Tugas Anreg 7

Moch Rizam

2024-03-06

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

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

```
## 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)
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

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

Data

```
data <- read_xlsx("C:/Users/LENOVO/Downloads/Documents/data anreg.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    52     8
## 14    60     4
## 15    65     6
```

```
head(data)
```

```
## # A tibble: 6 × 2
##        x     y
##    <dbl> <dbl>
## 1     2    54
## 2     5    50
## 3     7    45
```

```
## 4    10    37
## 5    14    35
## 6    19    25
```

Model

```
model <- lm(y~., data)
model
```

```
##
## Call:
## lm(formula = y ~ ., data = data)
##
## Coefficients:
## (Intercept)            x
##     46.4978      -0.7554
```

```
summary(model)
```

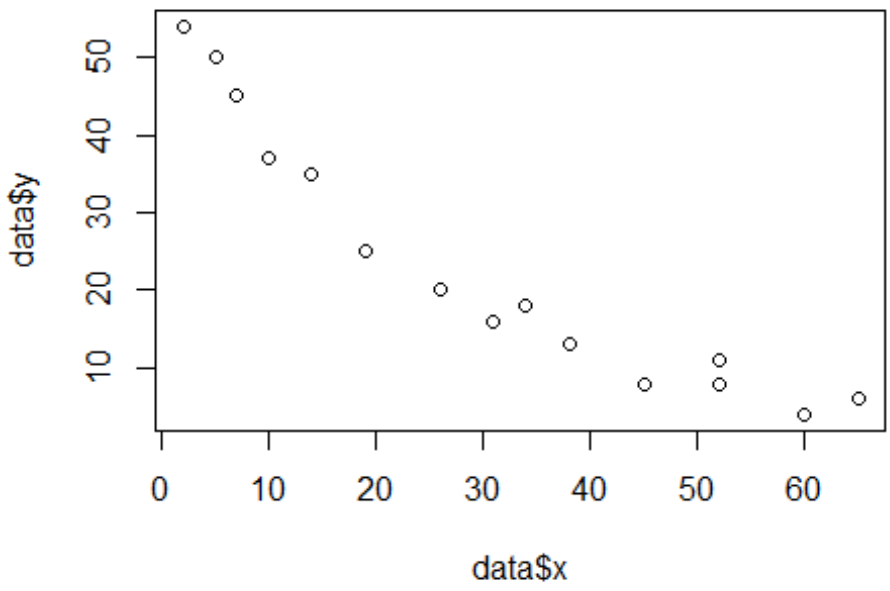
```
##
## Call:
## lm(formula = y ~ ., data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1459 -4.6502 -0.9227  3.7854  9.0130
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 46.49777    2.75973   16.85 3.27e-10 ***
## x           -0.75536    0.07515  -10.05 1.69e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.88 on 13 degrees of freedom
## Multiple R-squared:  0.886,  Adjusted R-squared:  0.8772
## F-statistic:   101 on 1 and 13 DF,  p-value: 1.695e-07
```

Model Regresi :
$\hat{Y} = 46.49777 - 0{,}75536X + e$ model ini belum bisa kita pastikan menjadi suatu model terbaik karena belum memenuhi serangkaian uji asumsi, maka dari itu diperlukan eksplorasi kondisi dan pengujian asumsi dan normalitas untuk menghasilkan model terbaik.

Exploration Of Condition

Exploration of Plot X and Y

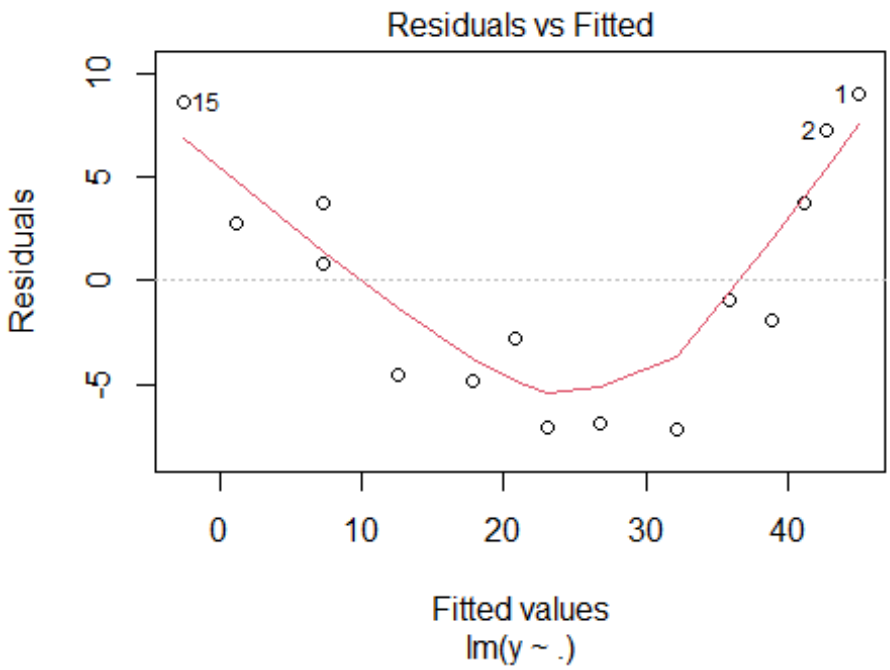
```
plot(x=data$x,y=data$y)
```



Hasil plot ini memiliki hubungan yang tidak linier.

Exploration Of Residual Plot X and Y

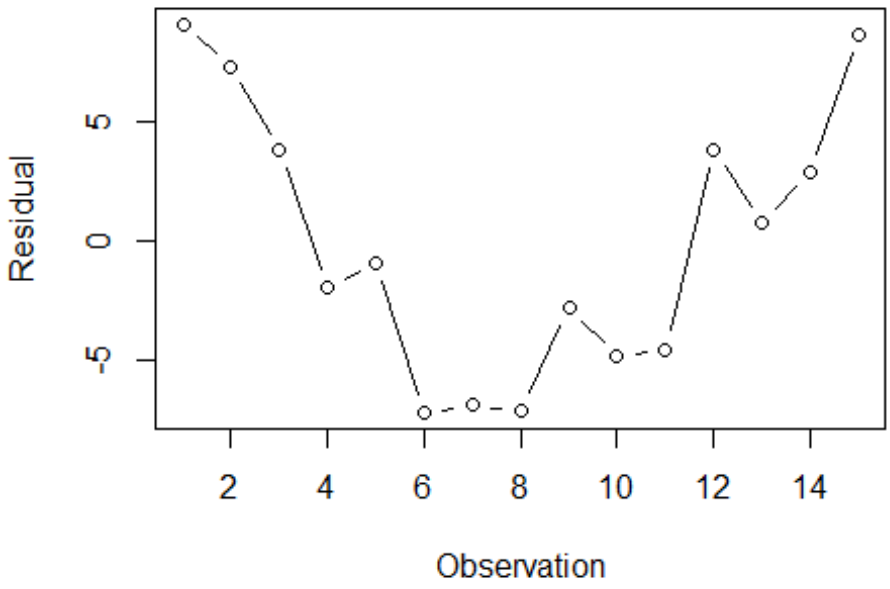
```
plot(model,1)
```



- sisaan menyebar di sekitar 0, maka dari itu galat = 0. - lebar pitanya sama dengan setiap nilai dugaan, sehingga ragamnya homogen. - plotnya membentuk kurva, sehingga modelnya tidak pas dan dibutuhkan transformasi.

Exploration Of Residual Plot Sequences

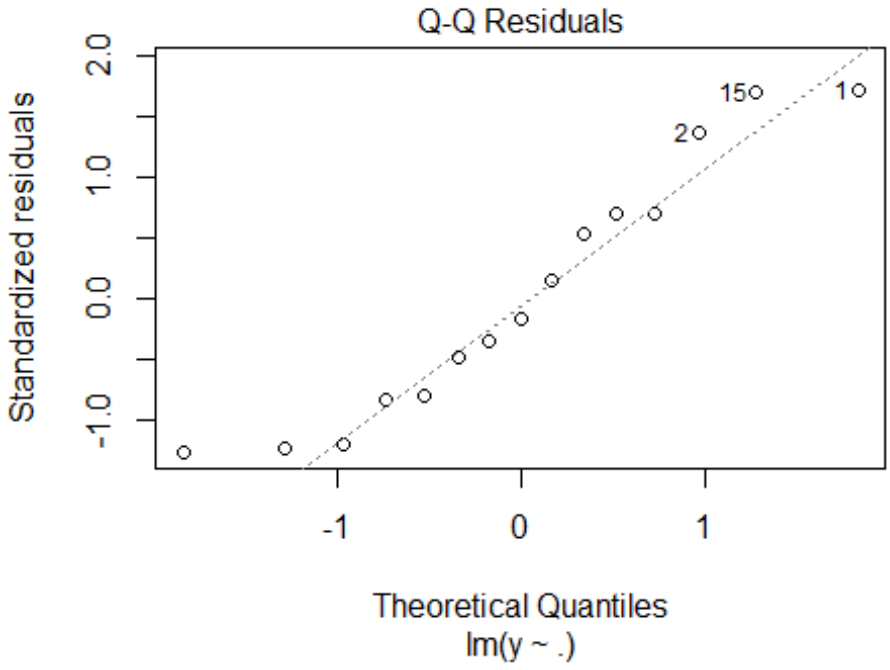
```
plot(x = 1:dim(data)[1],
     y = model$residuals,
     type = 'b',
     ylab = "Residual",
     xlab = "Observation")
```



Sebaran membeuk kurva dan tidak saling bebas.

Exploration Of Residual Normality and QQ-Plot

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



