

# Tugas Individu Analisis Regresi

HARIYOL\_G1401221009

2024-03-06

## Membaca Data

```
Anreg <- read.csv("C:/#2 SEMESTER 4 IPB UNIVERSITY/#MATA KULIAH/ANALISIS  
REGRESI/Tugas Individu.csv", sep = ";")
```

```
Anreg
```

```
##      X  Y  
## 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
```

```
Y<-Anreg$Y
```

```
X<-Anreg$X
```

```
n <- nrow(data)
```

```
model <- lm(Y~X, Anreg)
```

```
summary(model)
```

```
##
```

```
## Call:
```

```
## lm(formula = Y ~ X, data = Anreg)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -7.1628 -4.7313 -0.9253  3.7386  9.0446
```

```
##
```

```
## Coefficients:
```

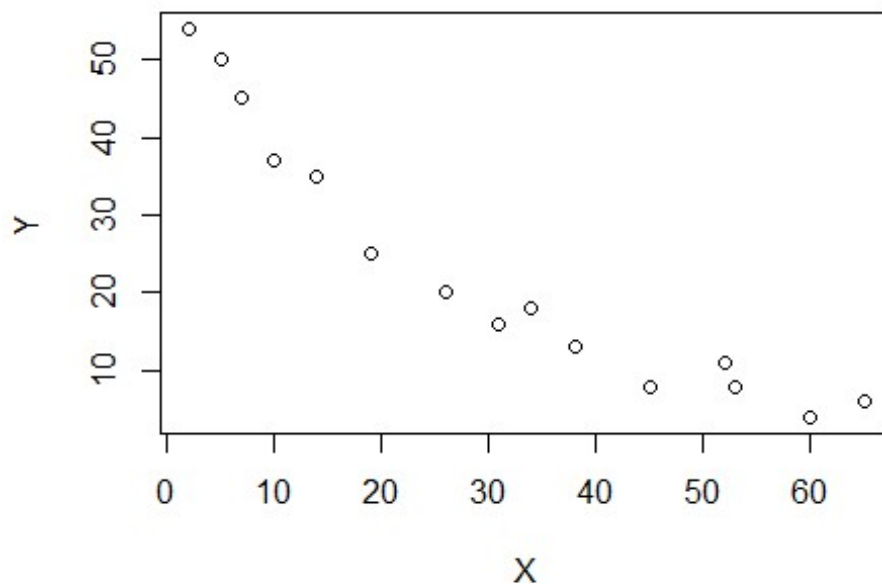
```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 46.46041    2.76218   16.82 3.33e-10 ***  
## X          -0.75251    0.07502  -10.03 1.74e-07 ***
```

```
## ---
```

```
## 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
```

## Eksplorasi Data

```
y.bar <- mean(Y)
plot(X,Y)
```



Berdasarkan Scatter plot diatas, hubungan antara X dan Y tidak linear. Hubungan antara X dan Y membentuk pola ekponensial

## Uji Formal

### Normalitas:Kolmogorov-Smirnov

Uji Ini memiliki hipotesis sebagai berikut,

$H_0$ :  $N$ (sisaan menyebar Normal)

$H_1$ :  $N$ (Sisaan tidak menyebar Normal)

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

```
##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  sisaan_model
## D = 0.12432, p-value = 0.7701

ifelse(norm_model$p.value < 0.05, "Sisaan Tidak Menyebar Normal", "Sisaan
Menyebar Normal")

## [1] "Sisaan Menyebar Normal"
```

Nilai p-value > 0,05 bermakna Tak Tolak H0 . Hal ini mengindikasikan bahwa dalam taraf nyata 5%, tidak cukup bukti untuk menyatakan bahwa sisaan tidak menyebar Normal.

