# Tugas Individu Analisis Regresi Penanganan Kondisi Tak Standar

Haidar Ramdhani – G140122101005

2024-03-03

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.2
```

```
data <- read_xlsx("D:\\Campss\\Season 4\\Anreg\\7\\Tugas Individu.xlsx",sheet
= "Sheet3")
(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 <- (lm(Y~X, 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
```

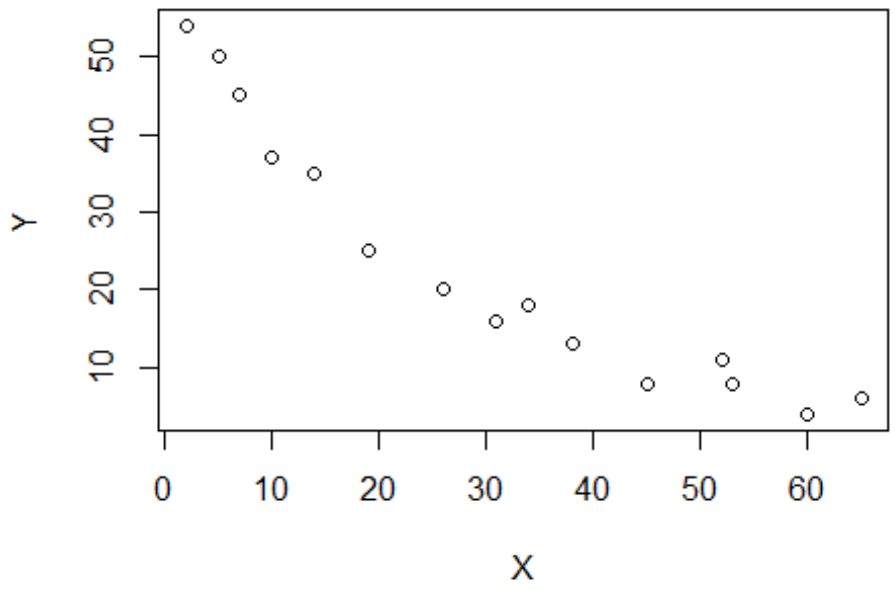
```
mean(model$residuals)
```