```r
shapiro.test(model$residual)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residual
## W = 0.92411, p-value = 0.2225
```

Datanya cenderung membentuk garis lurus walaupun tidak sepenuhnya lurus ada yang sedikit menjauh dari garis, sehingga data menyebar normal.

Uji Formal

Uji formal asumsi ini diharapkan nilai p-value > 0.05 dengan kesimpulan tidak tolak H0. Nilai harapan sisaan sama dengan 0, H0 ; Nilai harapan sisaan = 0.

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

```
##
##  One Sample t-test
##
## data:  model$residuals
## t = 1.9231e-16, df = 14, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -3.138014  3.138014
## sample estimates:
##    mean of x
## 2.813721e-16
```

Uji ini menunjukan hasil p-value = 1 > alpha = 0.05, maka tidak tolak H0, nilai harapan sisaan = 0 pada taraf 0.05 atau 5% dan asumsi terpenuhi.

Ragam Sisaan Homogen

H0 : var[e] = sigma2I (ragam sisan homogen) H1 : var[e]! = sigma2I (ragam siaan tidak homogen)

```r
homogen = lm(formula = abs(model$residuals)~x,
             data = data)
summary(homogen)
```

```
##
## Call:
## lm(formula = abs(model$residuals) ~ x, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2100 -1.7039 -0.0244  2.0409  4.4581
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.40449    1.30362   4.146  0.00115 **
## x           -0.01941    0.03550  -0.547  0.59371
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.778 on 13 degrees of freedom
## Multiple R-squared:  0.02249,    Adjusted R-squared:  -0.0527
## F-statistic: 0.2991 on 1 and 13 DF,  p-value: 0.5937
```

```
bptest(model)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 0.43085, df = 1, p-value = 0.5116
```

```
ncvTest(model)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.1636521, Df = 1, p = 0.68582
```

Karena p-value 0.68582 > alpha = 0.05, maka tolak H0, ragam sisaan homogen pada taraf nyata 5%.

Sisaan Saling Bebas

H0 : E[ei, ej] = 0 (sisaan saling bebas/tidak ada autokorelasi) H1 : E[ei, ej]! = 0 (sisaan tidak saling bebas/ada autokorelasi)

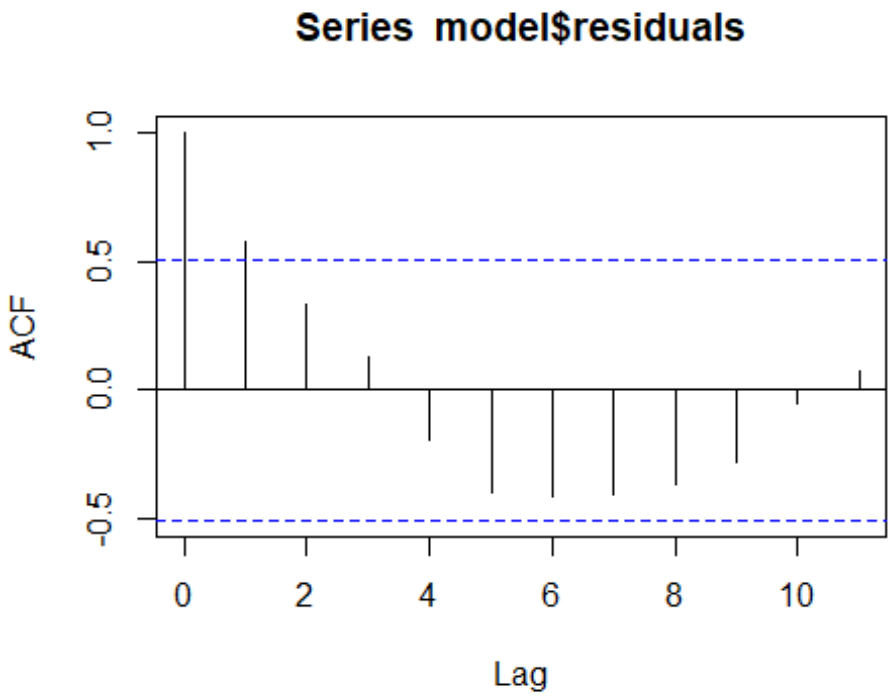
```
runs.test(model$residuals)
```

```
##
##  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 = 0.50138, p-value = 1.85e-05
## alternative hypothesis: true autocorrelation is greater than 0
```

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



Karena p-value = 1.85e-05 pada DW test < alpha = 0.05, mmaka tak tolak H0, sisaan tidak saling bebas pada taraf nyata 5% maka dari itu asumsi tidak terpenuhi.

Uji Formal Normalitas Sisaan

H0 : N (sisaan menyebar Normal) H1 : N (sisaan tidak menyebar Normal)

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

```
##
##  Shapiro-Wilk normality test
##
## data:  model$residuals
## W = 0.92411, p-value = 0.2225
```