### Homogenitas : Breusch-Pagan

Uji ini memiliki hipotesis sebagai berikut,

$$H_0: \text{var}[\epsilon] = \sigma^2 I(\text{Ragam Homogen})$$

$$H_1: \text{var}[\epsilon] \neq \sigma^2 I(\text{Ragam tidak Homogen})$$

```
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 4.3.2
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

(homogen_model <- bptest(model))

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

ifelse(homogen_model$p.value < 0.05, "Ragam Tidak Homogen", "Ragam Homogen")

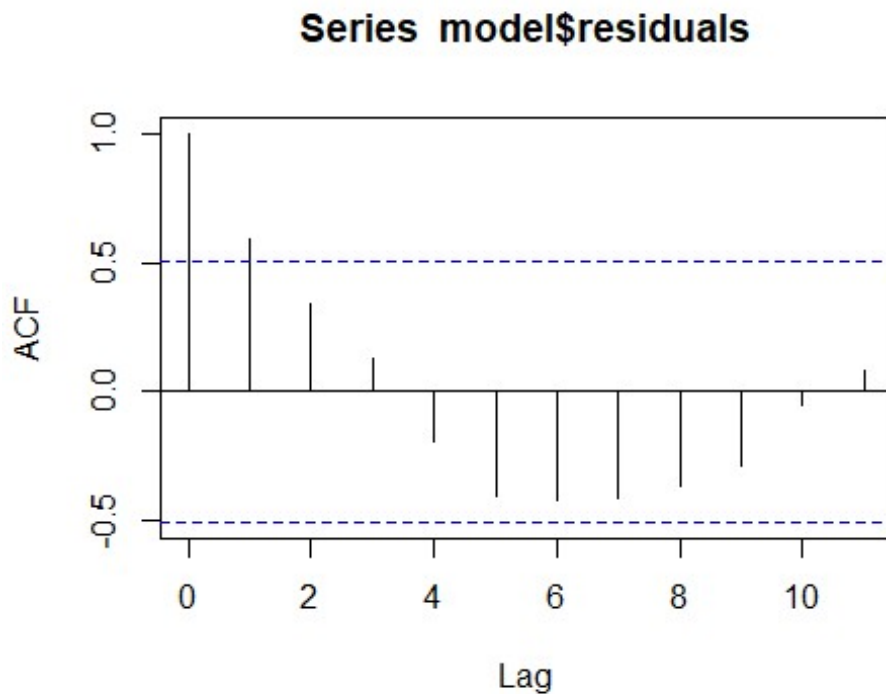
##
##      BP
## "Ragam Homogen"
```

Nilai p-value > 0,05 bermakna tak tolak H0 . Hal ini mengindikasikan bahwa dalam taraf nyata 5%, tidak cukup bukti untuk menyatakan bahwa ragam sisaan tidak homogen.

### Autokorelasi

```
dwtest(model)
```

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



#Berdasarkan grafik diatas, autokorelasi pada lag 1 adalah 0.5 dan pada lag 2 adalah 0.4. Kedua nilai tersebut melebihi batas kepercayaan 95%, menandakan bahwa autokorelasi pada lag 1 dan 2 signifikan. Ini mengindikasikan ketidakpenuhan asumsi Gauss-Markov, khususnya asumsi non-autokorelasi. Hasil p-test dari Uji Durbin-Watson juga menunjukkan nilai kurang dari 0.05, memperkuat temuan tersebut.