```
## [1] -7.254614e-16
```

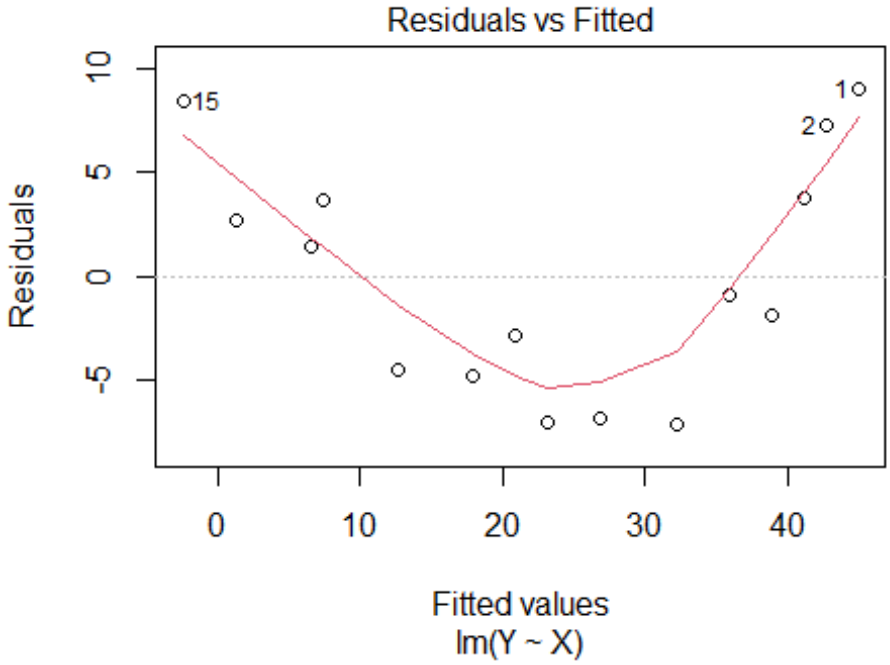
```
model$residuals
```

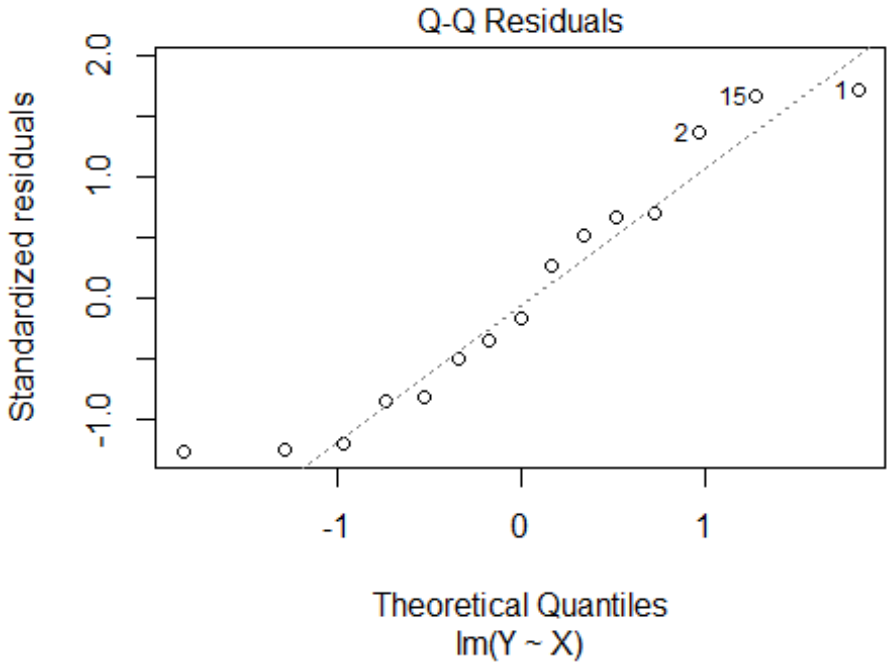
```
##          1          2          3          4          5          6
7
##  9.0446035  7.3021275  3.8071435 -1.9353325 -0.9253005 -7.1627605 -
6.8952045
##          8          9         10         11         12         13
14
## -7.1326645 -2.8751405 -4.8651085 -4.5975525  3.6700035  1.4225115
2.6900675
##         15
##  8.4526075
```

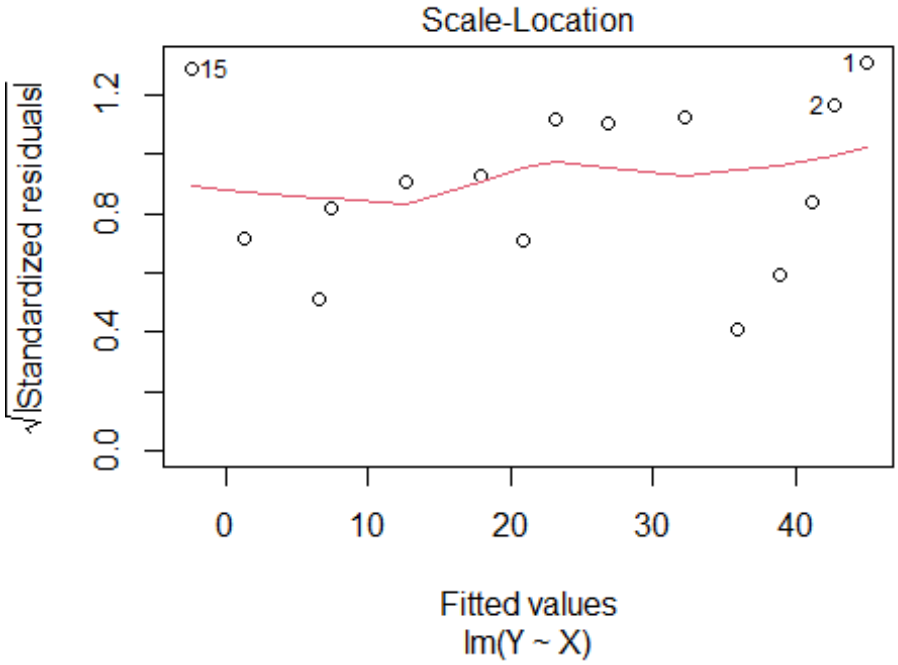
```
plot(data)
```

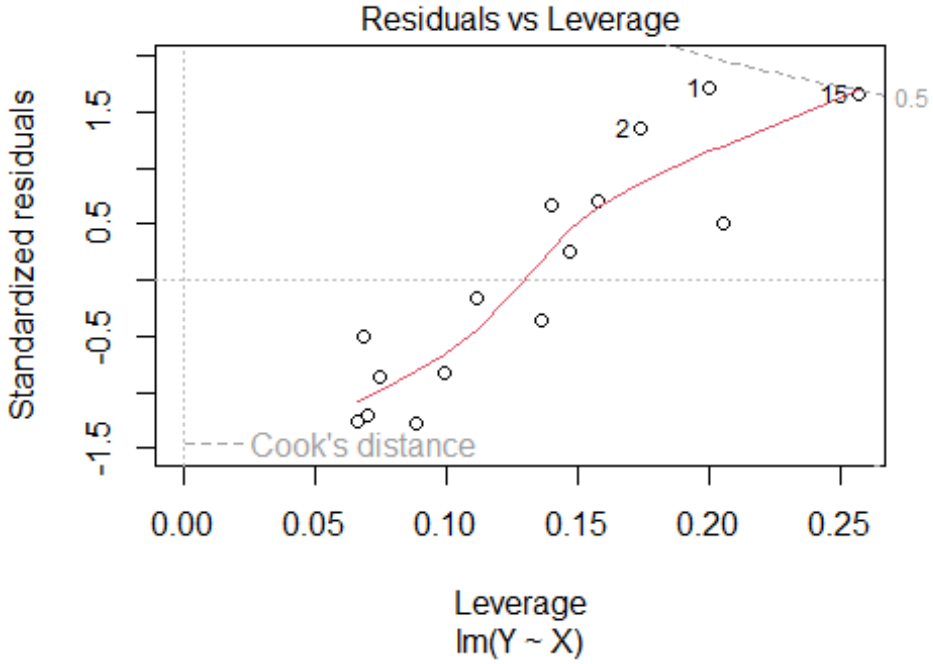



