

Pertemuan 7 Analisis Regresi

Nabil Bintang Prayoga/G1401221017

2024-03-05

Data

Inisialisasi 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)
```

```
## 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)
```

```
## 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
```

```
library(car)
```

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

```
##
```

```
## Attaching package: 'car'
```

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

```
##
```

```
##      recode
```

```
library(randtests)
```

```
library(lmtest)
```

Import Data

```
data <- read_xlsx("D:\\KULIAHH\\SEMESTER 4\\ANREG\\DATA ANREG.xlsx")
data
```

```
## # A tibble: 15 x 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 Awal

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

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

Diperoleh model sebagai berikut

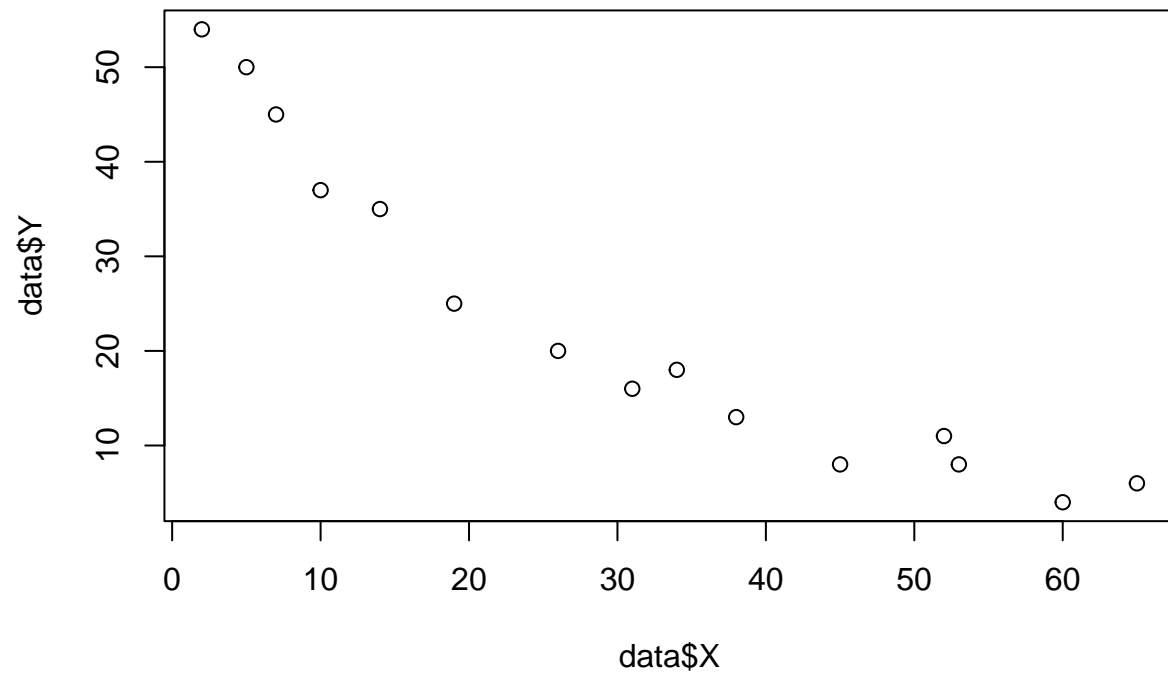
$$\hat{Y} = 46.46041 - 0.7525X + e$$

Hasil tersebut belum bisa dipastikan menjadi model terbaik karena belum melalui serangkaian uji asumsi, sehingga diperlukan eksplorasi kondisi dan pengujian asumsi Gaus Markov dan normalitas untuk menghasilkan model terbaik

