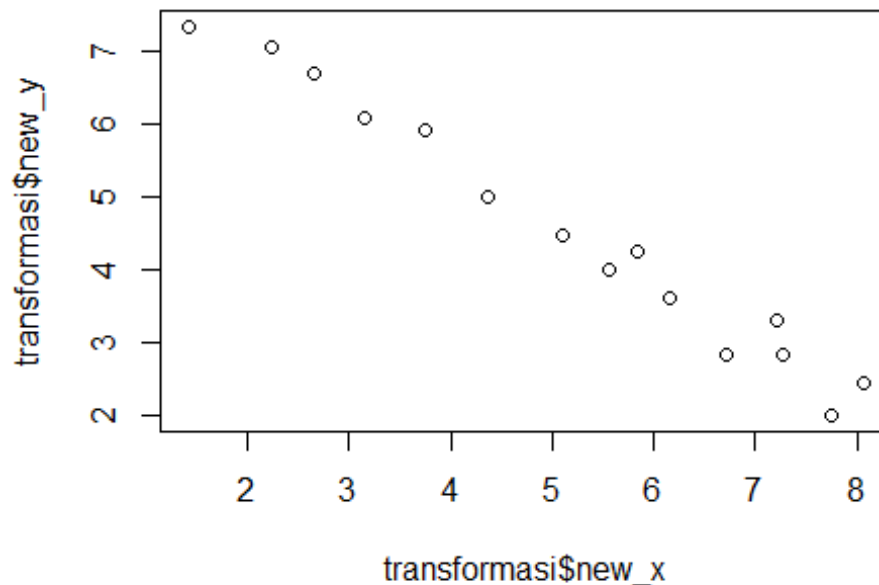
##      new_x    new_y
## 1  1.414214  7.348469
## 2  2.236068  7.071068
## 3  2.645751  6.708204
## 4  3.162278  6.082763
## 5  3.741657  5.916080
## 6  4.358899  5.000000
## 7  5.099020  4.472136
```



```
## 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
```

Plot Baru

```
plot(transformasi$new_x, transformasi$new_y)
```



Model Linear Baru

```
ModelBaru <- lm(transformasi$new_y ~ transformasi$new_x, data=transformasi)
summary (ModelBaru)
```

```
##
## Call:
## lm(formula = transformasi$new_y ~ transformasi$new_x, data = transformasi)
##
## 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 ***
```

```
## transformasi$new_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

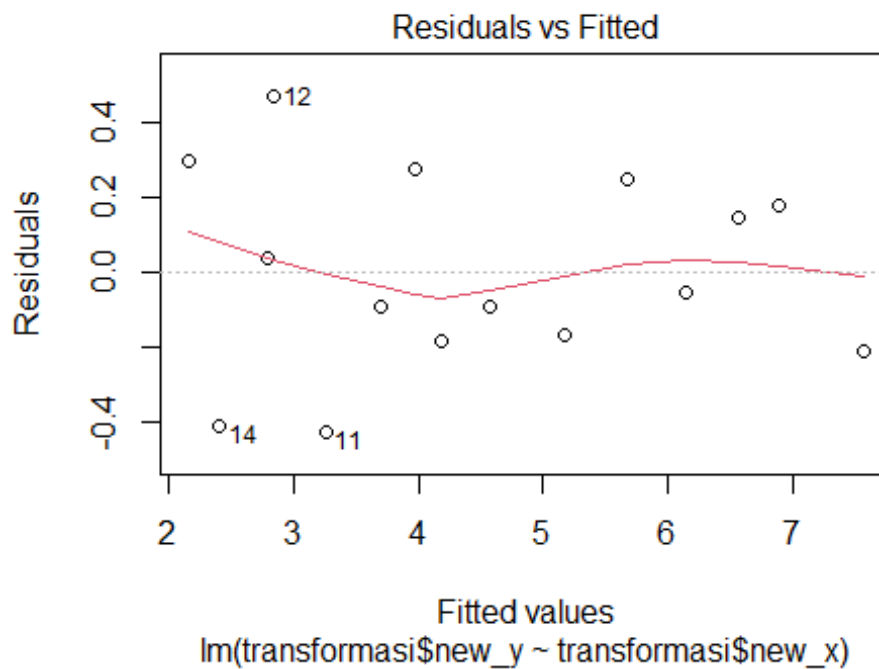
ModelBaru

##
## Call:
## lm(formula = transformasi$new_y ~ transformasi$new_x, data = transformasi)
##
## Coefficients:
##          (Intercept)  transformasi$new_x
##              8.7125          -0.8134
```

Eksplorasi Kondisi Gauss-Markov, Pemeriksaan dengan Grafik

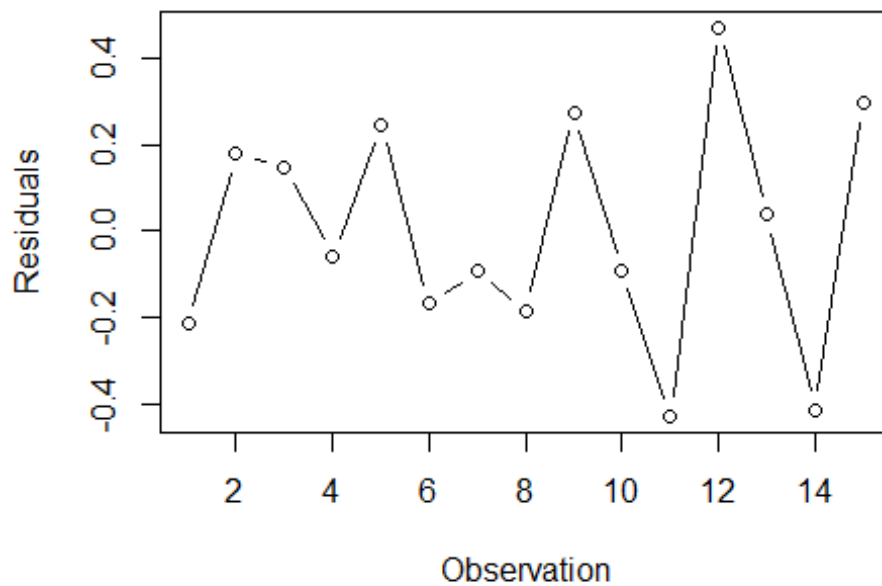
Plot Sisaan vs Y duga