```
plot(model)
```


Residuals vs Fitted
Residuals
Fitted values
lm(Y ~ X)
15
1
2
10
5
0
-5
0
10
20
30
40


Q-Q Residuals
Standardized residuals
Theoretical Quantiles
lm(Y ~ X)
15
1
2
2.0
1.0
0.0
-1.0
-1
0
1


Scale-Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
15
1
2
0
10
20
30
40
Fitted values
lm(Y ~ X)


Residuals vs Leverage
Standardized residuals
-1.5
-0.5
0.5
1.5
1
15
0.5
2
Cook's distance
0.00
0.05
0.10
0.15
0.20
0.25
Leverage
lm(Y ~ X)

## Uji Asumsi

### 1. Gauss Marcov

*a) Nilai harapan galat = 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
```

Menurut uji-t, diperoleh kesimpulan terima H0 dari p-value > 0.05, sehingga asumsi bahwa rataan galat = 0 terpenuhi

*b) Ragam galat homogen*

```
plot(model$residuals)
abline(a = mean(model$residuals), b = 0, col = "blue")

library(lmtest)
```

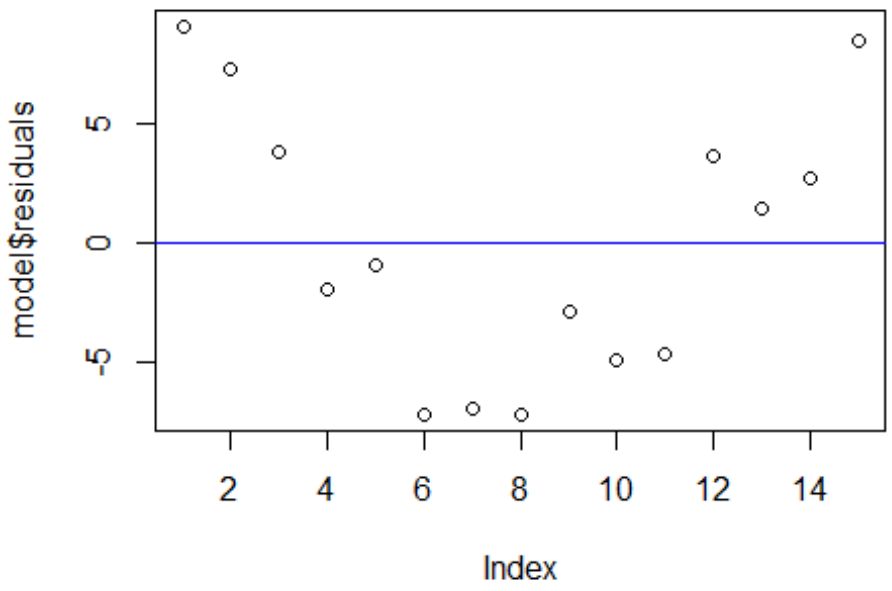
```
## Warning: package 'lmtest' was built under R version 4.3.2
```

```
## 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
```




```
bptest(model)

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

library(car)

## Loading required package: carData

ncvTest(model)

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

Menurut Breusch-Pagan Test dan Non-constant Variance Score Test, diperoleh kesimpulan terima H0, sehingga dapat disimpulkan bahwa asumsi ragam galat homogen terpenuhi. Namun, terdapat pola pada ragam sehingga perlu ditangani dengan cara ditransformasi.

*c) Autokorelasi*