Eksplorasi Kondisi

Plot Hubungan X dan Y

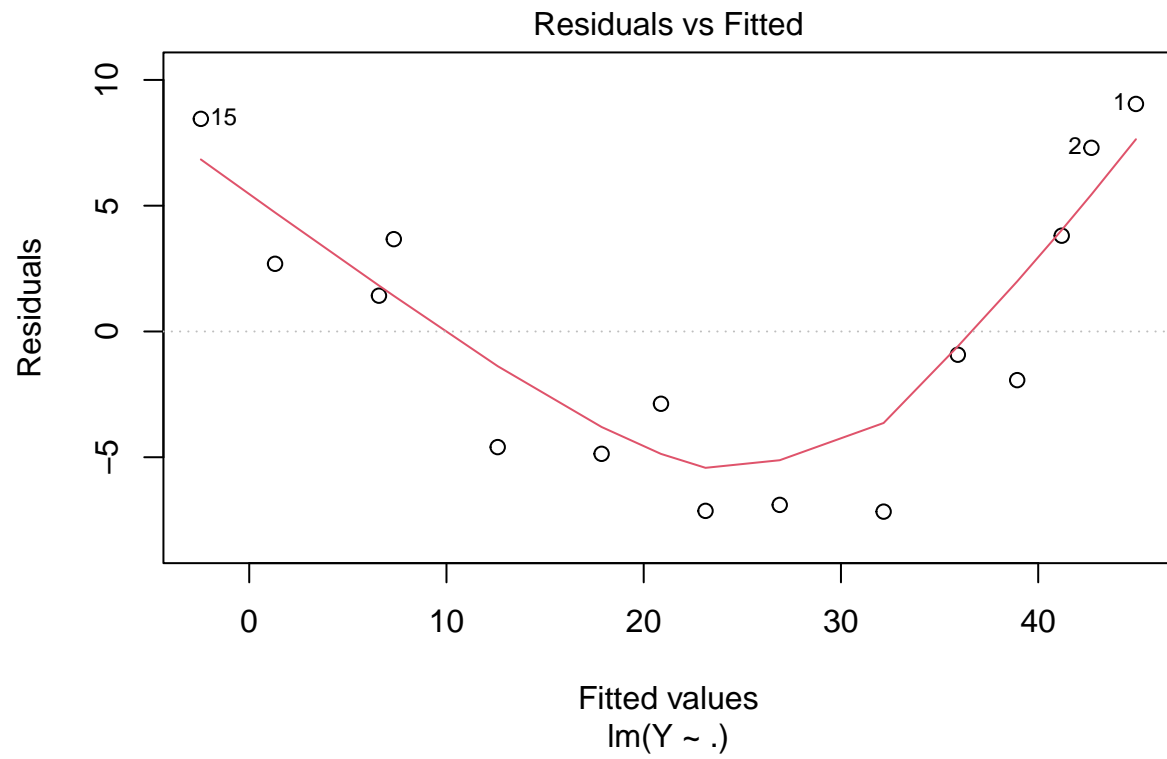
```
plot(x = data$X, y = data$Y)
```



Hasil plot hubungan X dan Y menggambarkan hubungan yang tidak linier dan cenderung membentuk parabola