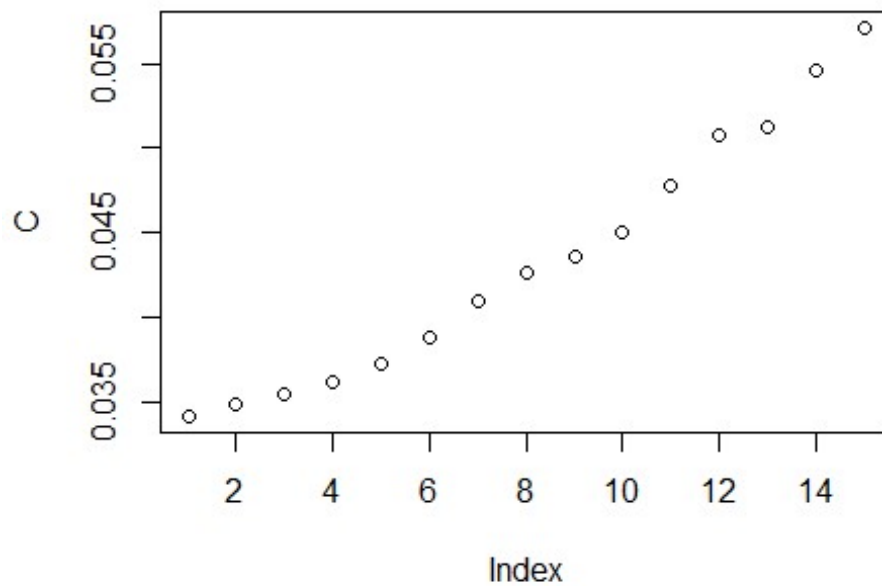
### Penanganan Kondisi Tak Standar

#### Transformasi Weighted Least Square

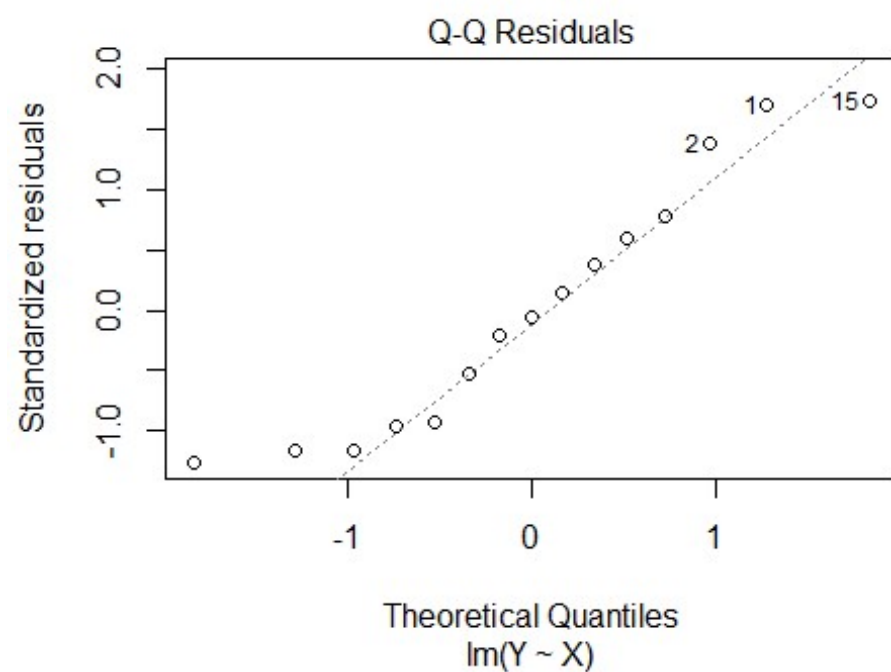
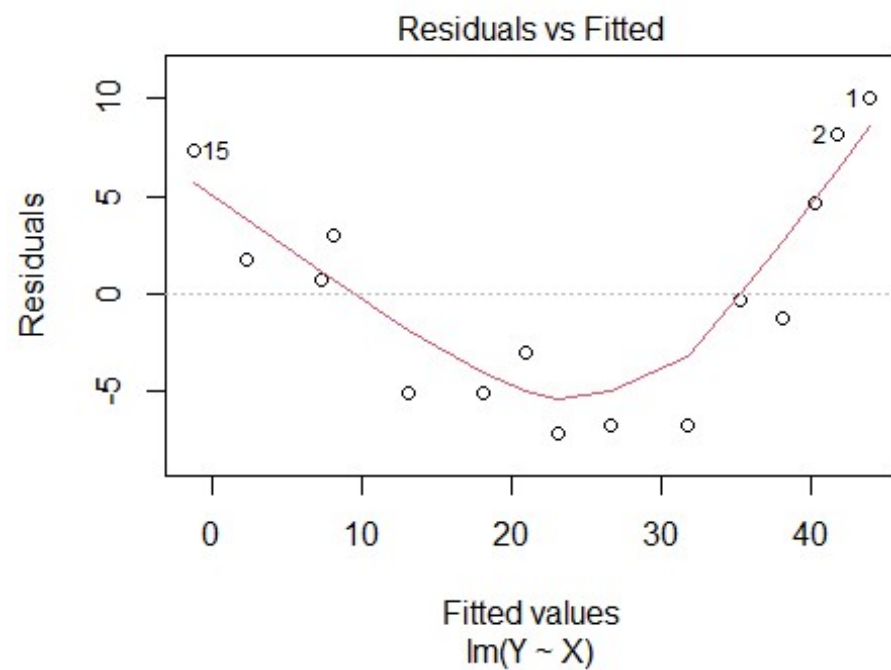
```
A <- abs(model$residuals)  
B<- model$fitted.values  
fit <- lm(A ~ B, Anreg)  
C <- 1 / fit$fitted.values^2  
C
```