```
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
```

Menurut Runs test dan Durbin-Watson test diperoleh kesimpulan tolak H0, maka dapat disimpulkan bahwa terdapat autokorelasi

### 2. Galat menyebar normal

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

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

```
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
```

Kedua uji menunjukkan kesimpulan terima H0 karena p-value > 0.05, sehingga dapat disimpulkan bahwa galat menyebar normal.

### Penanganan Kondisi Tak Standar dengan Transformasi

Ada asumsi yang dilanggar, maka perlu penanganan. Sebab plot data membentuk pola eksponensial dengan model regresinya adalah $Y = \alpha e^{\beta x}$, maka perlu dilakukan transformasi di mana $Y^* = ln(y)$, $\beta_0 = ln(\alpha)$, dan $\beta_1 = \beta$

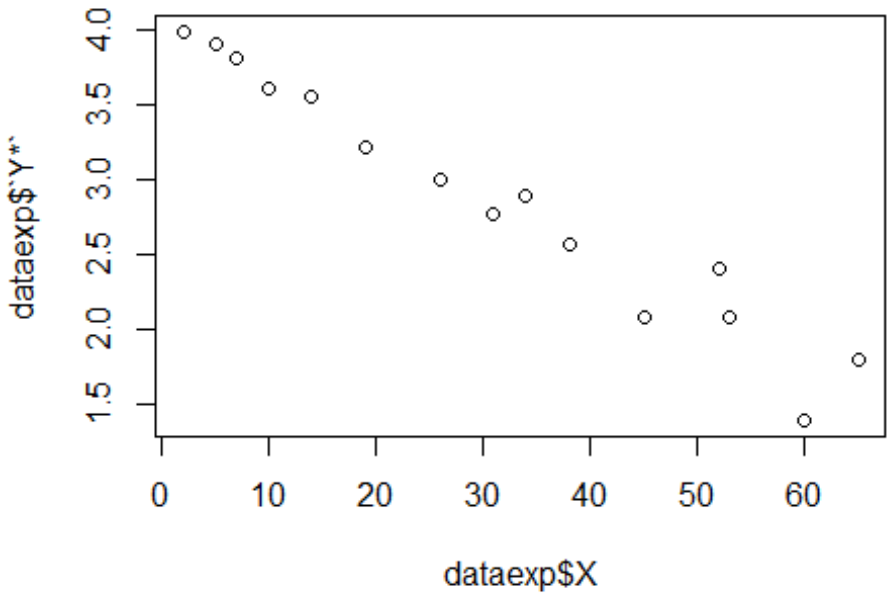
Selain itu, plot data juga dapat membentuk pola polinomial bentuk kuadrat (parabola) dengan model regresinya adalah $Y = \beta_0 + \beta_1 x + \beta_2 x^2$, sehingga perlu ditransformasi di mana $Y^* = \sqrt{Y}$ dan $X^* = \sqrt{X}$

## Transformasi Eksponensial