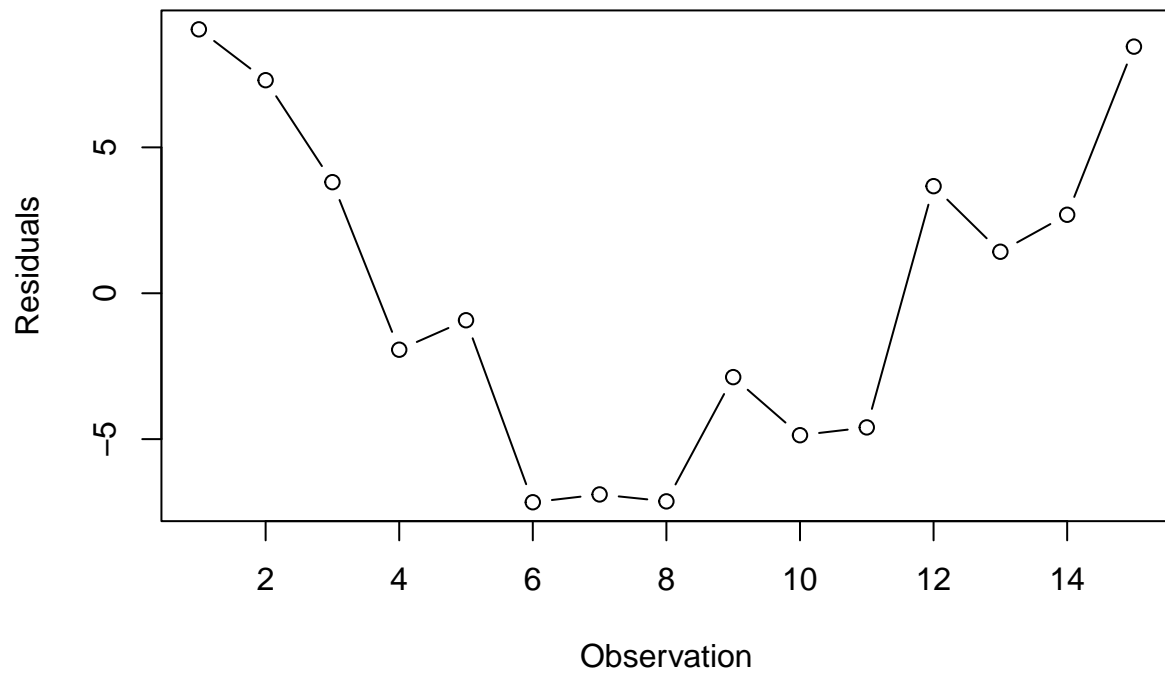
Plot Sisaan Vs Y duga

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



Plot Sisaan Vs Urutan

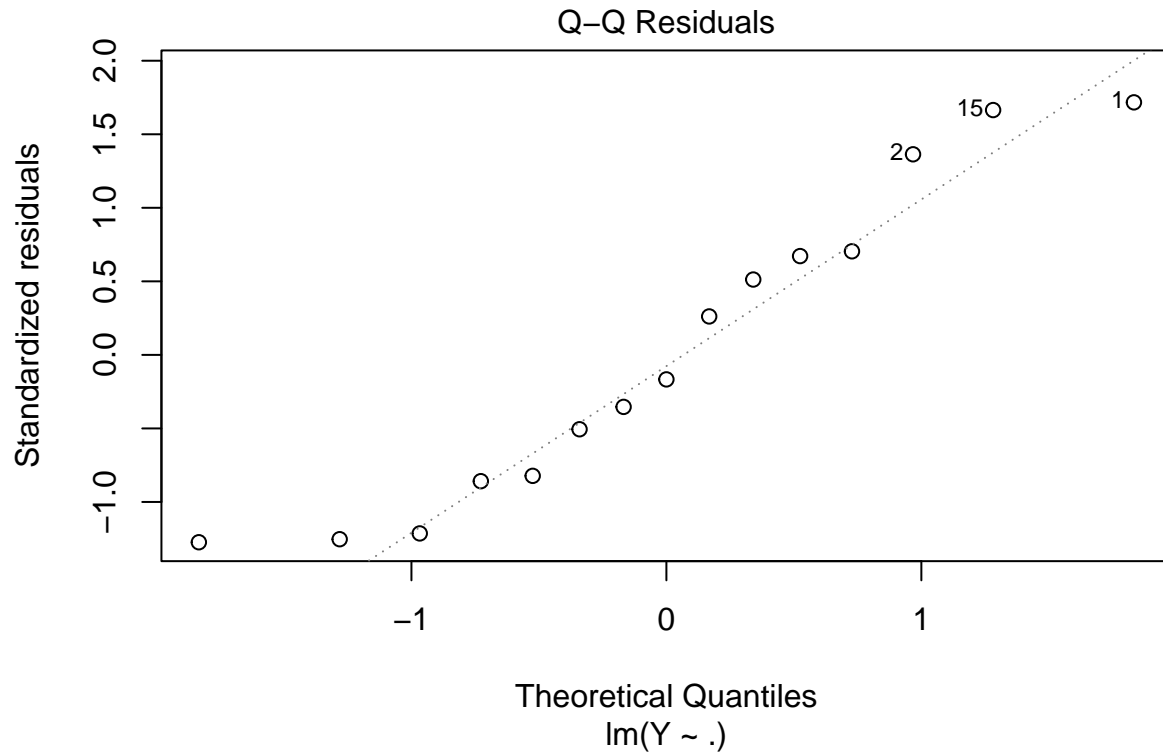
```
plot(x = 1:dim(data)[1],
     y = model$residuals,
     type = 'b',
     ylab = "Residuals",
     xlab = "Observation")
```