```
sisaan_model <- resid(model)
(norm_model <- lillie.test(sisaan_model))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  sisaan_model
## D = 0.12011, p-value = 0.8111
```

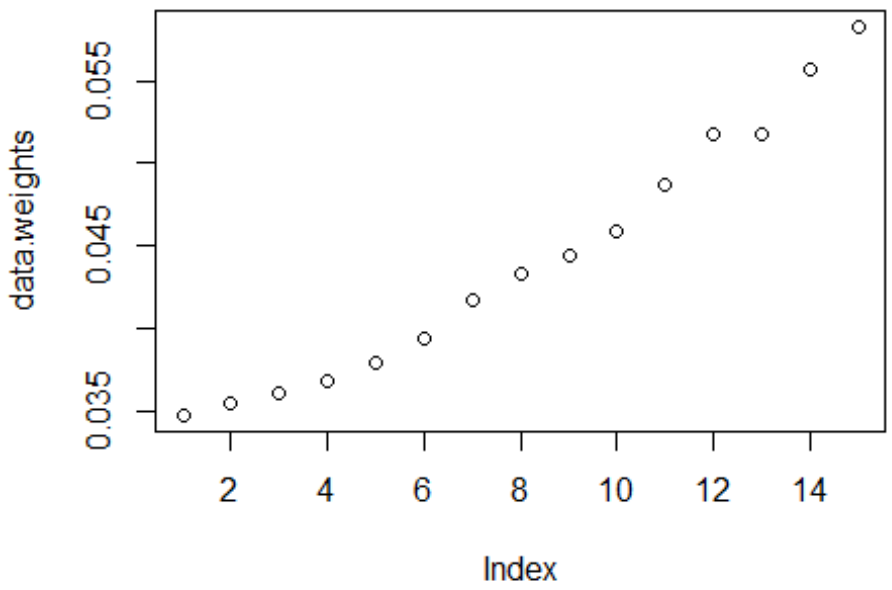
Uji normalitas digunakan bertujuan untuk mendeteksi normalitas sisaan dengan uji shapiro.test dan kolmogrovsmnirov. Krena p-value = 0.8111 > alpha = 0.05, maka tak tolak H0, sehingga menyebar mormal pada taraf 5%.

Metode Weighted Least Squares

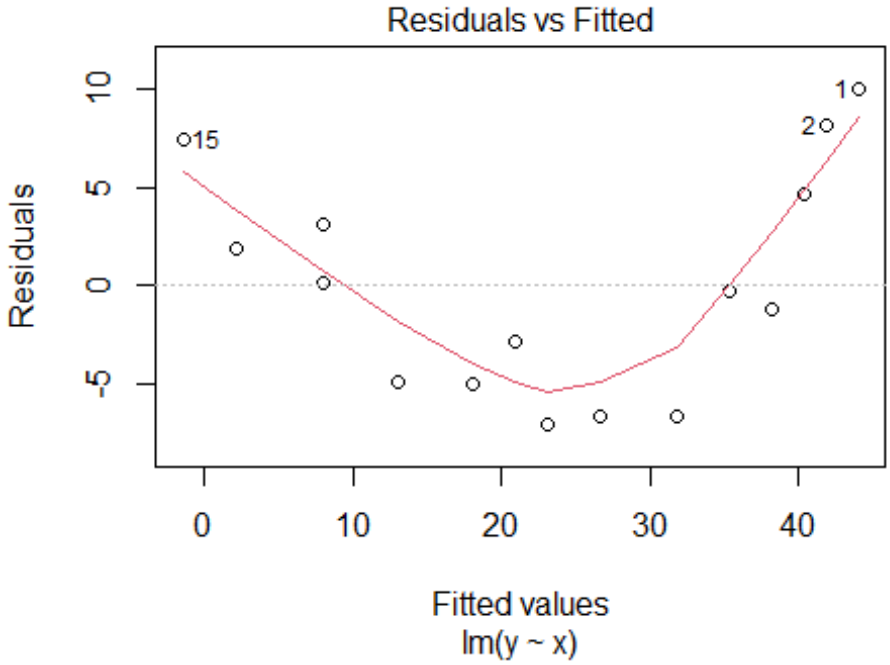
```
resid_abs <- abs(model$residuals)
fitted_val <- model$fitted.values