```
dataexp <- data
dataexp$`Y*` <- log(data$Y)
dataexp
```

```
## # A tibble: 15 × 3
##        X     Y  `Y*`
##    <dbl> <dbl> <dbl>
##  1     2    54  3.99
##  2     5    50  3.91
##  3     7    45  3.81
##  4    10    37  3.61
##  5    14    35  3.56
##  6    19    25  3.22
##  7    26    20  3.00
##  8    31    16  2.77
##  9    34    18  2.89
## 10    38    13  2.56
## 11    45     8  2.08
## 12    52    11  2.40
## 13    53     8  2.08
## 14    60     4  1.39
## 15    65     6  1.79
```

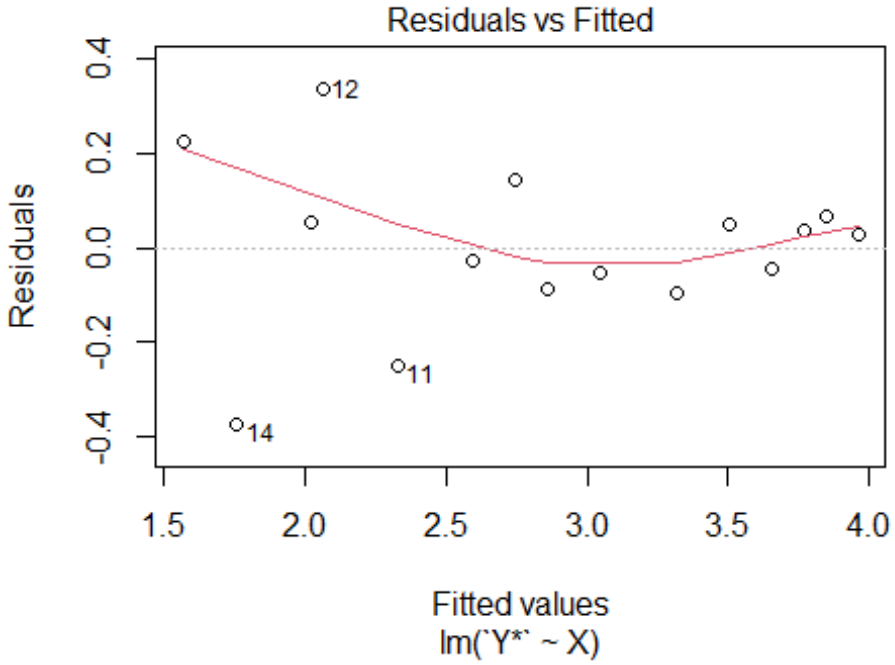
```
plot(dataexp$`X`, dataexp$`Y*`)
```

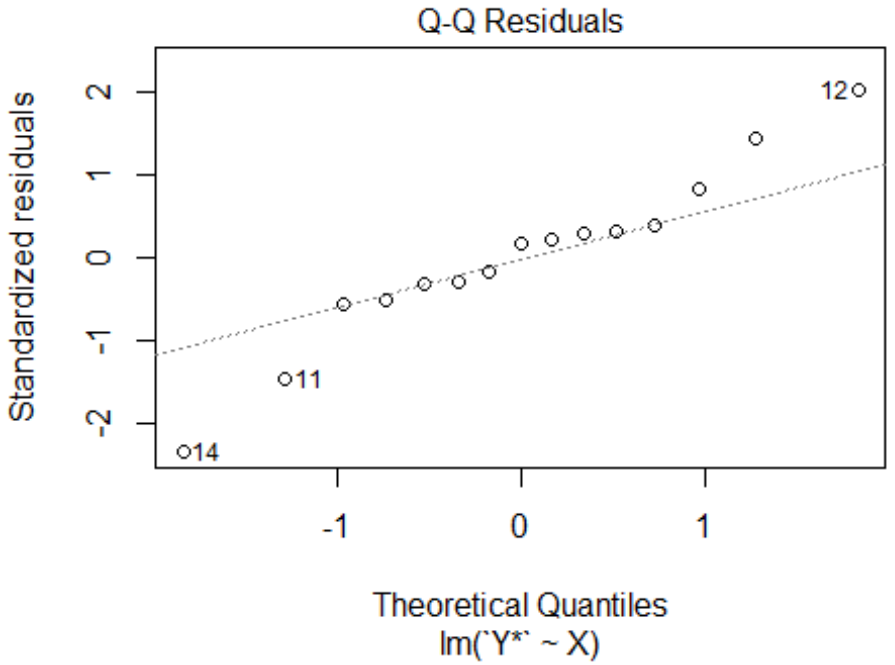