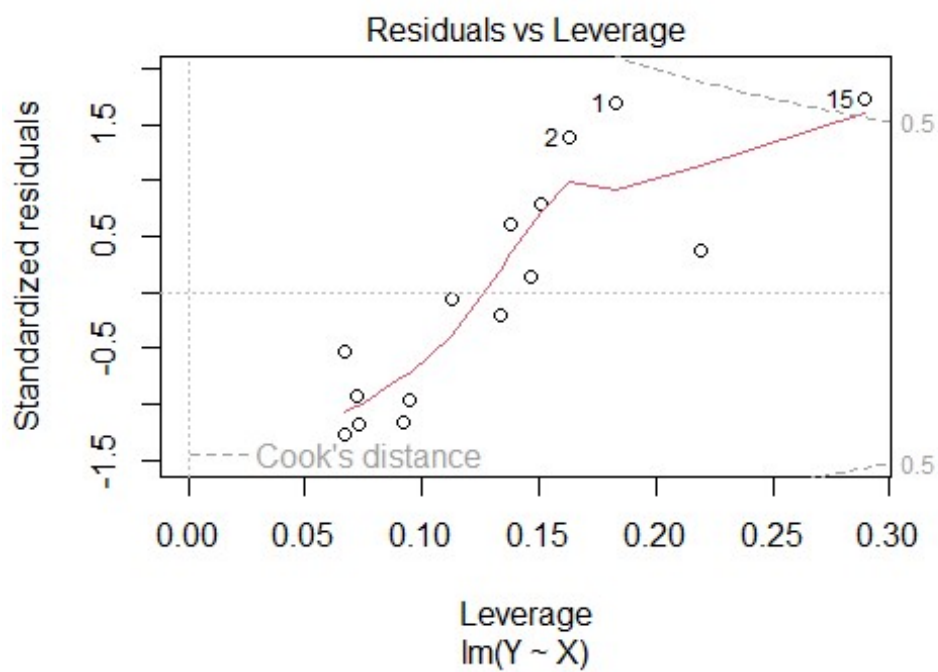
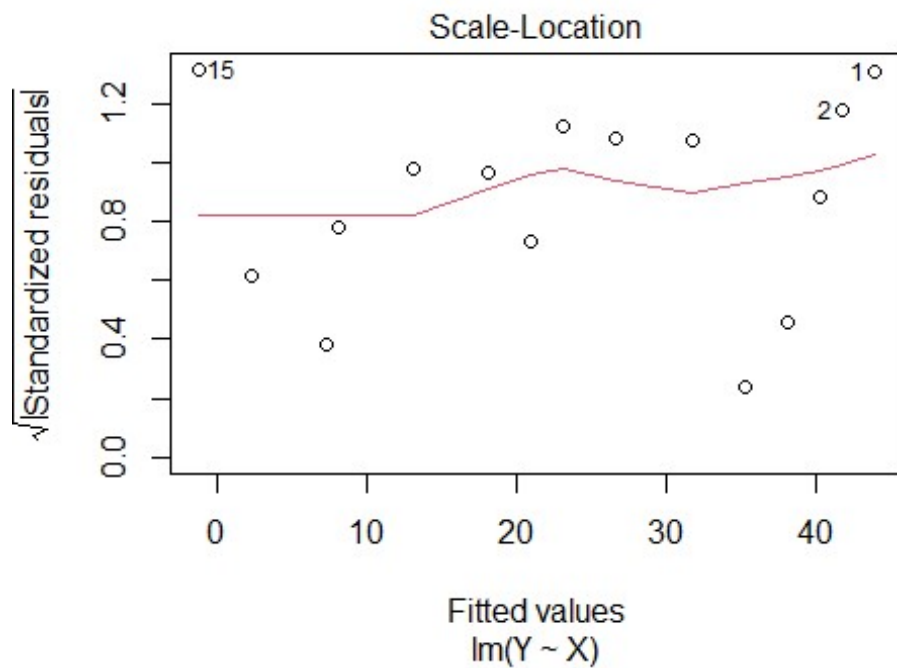
```
##           1           2           3           4           5           6
7
## 0.03414849 0.03489798 0.03541143 0.03620311 0.03730067 0.03874425
0.04091034
##           8           9          10          11          12          13
14
## 0.04257072 0.04361593 0.04507050 0.04779711 0.05077885 0.05122749
0.05454132
##           15
## 0.05710924
```

```
plot(C)
```



```
model2<- lm(Y~X, data=Anreg, weights = C)
plot(model2)
```





```
summary(model12)
```

```
##
```

```
## Call:
```

```
## lm(formula = Y ~ X, data = Anreg, weights = C)
##
## 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
```