fit <- lm(resid_abs ~ fitted_val, data)
data.weights <- 1 / fit$fitted.values^2
data.weights
```

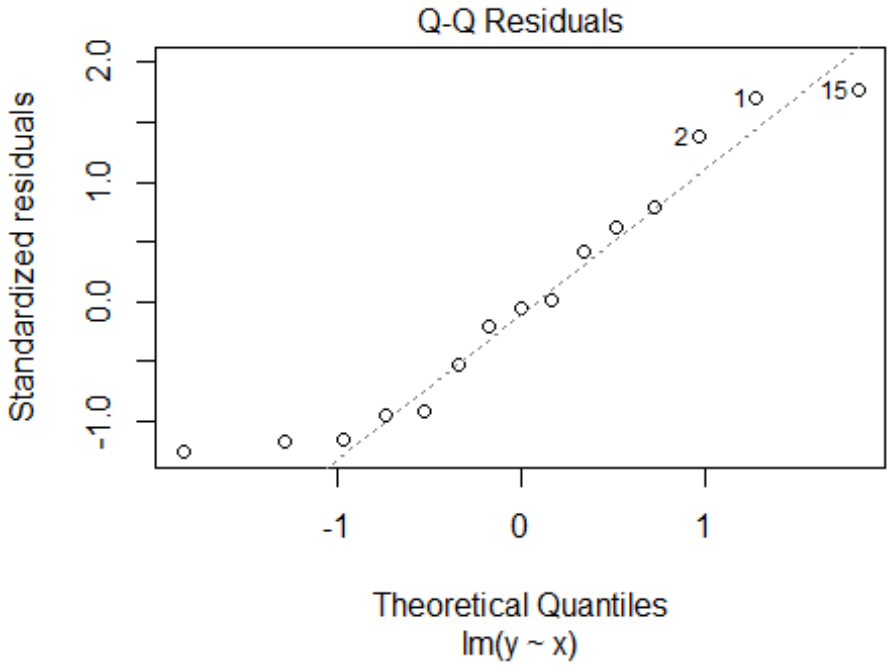
```
##          1          2          3          4          5          6
7
## 0.03473386 0.03550034 0.03602551 0.03683538 0.03795839 0.03943587
0.04165371
##          8          9         10         11         12         13
14
## 0.04335446 0.04442538 0.04591613 0.04871175 0.05177067 0.05177067
0.05563305
##         15
## 0.05827071
```

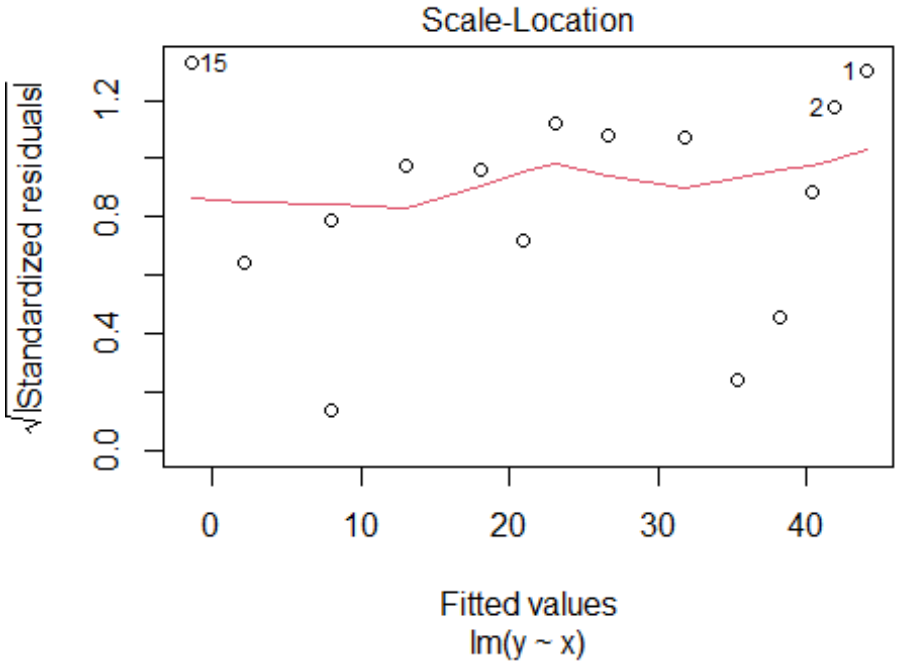
```
plot(data.weights)
```

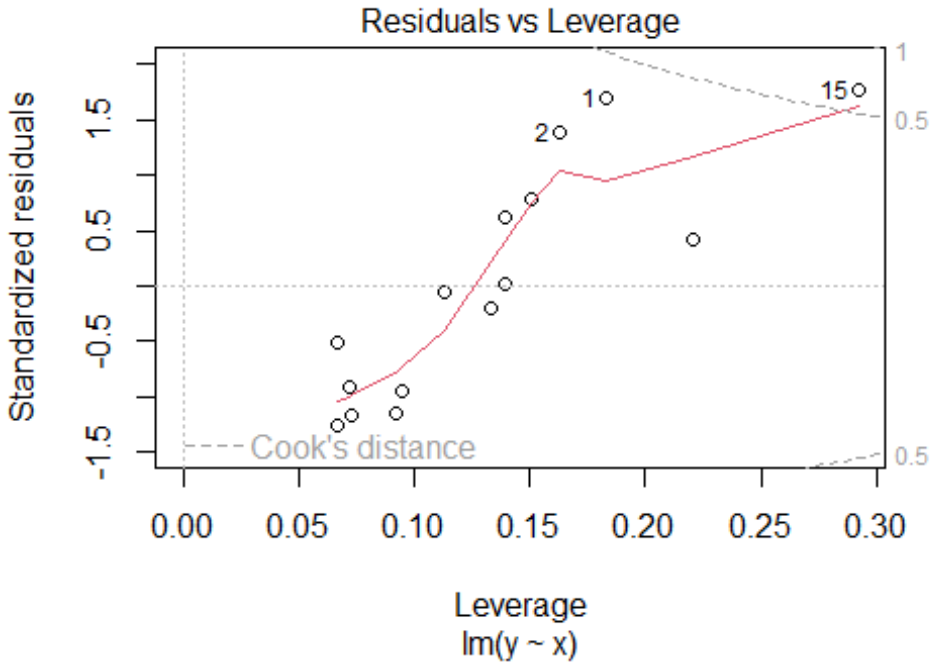