Tebaran membentuk pola kurva → sisaan tidak saling bebas, model tidak pas

Normalitas Sisaan dengan QQ-Plot

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



Uji Formal Asumsi

p-value < 0.05 tolak H_0

Kondisi Gaus Markov ### 1. Nilai Harapan Sisaan sama dengan Nol H_0 : Nilai harapan sisaan sama dengan nol

H_1 : Nilai harapan sisaan tidak sama dengan nol

```
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
```

Karena p-value = 1 > alpha = 0.05, maka tak tolak H_0 , nilai harapan sisaan sama dengan nol

2. Ragam Sisaan Homogen

H0: Ragam sisaan homogen

H1: Ragam sisaan tidak homogen

```
kehomogenan = lm(formula = abs(model$residuals) ~ X, # y: abs residual
  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
```

```
ncvTest(model)
```

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

Karena $p\text{-value} = 0.4674 > \alpha = 0.05$, maka tak tolak H0, ragam sisaan homogen

3. Sisaan Saling Bebas

H0: Sisaan saling bebas

H1: Sisaan tidak saling bebas


```
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
```

Karena $p\text{-value} = 1.333e-05$ (pada DW test) $< \alpha = 0.05$, maka tolak H_0 , sisaan tidak saling bebas, asumsi tidak terpenuhi

Uji Formal Normalitas Sisaan

H_0 : Sisaan menyebar normal

H_1 : Sisaan tidak menyebar normal

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

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

Karena $p\text{-value} = 0.226 > \alpha = 0.05$, maka tak tolak H_0 , sisaan menyebar normal

