

# Tugas Individu Analisis Regresi

Muhammad Haikal Rasyadan

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(ggribes)
library(ggplot2)
library(datarium)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats   1.0.0      v stringr   1.5.0
## v lubridate 1.9.3      v tibble   3.2.1
## v purrr     1.0.1      v tidyr    1.3.0
## v readr     2.1.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x 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)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```

```
library(plotly)
```

```
##
## 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(dplyr)
library(lmtest)
```

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

```
library(stats)
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:purrr':
##
##   some
##
## The following object is masked from 'package:dplyr':
##
##   recode
```

## Data

```
tugas <- read_excel("/Users/user/Downloads/Anreg Individu.xlsx")
y <- tugas$Y
x <- tugas$X
n <- nrow(tugas)
n
```

```
## [1] 15
```

```
data<-data.frame(cbind(y,x))
head(tugas)
```

```
## # A tibble: 6 x 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
```

## Model Regresi (awal)

```
model <- lm(y~x, tugas)
model
```

```
##
## Call:
## lm(formula = y ~ x, data = tugas)
##
## Coefficients:
## (Intercept)          x
##    46.4604      -0.7525
```

```
summary(model)
```

```
##
## Call:
## lm(formula = y ~ x, data = tugas)
##
## 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
```

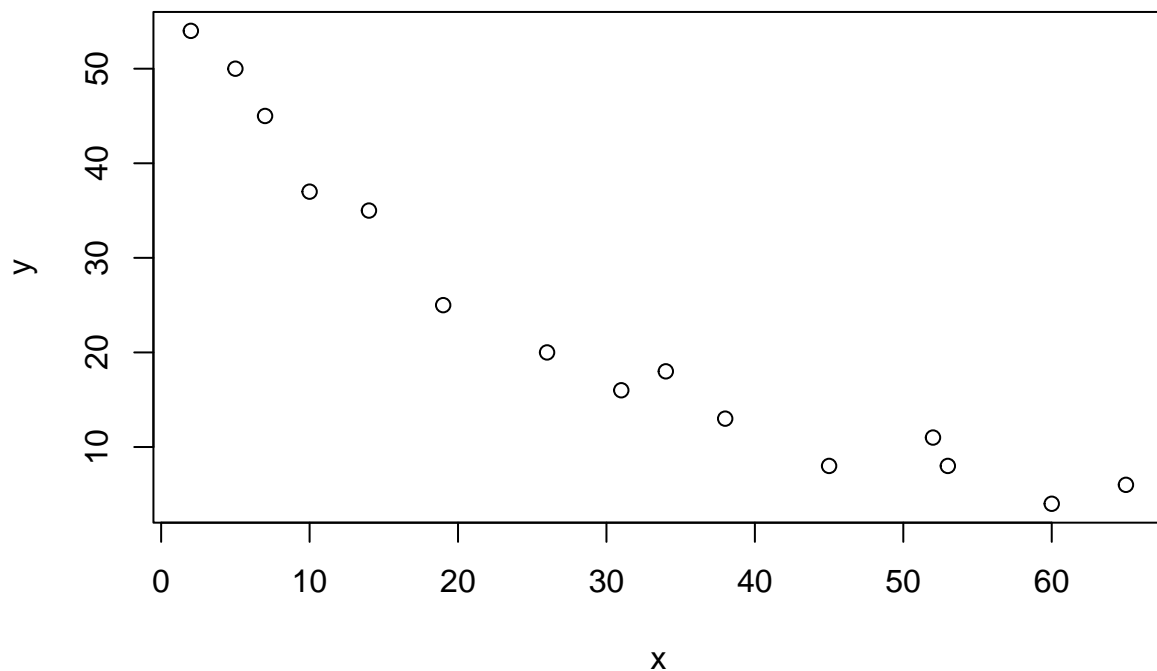
$$\hat{Y} = 46.46041 - 0.75251X + e$$

Persamaan Model di atas belum dapat dianggap sebagai model terbaik karena belum mengikuti serangkaian uji untuk memenuhi asumsi Gauss-Markov dan normalitas, yang merupakan prasyarat untuk mendapatkan model terbaik. Oleh karena itu, diperlukan serangkaian uji untuk menilai kualitas model. Jika uji tersebut menunjukkan bahwa asumsi belum terpenuhi, maka akan dilakukan transformasi pada data untuk memastikan kualitas model yang lebih baik.