```
model.weighted <- lm(y~x, data = data, weights = data.weights)
plot(model.weighted)
```


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





Hasil ekplorasi ini menggambarkan jika data hasil pembobotan belum mmenuhi uji asumsi.

Model WLS

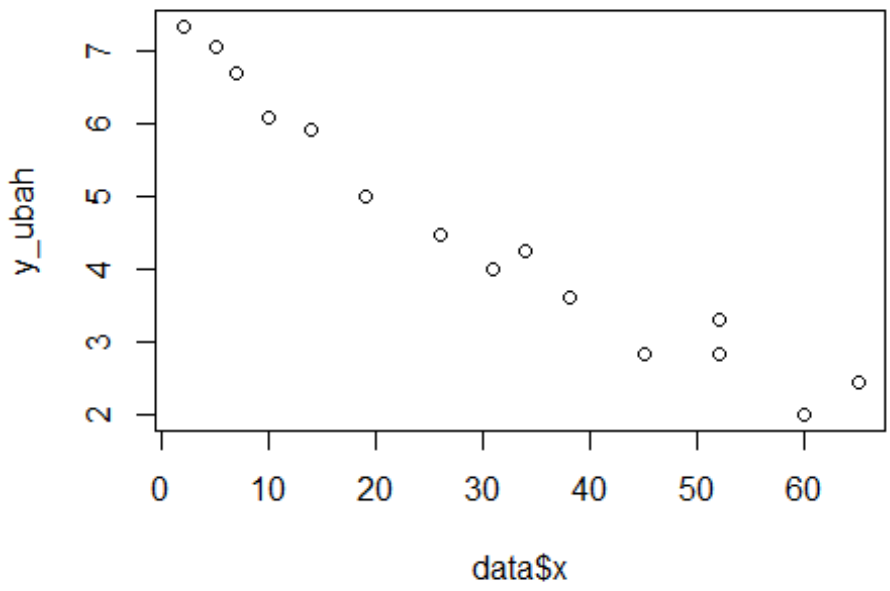
```
model.lmw <- lm(y~x,
data = data,
weights = data.weights)
summary(model.lmw)
```