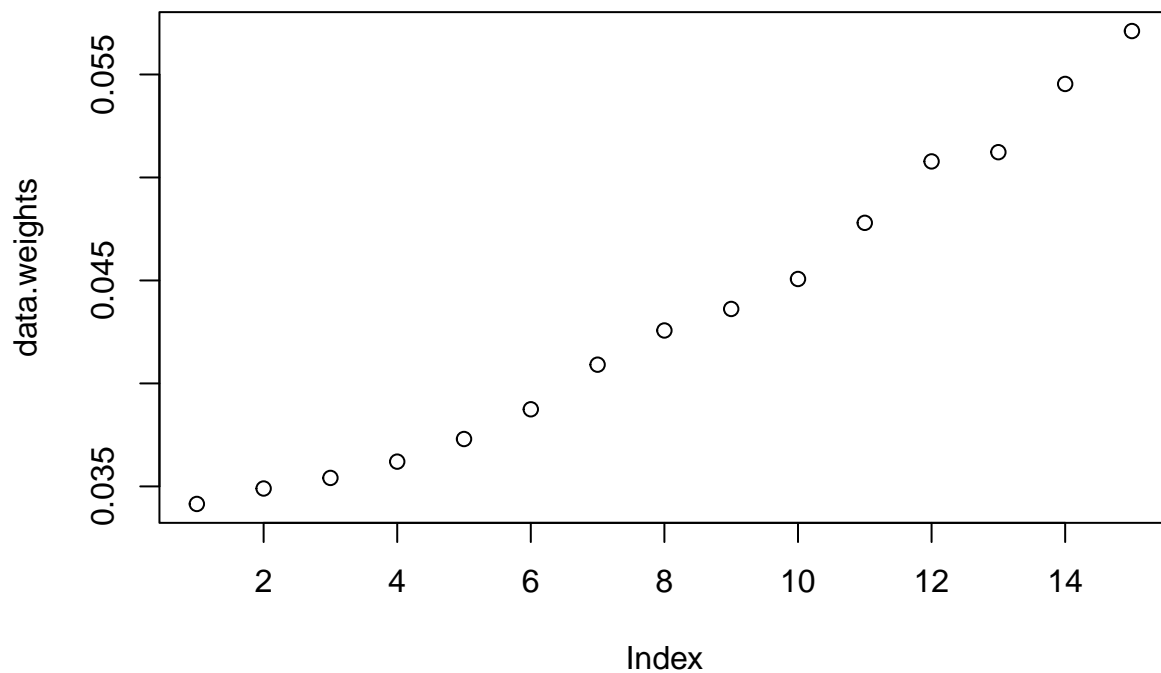
Metode Weighted Least Squares

Pembobotan Data

```
resid_abs <- abs(model$residuals)  
fitted_val <- model$fitted.values  
fit <- lm(resid_abs ~ fitted_val)  
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 WLS

Langkah ini hanya untuk membandingkan saja, karena sisaan ragam sudah homogen maka sebenarnya tidak perlu dilakukan pembobotan dan langsung transformasi data saja

```
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:
```

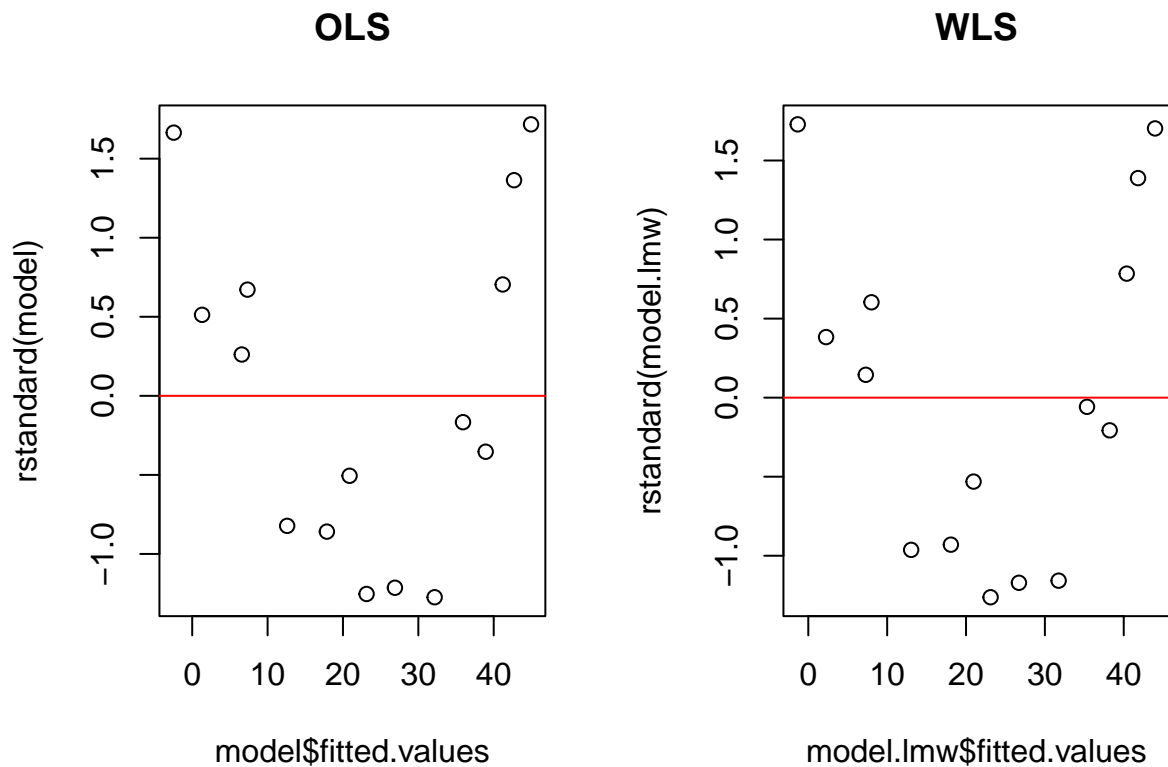
```
##      Min      1Q   Median      3Q      Max
## -1.46776 -1.09054 -0.06587  0.77203  1.85309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 45.41058    2.90674   15.623 8.35e-10 ***
## X           -0.71925    0.07313   -9.835 2.18e-07 ***
## ---
## Signif. 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
```

Diperoleh model sebagai berikut

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

Perbandingan OLS Vs WLS

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