WLS belum efektif karena belum memenuhi asumsi Gauss-Markov

### TRANSFORMASI AKAR PADA x, y, ATAU X DAN Y

```
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 'readr' was built under R version 4.3.2
## Warning: package 'stringr' was built under R version 4.3.2
## Warning: package 'forcats' 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      1.1.3      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ 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()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(ggribes)

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



```

library(GGally)

## Warning: package 'GGally' was built under R version 4.3.2

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

library(plotly)

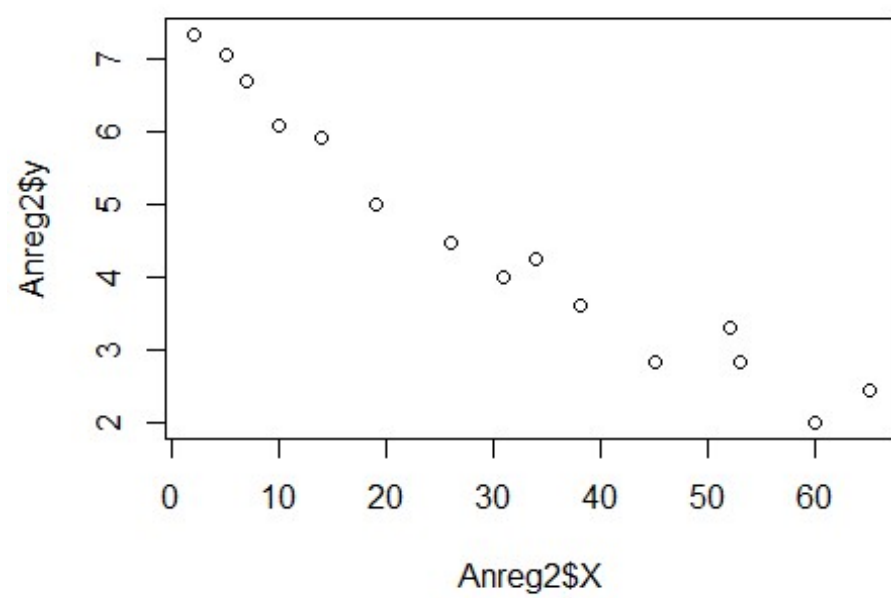
## Warning: package 'plotly' was built under R version 4.3.3