```
##
## Call:
## lm(formula = y ~ x, data = data, weights = data.weights)
##
## Weighted Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47099 -1.08463 -0.06483  0.78958  1.86509
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.43585    2.90792   15.62 8.34e-10 ***
## x           -0.72165    0.07334   -9.84 2.17e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.214 on 13 degrees of freedom
## Multiple R-squared:  0.8816, Adjusted R-squared:  0.8725
## F-statistic: 96.82 on 1 and 13 DF,  p-value: 2.169e-07
```

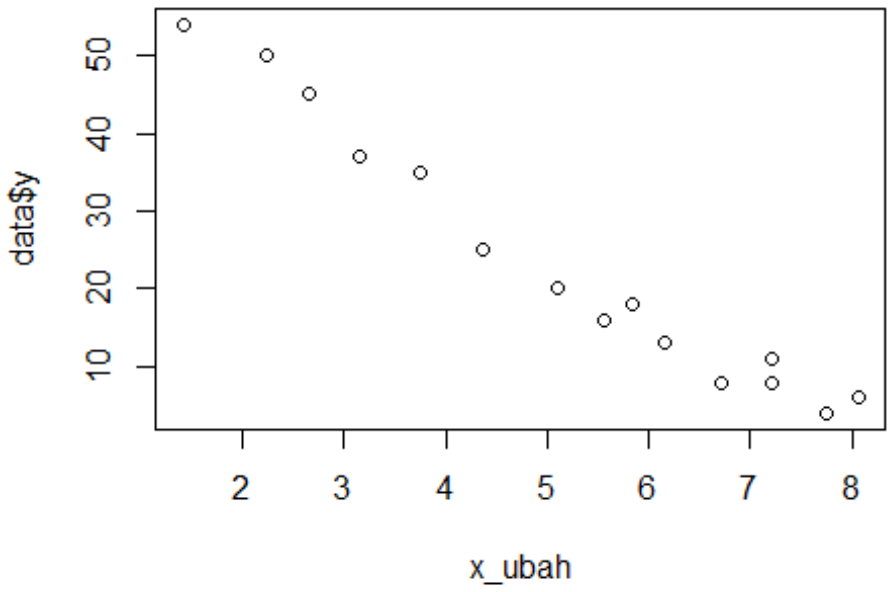
Model regresi : $/hatY = 45.43575 - 0{,}72165X + e$ ## Penyesuaian Data

Transformasi

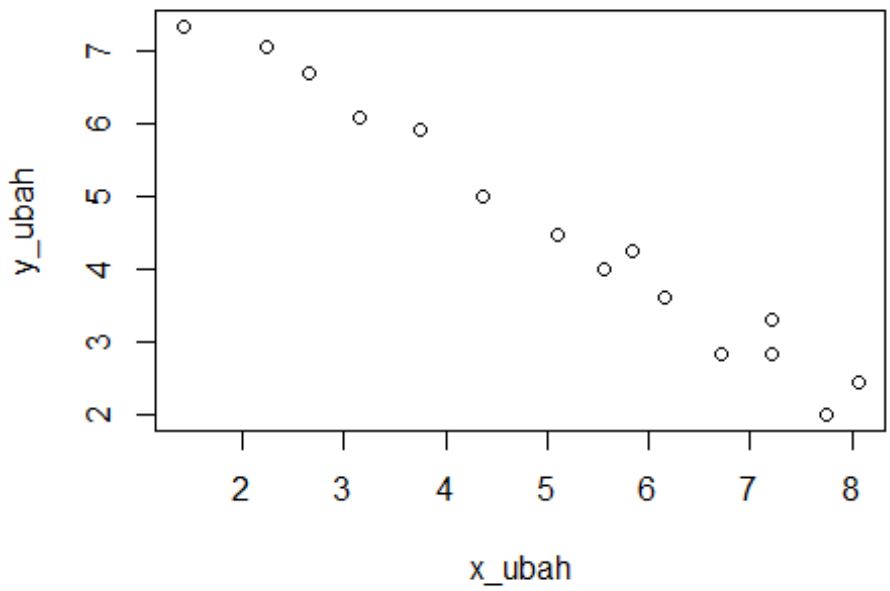
```
y_ubah = sqrt(data$y)
x_ubah = sqrt(data$x)
plot(x = data$x,y = y_ubah)
```




```r
plot(x = x_ubah, y = data$y)
```




```r
plot(x = x_ubah, y = y_ubah)
```




```
data.sqrt <- data.frame(x_ubah, y_ubah)
```

Karena hubungan X dan Y cenderung membentuk sebuah parabola dan nilai B1 < 0, maka data dapat ditransformasi dengan mengecilkan nilai X dan/atau Y dengan membentuknya menjadi pangkat setengah atau akar dari data asli.

Model Pemeriksaan Asumsi

x_ubah dengan y

```
model1 = lm(formula = data$y ~ x_ubah)
summary(model1)
```

```
##
## Call:
## lm(formula = data$y ~ x_ubah)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4310 -2.7959  0.7218  1.9806  5.3327
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.2861     2.2656   27.93 5.42e-13 ***
## x_ubah       -7.7669     0.4091  -18.98 7.32e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 3.25 on 13 degrees of freedom
## Multiple R-squared:  0.9652, Adjusted R-squared:  0.9625
## F-statistic: 360.4 on 1 and 13 DF,  p-value: 7.322e-11
```

```
dwtest(model1)
```

```
##
##  Durbin-Watson test
##
## data:  model1
## DW = 1.1529, p-value = 0.017
## alternative hypothesis: true autocorrelation is greater than 0
```

x dengan y_ubah