```
par(mfrow=c(1,1))
```

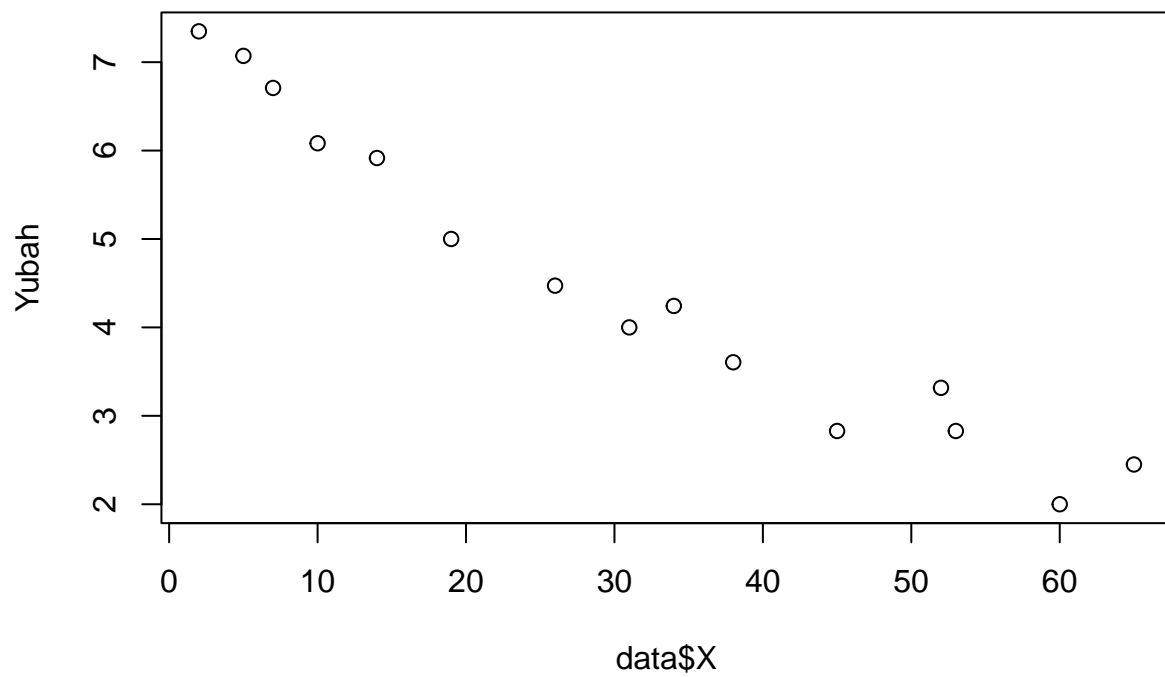
Karena data telah memiliki ragam sisaan yang homogen, maka model dan plot yang terbentuk tidak berubah signifikan sehingga model masih dikatakan belum pas. Sebab metode WLS umumnya dilakukan untuk data yang memiliki ragam sisaan tidak homogen.

Penyesuaian Data

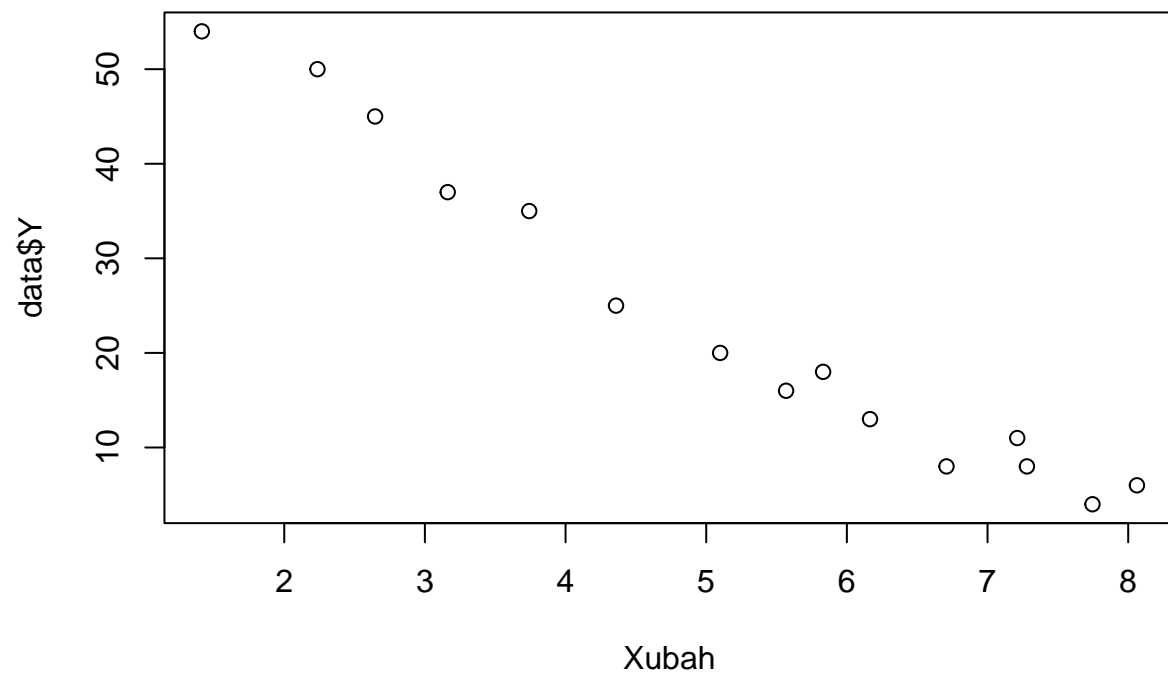
Transformasi Data