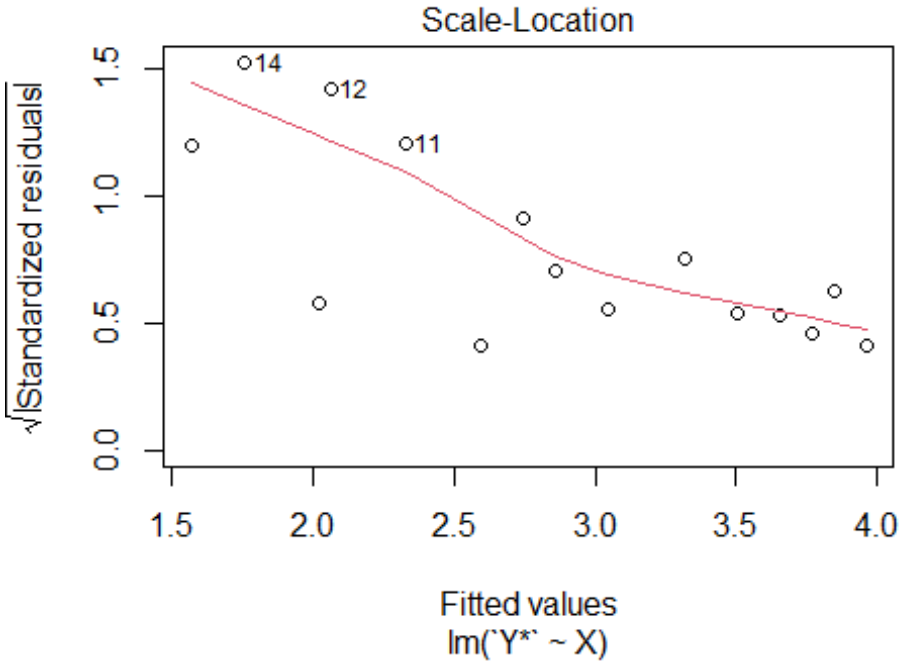
```r
model.exp <- lm(`Y*`~X, dataexp)
summary(model.exp)
```

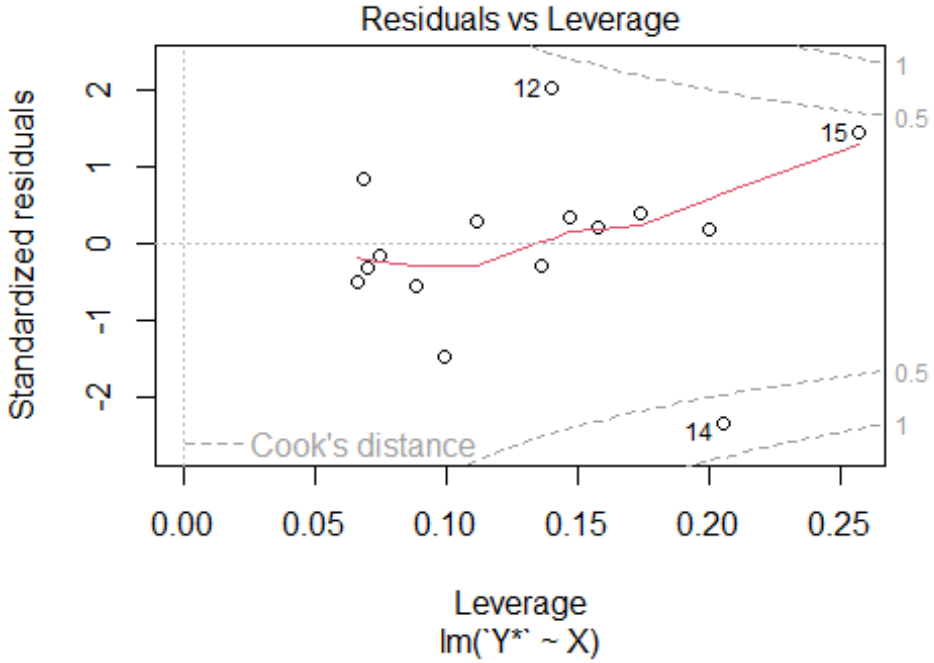
```
##
## Call:
## lm(formula = `Y*` ~ X, data = dataexp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.37241 -0.07073  0.02777  0.05982  0.33539
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.037159   0.084103   48.00 5.08e-16 ***
## X           -0.037974   0.002284  -16.62 3.86e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1794 on 13 degrees of freedom
## Multiple R-squared:  0.9551, Adjusted R-squared:  0.9516
## F-statistic: 276.4 on 1 and 13 DF,  p-value: 3.858e-10
```