##
## 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

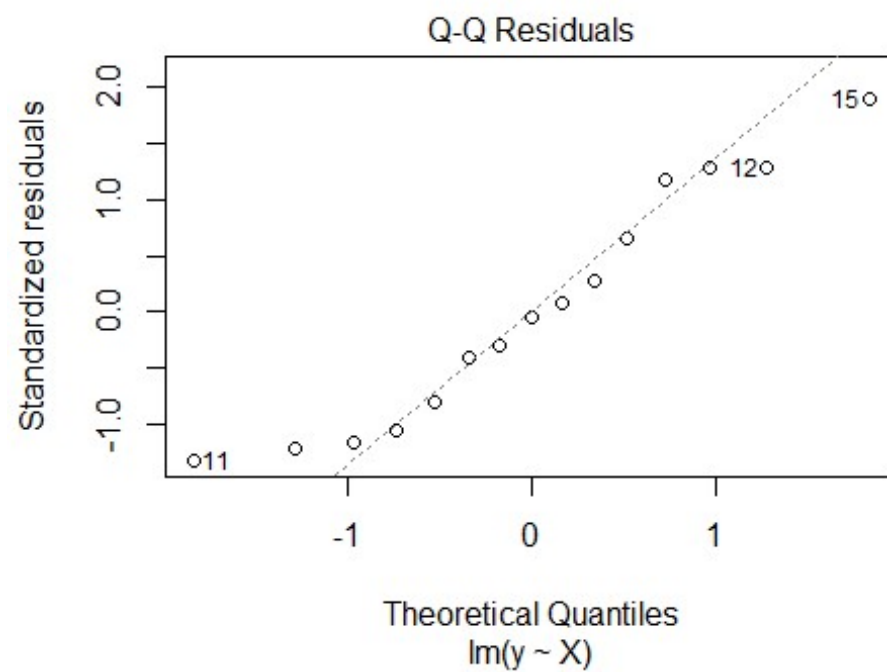
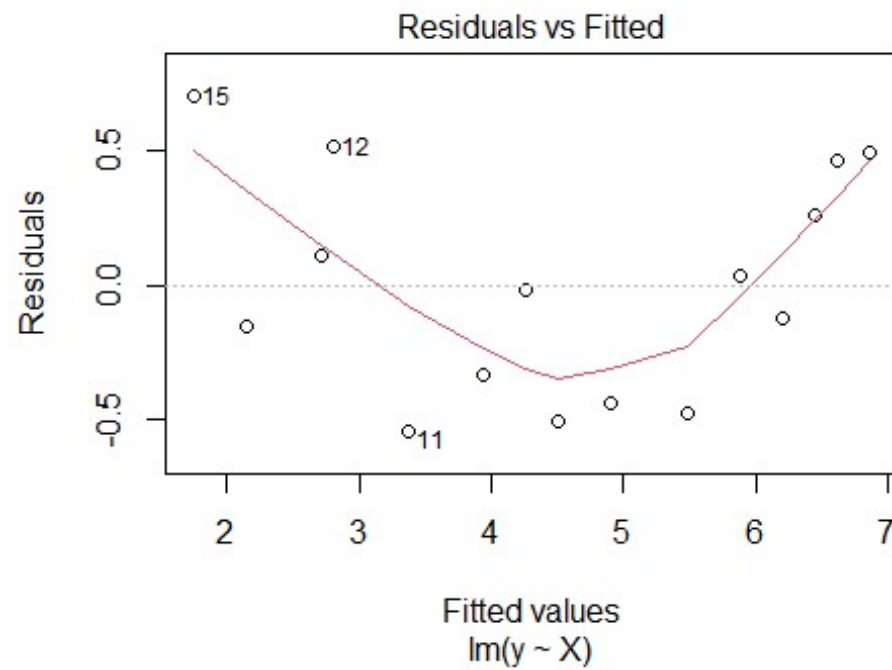
library(dplyr)
library(lmtest)
library(stats)