## Eksplorasi Data

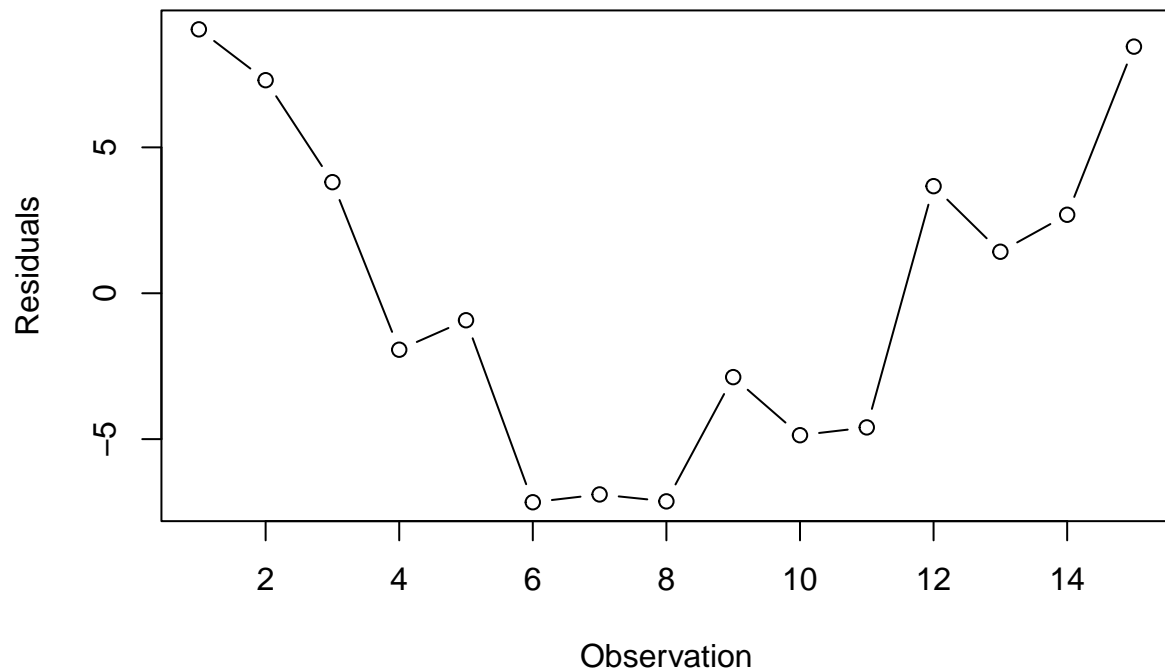
### Hubungan X dan Y

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



### Residual dan Urutan

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



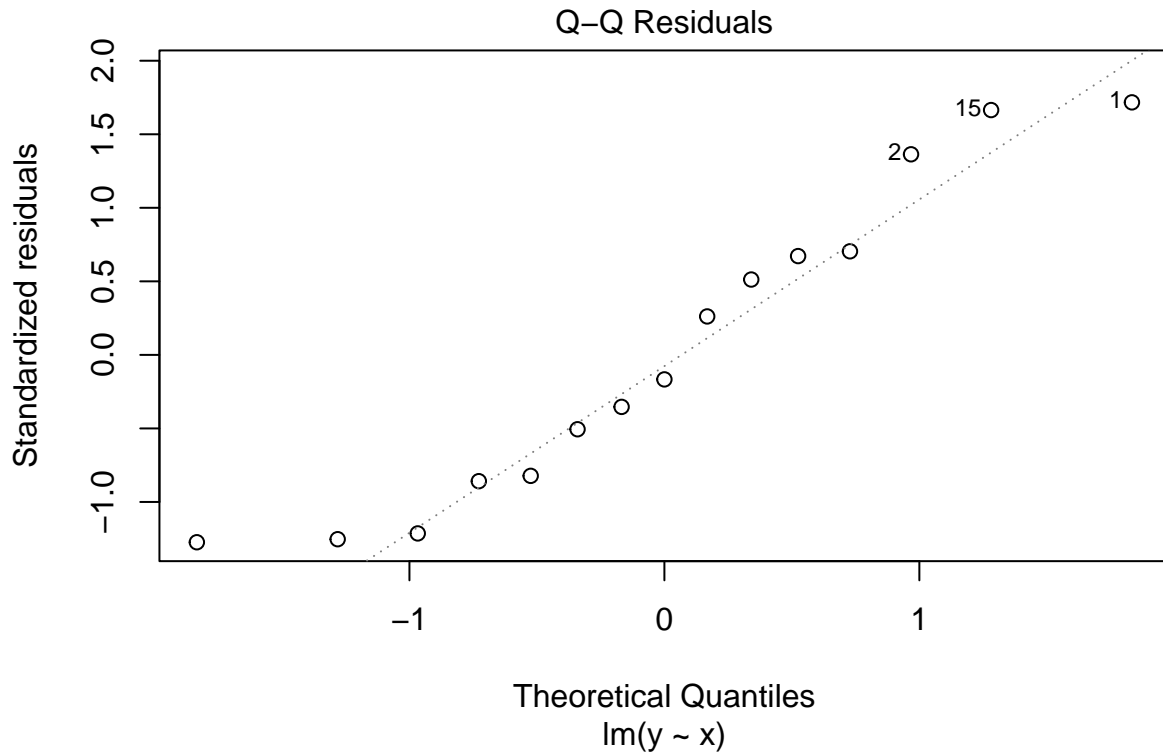
dapat terlihat berdasarkan plot tersebut bahwa plot tersebut membentuk pola kurva, sehingga residual tidak saling bebas dan model menjadi kurang baik

## Uji Normalitas

$H_0$  = residual menyebar normal

$H_1$  = residual tidak menyebar normal

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



### Uji Kolmogorov-Smirnov

```
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
```

hasil p-value dari uji Kolmogorov-Smirnov lebih besar dari 0.05, maka dapat disimpulkan bahwa data dalam model menyebar secara normal

Uji Shapiro Wilk

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

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

hasil p-value dari Uji Kolmogorov-Smirnov dan Uji Shapiro Wilk lebih besar dari 0.05, maka dapat disimpulkan bahwa data dalam model menyebar secara normal

## Pemeriksaan Asumsi

Uji Formal Kondisi Gauss-Markov

### Uji-t

$H_0$  = Nilai harapan residual sama dengan 0

$H_1$  = Nilai harapan residual tidak sama dengan 0

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

```
##
## One Sample t-test
##
## data: model$residuals
## t = -1.6158e-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
## -2.368476e-16
```

Dapat dilihat berdasarkan uji t bahwa p-value lebih dari 0.05 maka tak tolak  $H_0$ , yaitu nilai harapan residual sama dengan nol.