```
model2 = lm(formula = x_ubah ~ data$x)
summary(model2)
```

```
##
## Call:
## lm(formula = x_ubah ~ data$x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.86733 -0.23103  0.08191  0.32116  0.42102
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 2.081840   0.185910   11.20 4.77e-08 ***
## data$x      0.099852   0.005062   19.73 4.52e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3961 on 13 degrees of freedom
## Multiple R-squared:  0.9677, Adjusted R-squared:  0.9652
## F-statistic: 389.1 on 1 and 13 DF,  p-value: 4.524e-11
```

```
dwtest(model2)
```

```
##
##  Durbin-Watson test
##
## data:  model2
## DW = 0.28639, p-value = 2.783e-08
## alternative hypothesis: true autocorrelation is greater than 0
```

```
model3 = lm(formula = y_ubah ~ x_ubah)
summary(model3)
```

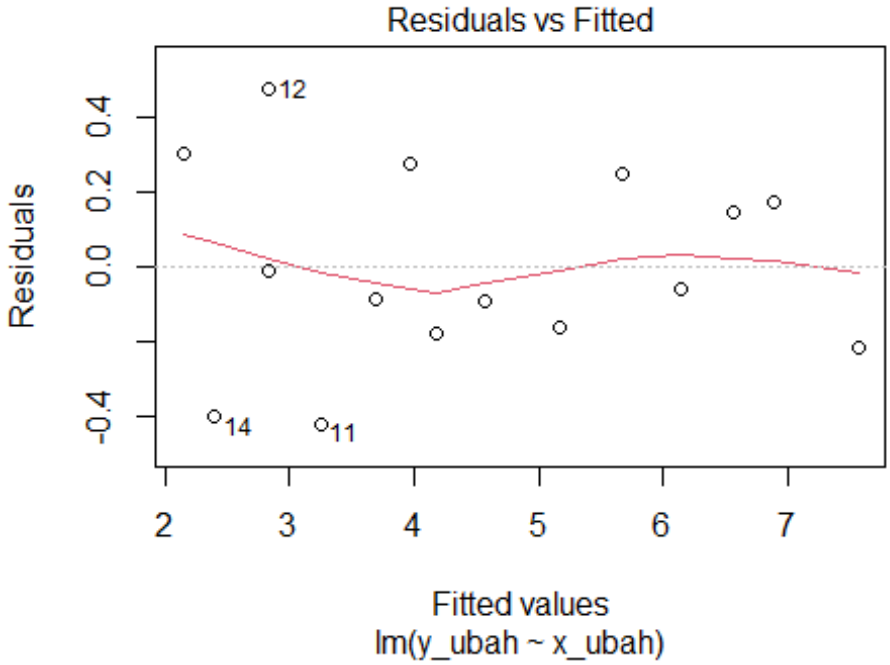
```
##
## Call:
## lm(formula = y_ubah ~ x_ubah)
```

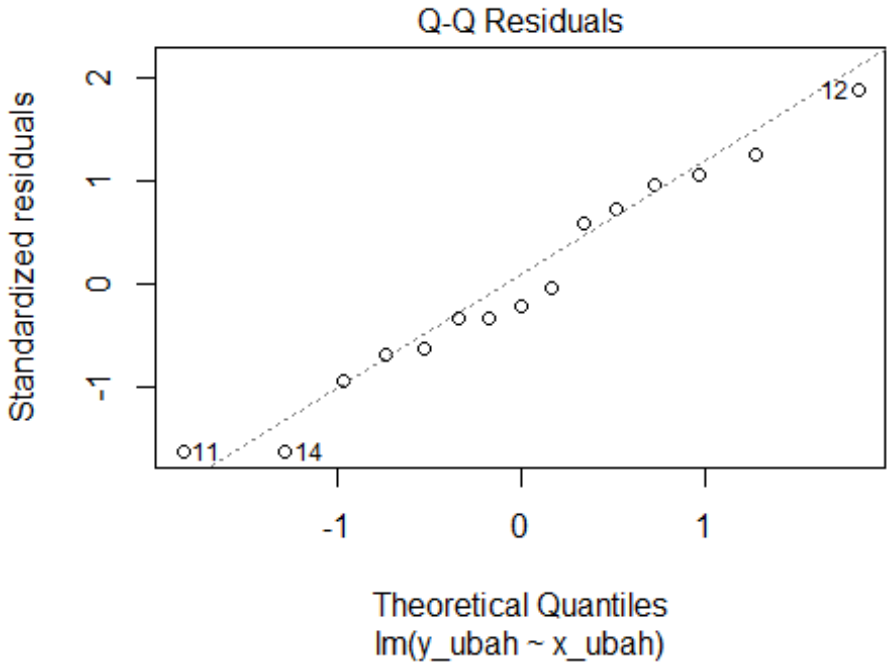
```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42097 -0.17193 -0.05751  0.21192  0.47723
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.71838    0.19109   45.62  9.8e-16 ***
## x_ubah      -0.81527    0.03451  -23.63  4.6e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2741 on 13 degrees of freedom
## Multiple R-squared:  0.9772, Adjusted R-squared:  0.9755
## F-statistic: 558.2 on 1 and 13 DF,  p-value: 4.597e-12
```

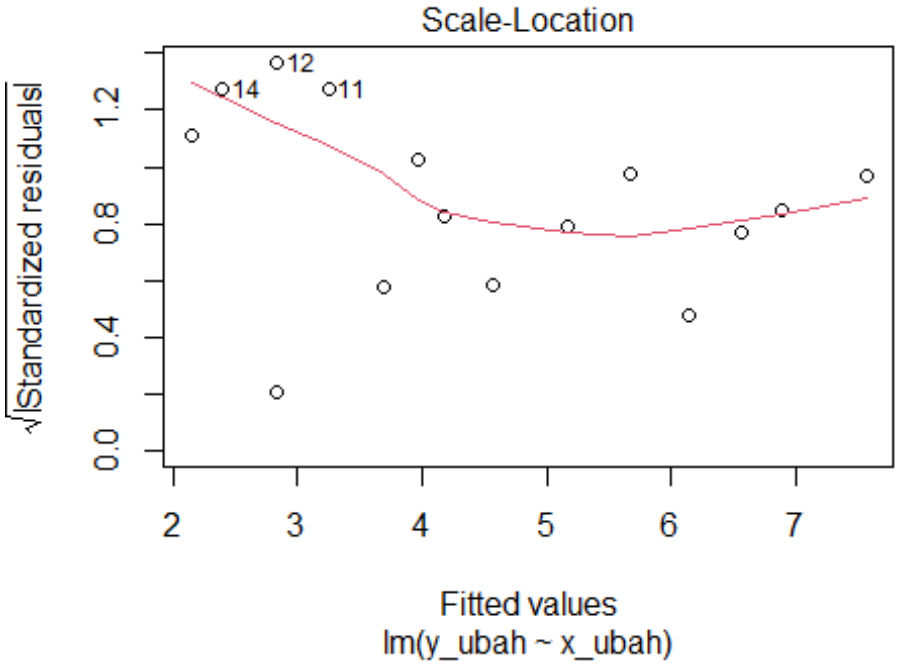
```
dwtest(model3)
```

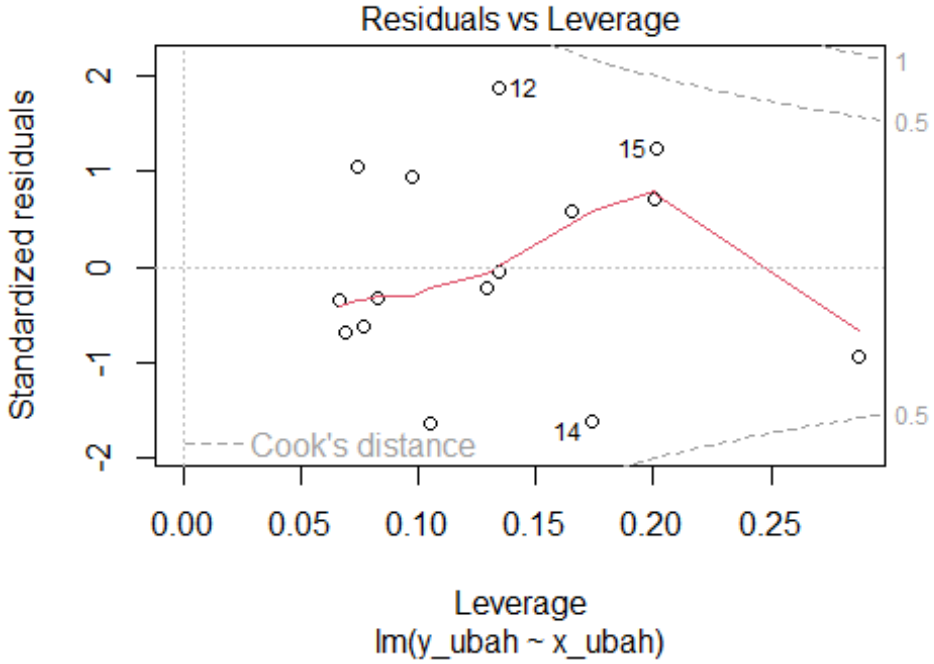
```
##
##  Durbin-Watson test
##
## data:  model3
## DW = 2.6905, p-value = 0.8678
## alternative hypothesis: true autocorrelation is greater than 0
```