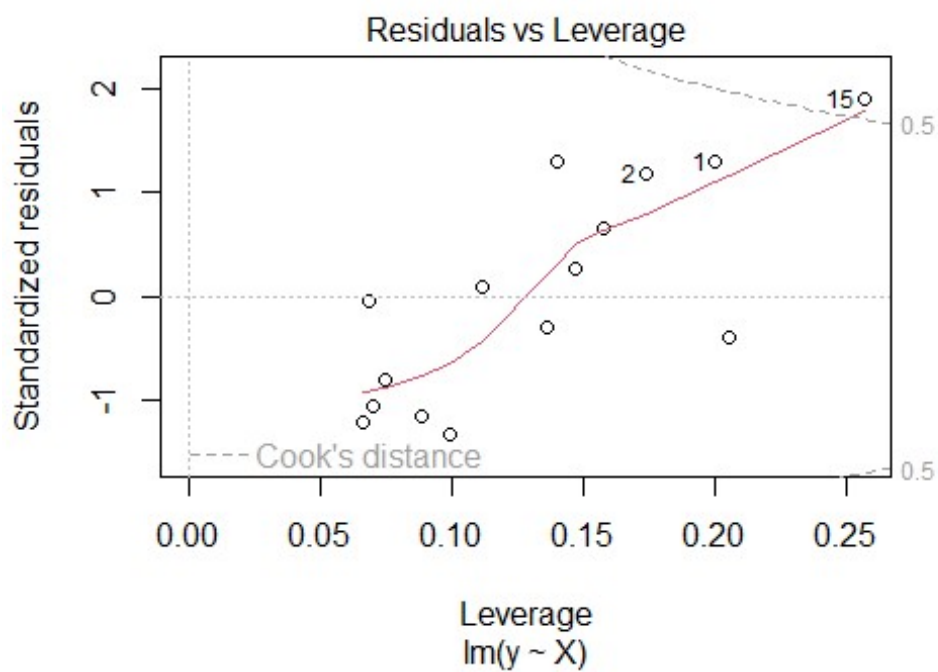
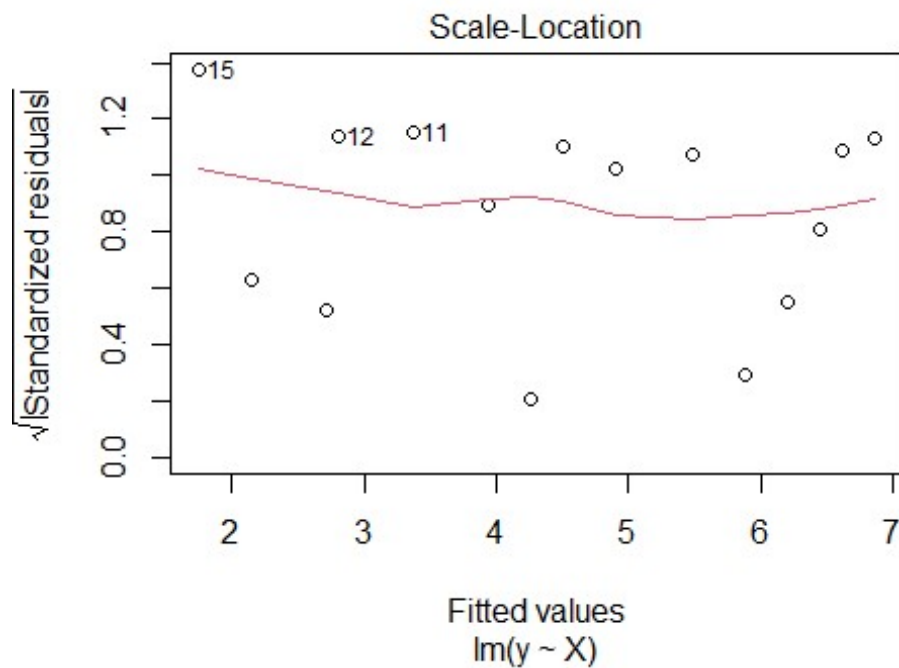
Anreg2 <- Anreg %>%
  mutate(y = sqrt(Y)) %>%
  mutate(x = sqrt(X))
model3 <- lm(y ~ X, data = Anreg2)
plot(x = Anreg2$X, y = Anreg2$y)

```