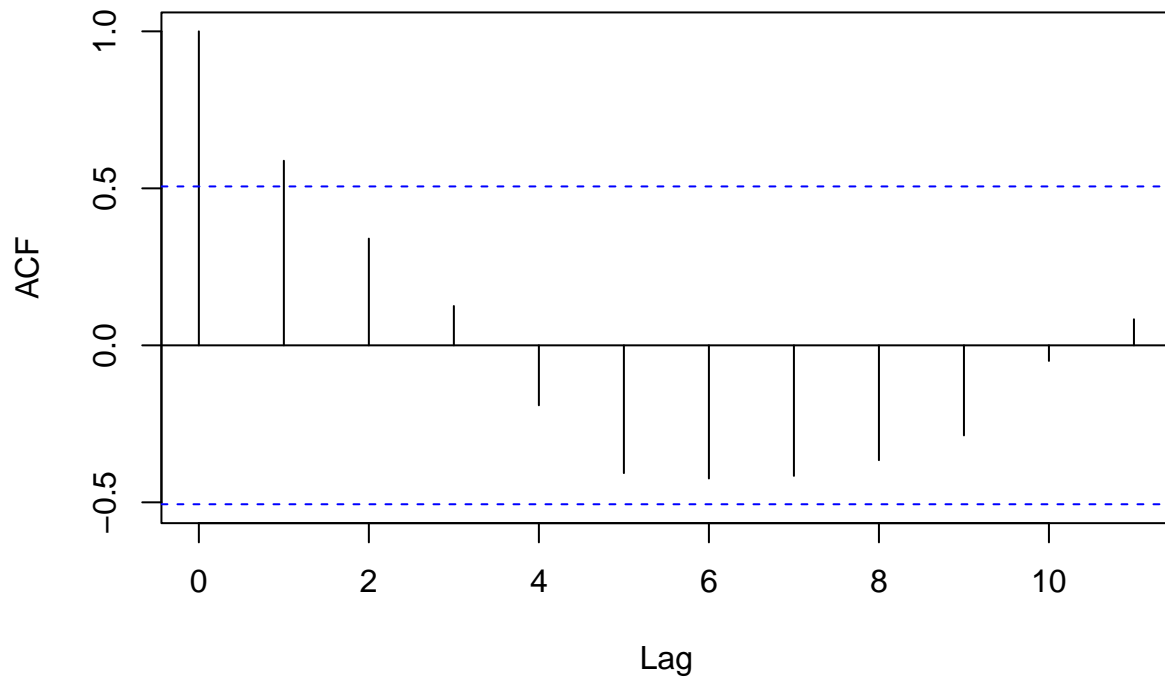
### Uji Autokorelasi

$H_0$  = residual saling bebas

$H_1$  = residual tak saling bebas

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

## Series model\$residuals



Pada plot ACF diatas, terlihat bahwa nilai autokorelasi pada lag 1 sebesar 0.5 dan pada lag 2 sekitar 0.4. Kedua nilai ini berada di luar batas kepercayaan 95%, menunjukkan bahwa autokorelasi pada kedua lag tersebut signifikan. Keberadaan autokorelasi yang signifikan menunjukkan bahwa asumsi Gauss-Markov tidak terpenuhi.

## Uji Durbin Watson

```
dwtest(model)
```

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

hasil p-value dari uji Durbin-Watson lebih rendah daripada 0.05 hal tersebut menunjukkan adanya autokorelasi positif dalam data, oleh karena itu uji Durbin Watson ini menguatkan tidak terpenuhinya asumsi Gauss-Markov, yaitu non-autokorelasi.