```r
plot(model.exp)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
-0.2
-0.4
12
11
14
1.5
2.0
2.5
3.0
3.5
4.0
Fitted values
lm(`Y*` ~ X)


Q-Q Residuals
Standardized residuals
2
1
0
-1
-2
12
11
14
-1
0
1
Theoretical Quantiles
lm(`Y*` ~ X)


Scale-Location
√|Standardized residuals|
14
12
11
0.0
0.5
1.0
1.5
1.5
2.0
2.5
3.0
3.5
4.0
Fitted values
lm(`Y*` ~ X)

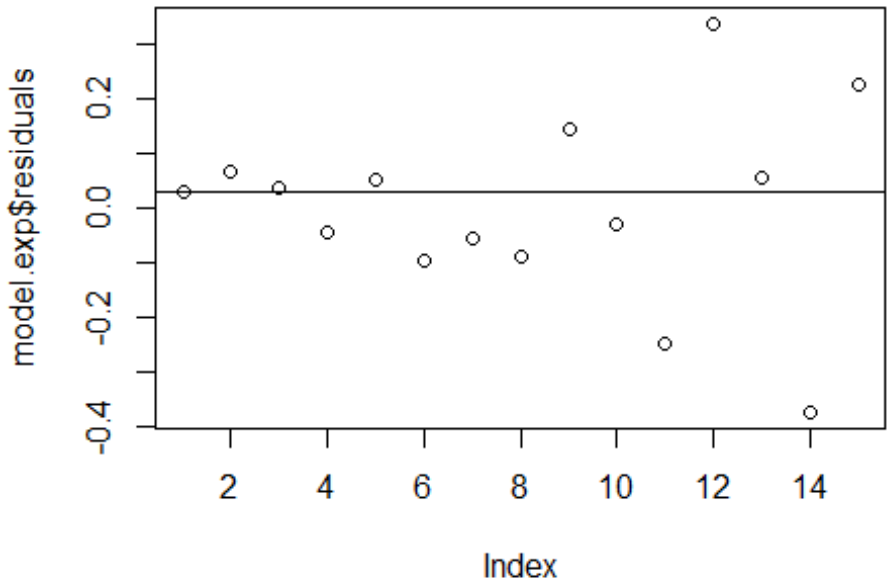

Residuals vs Leverage
Standardized residuals
12
15
14
Cook's distance
1
0.5
0.5
1
2
1
0
-1
-2
0.00
0.05
0.10
0.15
0.20
0.25
Leverage
lm(`Y*` ~ X)

## Pengujian Asumsi Model Transformasi Eksponensial

*1. Gauss-Marcov*

a) Nilai harapan galat = 0

```
plot(model.exp$residuals)
abline(a = model.exp$residuals, b = 0)
```




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

##
##  One Sample t-test
##
## data:  model.exp$residuals
## t = 0, df = 14, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.09572305  0.09572305
## sample estimates:
## mean of x
##         0
```

Menurut uji-t, diperoleh kesimpulan terima H0 dari p-value > 0.05, sehingga asumsi bahwa rataan galat = 0 terpenuhi

b) Ragam galat homogen

```
bptest(model.exp)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model.exp
## BP = 6.9535, df = 1, p-value = 0.008365
```

```
ncvTest(model.exp)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 8.095084, Df = 1, p = 0.0044385
```

Menurut Breusch-Pagan Test dan Non-constant Variance Score Test, diperoleh kesimpulan tolak H0, sehingga dapat disimpulkan bahwa asumsi ragam galat homogen tidak terpenuhi.

### c) Autokorelasi

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

```
##
##  Runs Test
##
## data:  model.exp$residuals
## statistic = 0.55635, runs = 9, n1 = 7, n2 = 7, n = 14, p-value = 0.578
## alternative hypothesis: nonrandomness
```

```
dwtest(model.exp)
```

```
##
##  Durbin-Watson test
##
## data:  model.exp
## DW = 2.7057, p-value = 0.8746
## alternative hypothesis: true autocorrelation is greater than 0
```

Menurut Runs test dan Durbin-Watson test diperoleh kesimpulan terima H0, maka dapat disimpulkan bahwa tidak terdapat autokorelasi

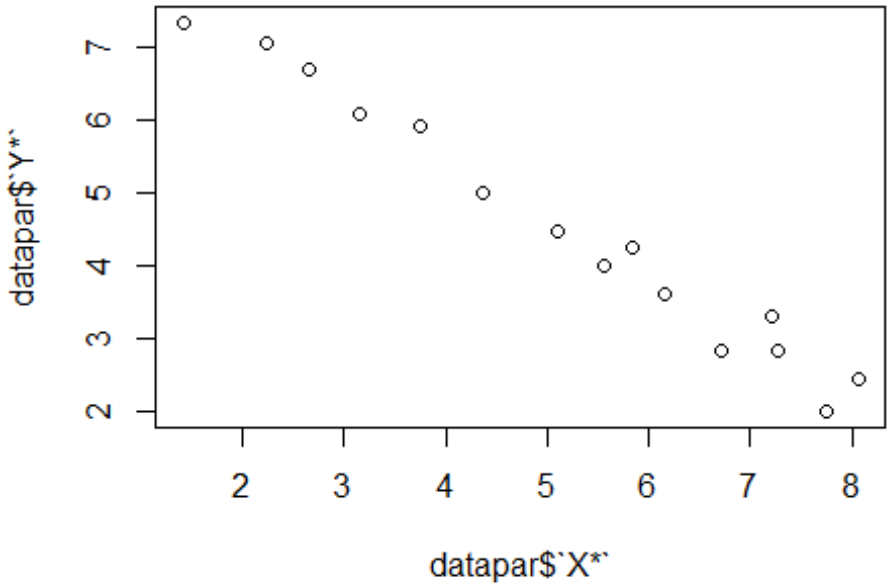
## Transformasi Parabola

```
datapar <- data
datapar$`Y*` <- sqrt(data$Y)
datapar$`X*` <- sqrt(data$X)
datapar
```

```
## # A tibble: 15 × 4
##        X     Y  `Y*`  `X*`
##    <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(datapar$`X*`, datapar$`Y*`)
```