```
plot(ModelBaru,1)
```



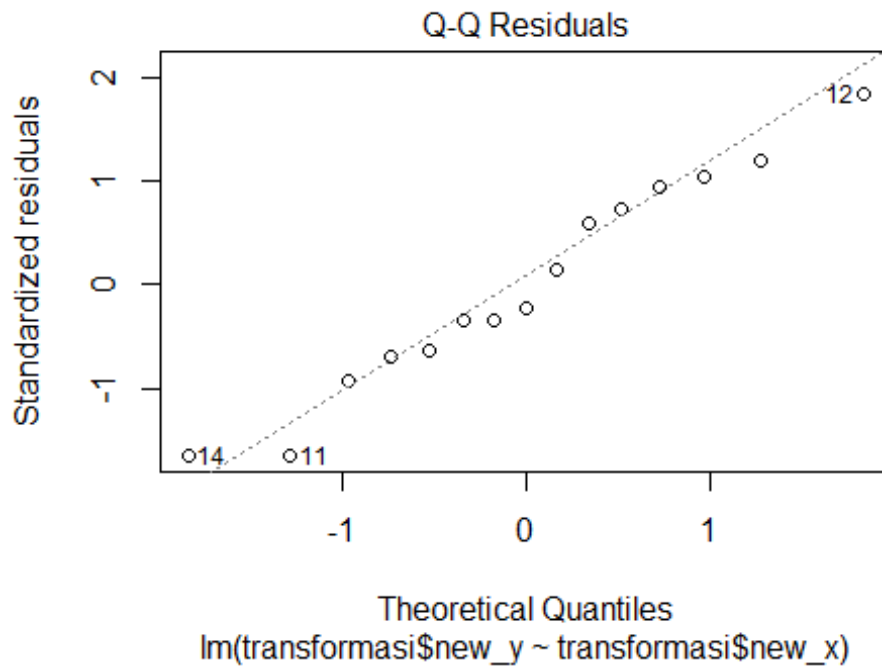
Plot Sisaan vs Urutan

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



Normalitas Sisaan dengan QQ-Plot

```
plot(ModelBaru,2)
```



Pengujian Asumsi

Uji Asumsi Gauss-Markov

1. Nilai Harapan Sisaan sama dengan nol

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

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

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

Dalam pengujian asumsi Gauss-Markov yang pertama, didapat bahwa p-value > 0.05, maka tak tolak H_0 , atau nilai harapan sisaan sama dengan nol

2. Ragam Sisaan Homogen (Homogenitas)

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

```
library(car)

## Warning: package 'car' was built under R version 4.3.3
## 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

ncvTest(ModelBaru)

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

Didapat bahwa p-value > 0.05, maka taktolak H_0 , atau ragam sisaan homogen untk setiap nilai x.

3. Sisaan Saling Bebas/Tidak Ada Autokorelasi

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

```
library (randtests)
runs.test(ModelBaru$residuals)

##
## Runs Test
##
## data: ModelBaru$residuals
## statistic = 0, runs = 8, n1 = 7, n2 = 7, n = 14, p-value = 1
## 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
```

Didapat bahwa p-value > 0.05, maka tolak H_0 , atau sisaan saling bebas(tidak terdapat autokorelasi)

Uji Normalitas Sisaan

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

```
shapiro.test(model$residuals)

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

Dari Uji Shapiro-Wilk, didapat hasil bahwa p-value > alpha, maka tak tolak H_0 , atau sisaan menyebar normal dengan alpha 0.05.

Kesimpulan

Pada model baru, diperoleh hasil bahwa memenuhi ketiga asumsi Gauss-markov serta sisaan menyebar normal. Berdasarkan transformasi yang dilakukan, maka akan diperoleh model regresi yang lebih efektif dengan semua asumsi telah terpenuhi dalam analisis regresi linear sederhana.

Model regresi setelah di transformasi adalah sebagai berikut :

$$Y^* = 8.71245 - 0.814X^* + e$$

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

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

Sehingga model terbaik untuk data ini adalah:

$$\hat{Y} = \left(8.71245 - 0.814X^{\frac{1}{2}} \right)^2 + e$$

Model regresi yang di peroleh menunjukkan hubungan kuadrat negatif antara Y dan X. Ketika X meningkat, Y cenderung akan menurun dengan kecepatan yang semakin cepat. Konstanta sebesar 8.71245 mewakili nilai Y ketika X sama dengan 0. Koefisien -0.0814 menunjukkan pengaruh perubahan X terhadap Y. Semakin besar nilai absolut koefisien, semakin besar pengaruh X terhadap Y.