## Uji Homoskedastisitas

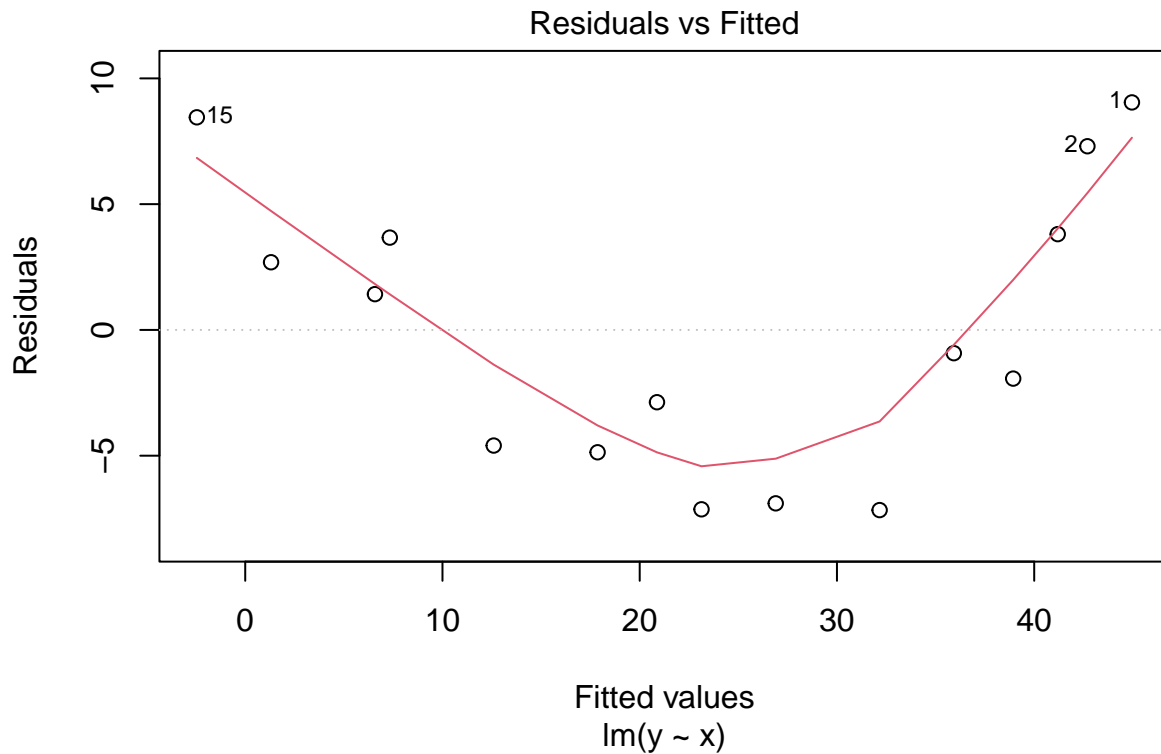
$$H_0 = \text{Ragam residual homogen}$$



$H_1$  = ragam residual tidak homogen

### 1&2 plot sisaan vs yduga

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



Berdasarkan grafik diatas, dapat diamati bahwa ragam residual tetap konstan pada berbagai tingkat nilai prediksi. Namun, terdapat kecenderungan peningkatan varian residual seiring dengan meningkatnya nilai prediksi. Fenomena ini mengindikasikan adanya kehomogenan dalam variabilitas residual.

Dikarenakan Model berbentuk kurva maka model tidak pas, oleh karena itu diperlukan suku-suku lain dalam model atau transformasi terhadap Y dan X.