```
Yubah = sqrt(data$Y)
Xubah = sqrt(data$X)

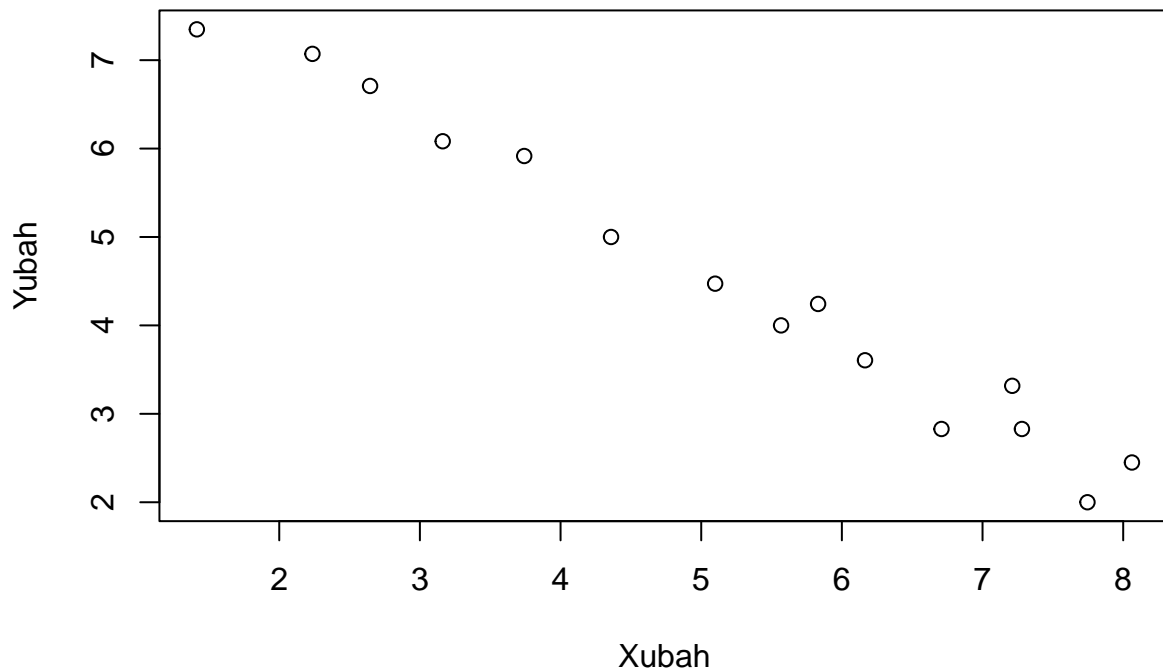
plot(x = data$X, y = Yubah)
```



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