```
plot(model3)
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
-0.4
12
14
11
2
3
4
5
6
7
Fitted values
lm(y_ubah ~ x_ubah)


Q-Q Residuals
Standardized residuals
2
1
0
-1
12
11
14
-1
0
1
Theoretical Quantiles
lm(y_ubah ~ x_ubah)






```
t.test(model3$residuals,mu = 0,conf.level = 0.95)

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

```
##
## data:  model3$residuals
## t = 2.035e-16, df = 14, p-value = 1
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.1462664  0.1462664
## sample estimates:
##    mean of x
## 1.387779e-17
```

```
ncvTest(model3)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 2.188518, Df = 1, p = 0.13904
```

```
sisaan.model3 <- resid(model3)
(norm.model3 <- lillie.test(sisaan.model3))
```

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  sisaan.model3
## D = 0.11952, p-value = 0.8166
```

karena p-value dari semua uji lebih dari 0.05 maka asumsi terpenuhi.

Kesmpulan

Hasil model terbaik dipenuhi ketika variabel X dan Y keduanya ditransformasi ke dalam bentuk akar atau pangkat 1/2 dan memenuhi semua asumsi dalam analisis regresi linier sederhana. Sehingga model untuk data ini adalah

$Y^{\frac{1}{2}}$ = 8.71838 - 0.81527

di transformasi menjadi

$$/hatY = (8.71838 - 0.81527)^2$$

Bisa kita interpretasi terhadap model tersebut menunjukkan hubungan yang terbalik antara Y dengan X sebagai hubungan kuadratik.