### Uji Breusch-Pagan

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

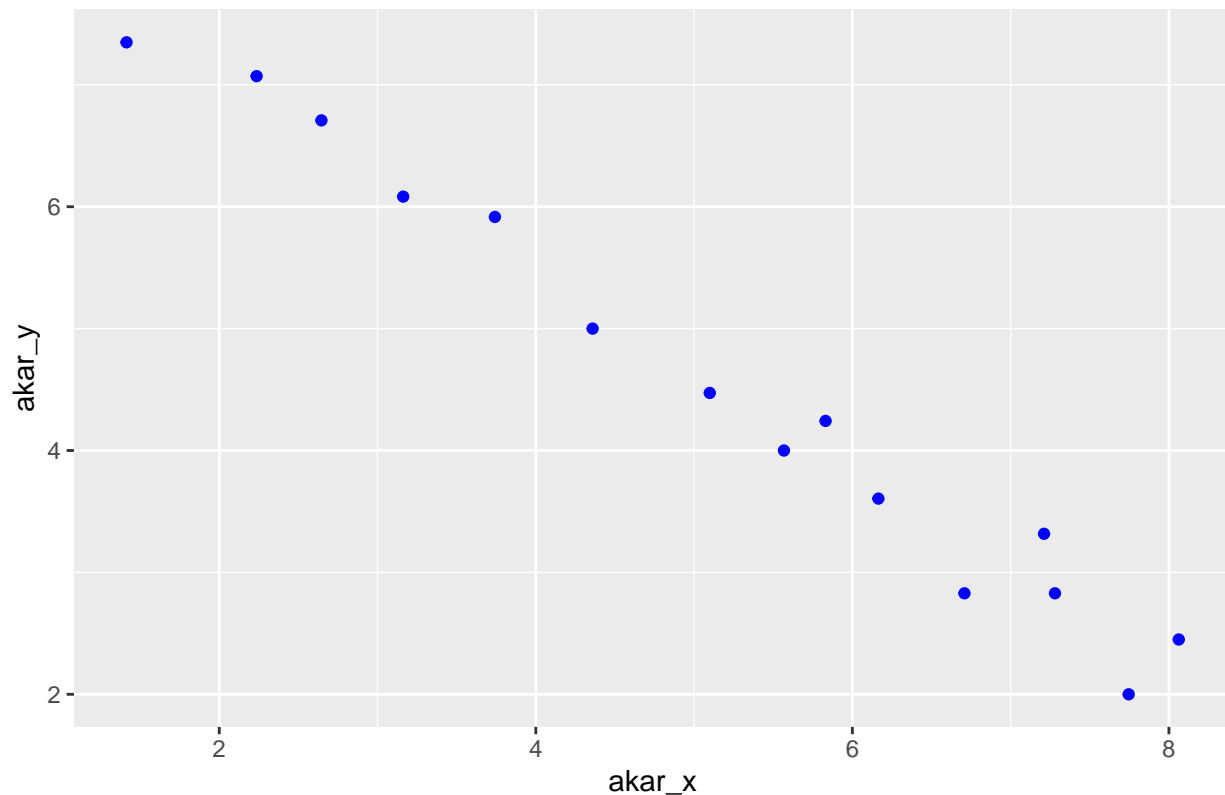
dihasilkan bahwa p-value dari uji Breusch-Pagan lebih besar daripada 0.05 sehingga tak tolak  $H_0$ . Residual dari sisaan bersifat homogen atau terdapat kehomogenan dalam variabilitas residual.

## Transformasi

transformasi y dan x diperkecil

```
akar_y<-sqrt(y)
akar_x<-sqrt(x)
dtrans1<-data.frame(akar_x,akar_y)
ggplot(tugas = dtrans1)+
  geom_point(aes(x=akar_x, y=akar_y),col="blue")+
  labs(title = "Transformasi memperkecil X dan Y")
```

## Transformasi memperkecil X dan Y



```
model_trans=lm(formula=akar_y~akar_x)
summary(model_trans)
```

```
##
## Call:
## lm(formula = akar_y ~ akar_x)
##
## 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 ***
## akar_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
```

## Uji lanjutan

### Uji Durbin Watson

```
dwtest(model_trans)
```

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

Didapatkan hasil p-value dari Uji Durbin-Watson lebih besar daripada 0.05 maka dapat disimpulkan bahwa data saling bebas atau tidak memiliki tidak autokorelasi.

## Kesimpulan

Berdasarkan perbandingan hasil uji sebelum dan sesudah transformasi, didapatkan kesimpulan bahwa model setelah transformasi lebih baik daripada model sebelum ditransformasi karena menghasilkan model yang lebih efektif dan memenuhi semua asumsi dalam regresi linear sederhana

Oleh karena itu, didapatkan model terbaik yaitu:

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

model regresi tersebut menjelaskan bahwa ketika nilai X meningkat maka Y menurun dengan cepat. nilai Y ketika X sama dengan 0 didapatkan hasil 8.71245. Diketahui juga dalam persamaan tersebut terdapat hubungan negatif antar peubah Y dan X.