```
model.par <- lm(`Y*`~`X*`, datapar)
summary(model.par)
```

```
##
## Call:
## lm(formula = `Y*` ~ `X*`, data = datapar)
##
## 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 ***
```

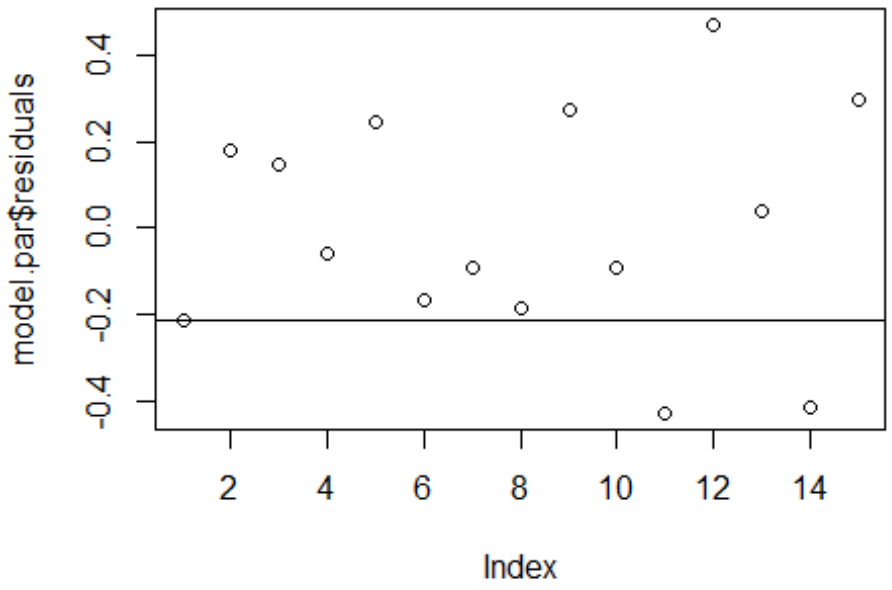
```
## `X*`        -0.81339    0.03445  -23.61 4.64e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2743 on 13 degrees of freedom
## Multiple R-squared:  0.9772, Adjusted R-squared:  0.9755
## F-statistic: 557.3 on 1 and 13 DF,  p-value: 4.643e-12
```

## Pengujian Asumsi Model Transformasi Parabola

*1. Gauss-Marcov*

a) Nilai harapan galat = 0

```
plot(model.par$residuals)
abline(a = model.par$residuals, b = 0)
```




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

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

Menurut uji-t, diperoleh kesimpulan terima H0 dari p-value > 0.05, sehingga asumsi bahwa rataan galat = 0 terpenuhi

### b) Ragam galat homogen

```
bptest(model.par)

##
##  studentized Breusch-Pagan test
##
## data:  model.par
## BP = 3.9621, df = 1, p-value = 0.04654

ncvTest(model.par)

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

Menurut Breusch-Pagan Test, diperoleh kesimpulan tolak H0, sementara untuk Non-constant Variance Score Test diperoleh kesimpulan terima H0. Jika melihat plot galat, terlihat bahwa sebarannya cukup homogen dan tidak berpola, sehingga dapat disimpulkan bahwa asumsi ragam galat homogen terpenuhi.

### c) Autokorelasi

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

##
##  Runs Test
##
## data:  model.par$residuals
## statistic = 0, runs = 8, n1 = 7, n2 = 7, n = 14, p-value = 1
## alternative hypothesis: nonrandomness

dwtest(model.par)

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

Menurut runs test dan Durbin-Watson test diperoleh kesimpulan terima H0, maka dapat disimpulkan bahwa tidak terdapat autokorelasi

### Perbandingan Transformasi Model Terbaik

Dari kedua pengujian transformasi model tersebut, diperoleh bahwa transformasi parabola memenuhi semua asumsi yang dilanggar sebelumnya. Transformasi ini juga memenuhi semua asumsi yang dibutuhkan, sementara untuk transformasi eksponensial terdapat satu asumsi yang dilanggar, yakni homogenitas ragam (ragam tidak homogen), sehingga model yang terbaik adalah model transformasi parabola

*Transformasi Balik Model Parabola*

```
b0 <- model.par$coefficients[[1]]
b1 <- model.par$coefficients[[2]]
b0;b1

## [1] 8.712454

## [1] -0.8133888
```

Oleh karena itu, transformasi dari model sebelumnya adalah

$$Y^* = 8.7124535 - 0.8133888X^*$$

dengan $Y^* = \sqrt{Y}$ dan $X^* = \sqrt{X}$ sehingga transformasi baliknya adalah:

$$\sqrt{Y} = 8.7124535 - 0.8133888\sqrt{X}$$

$$Y = \left(8.7124535 - 0.8133888X^{\frac{1}{2}}\right)^2$$