```
plot(x = Xubah, y = Yubah)
```



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. Terdapat perbedaan antara hasil plot hubungan Xubah dengan Y, X dengan Yubah, dan Xubah dengan Yubah sehingga perlu ditelusuri lebih lanjut untuk memperoleh model terbaik melalui pemeriksaan asumsi pada data dengan sisaan paling bebas

Model dan Pemeriksaan Asumsi

Xubah dengan Y

```
model1 = lm(formula = data$Y ~ Xubah)
summary(model1)
```

```
##
## Call:
## lm(formula = data$Y ~ Xubah)
##
## 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 ***
```

```
## Xubah          -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
```

Diperoleh model sebagai berikut

$$\hat{Y} = 63.2250 - 0.77481X + e$$

```
dwtest(model1)
```

```
##
## Durbin-Watson test
##
## data: model1
## DW = 1.1236, p-value = 0.01422
## alternative hypothesis: true autocorrelation is greater than 0
```

Karena p-value = 0.01422 (pada DW test) < alpha = 0.05, maka tolak H0, sisaan tidak saling bebas, asumsi tidak terpenuhi, bukan model terbaik

X dengan Yubah

```
model2 = lm(formula = Yubah ~ data$X)
summary(model2)
```

```
##
## Call:
## lm(formula = Yubah ~ data$X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53998 -0.38316 -0.01727  0.36045  0.70199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.015455   0.201677   34.79 3.24e-14 ***
## data$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
```

Diperoleh model sebagai berikut

$$\hat{Y} = 7.015455 - 0.081045X + e$$


```
dwtest(model2)
```

```
##
## Durbin-Watson test
##
## data: model2
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0
```

Karena p-value = 0.02493 (pada DW test) < alpha = 0.05, maka tolak H0, sisaan tidak saling bebas, asumsi tidak terpenuhi, bukan model terbaik

Xubah dengan Yubah

```
model3 = lm(formula = Yubah ~ Xubah)
summary(model3)
```

```
##
## Call:
## lm(formula = Yubah ~ Xubah)
##
## 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 ***
## Xubah       -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
```

Diperoleh model sebagai berikut

$$\hat{Y} = 8.71245 - 0.81339X + e$$

```
dwtest(model3)
```

```
##
## Durbin-Watson test
##
## data: model3
## DW = 2.6803, p-value = 0.8629
## alternative hypothesis: true autocorrelation is greater than 0
```

Karena p-value = 0.8629 (pada DW test) > alpha = 0.05, maka tak tolak H0, sisaan saling bebas

Kesimpulan dan Transformasi Balik

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

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

Jika ingin mengembalikan model untuk menjelaskan peubah respons sebelum ditransformasi, perlu dilakukan transformasi balik pada model yang dibentuk. Prosedur transformasi balik dilakukan dengan proses matematika biasa dengan mengutamakan pengubahan peubah respons terlebih dahulu.

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