```
plot(model3)
```





```
summary(model13)
```

```
##
```

```
## Call:
```

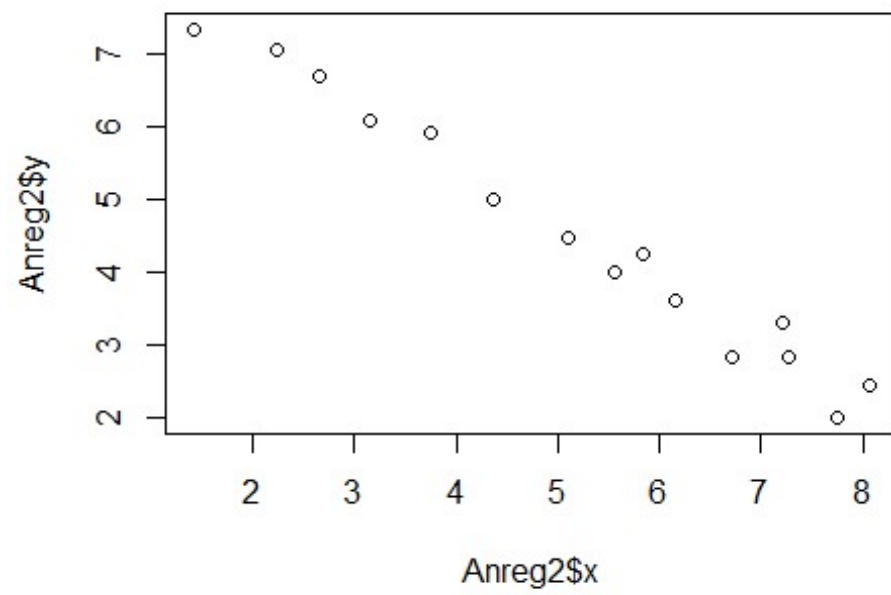
```
## lm(formula = y ~ X, data = Anreg2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53998 -0.38316 -0.01727  0.36045  0.70199
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.015455   0.201677   34.79 3.24e-14 ***
## X            -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
```

#### *UJI AUTOKORELASI MODEL REGRESI TRANSFORMASI*

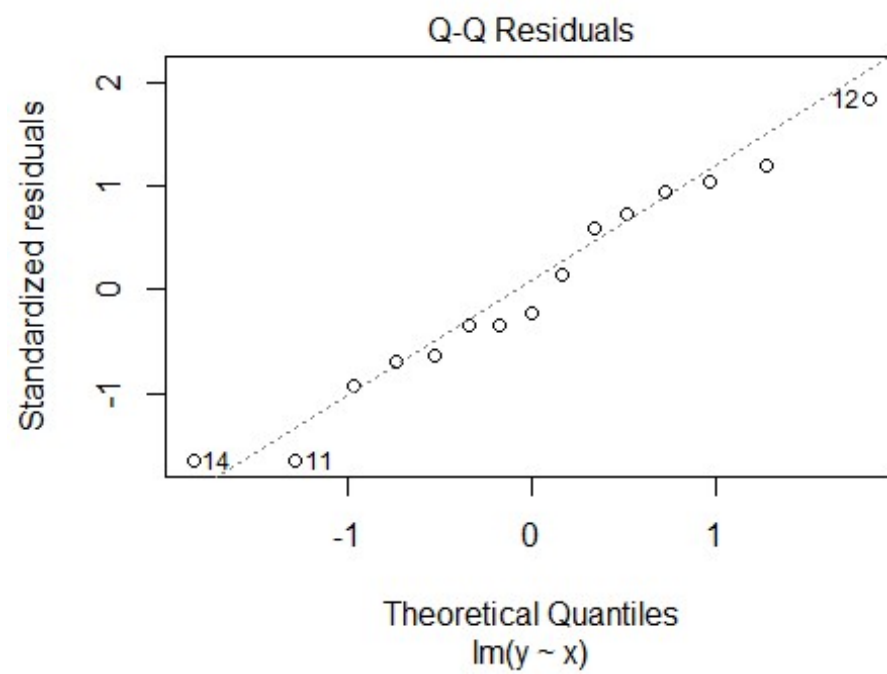
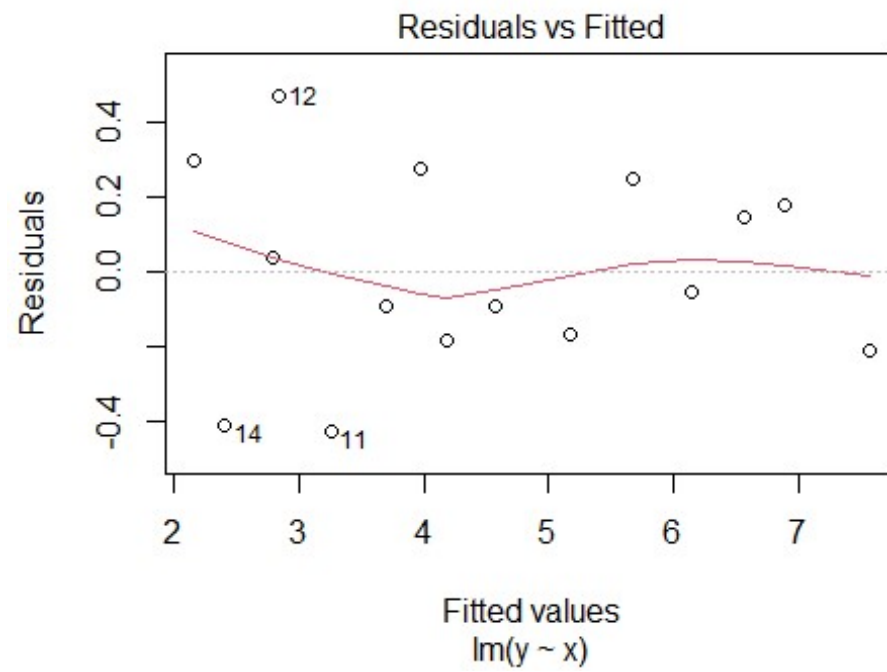
```
dwtest(model3)

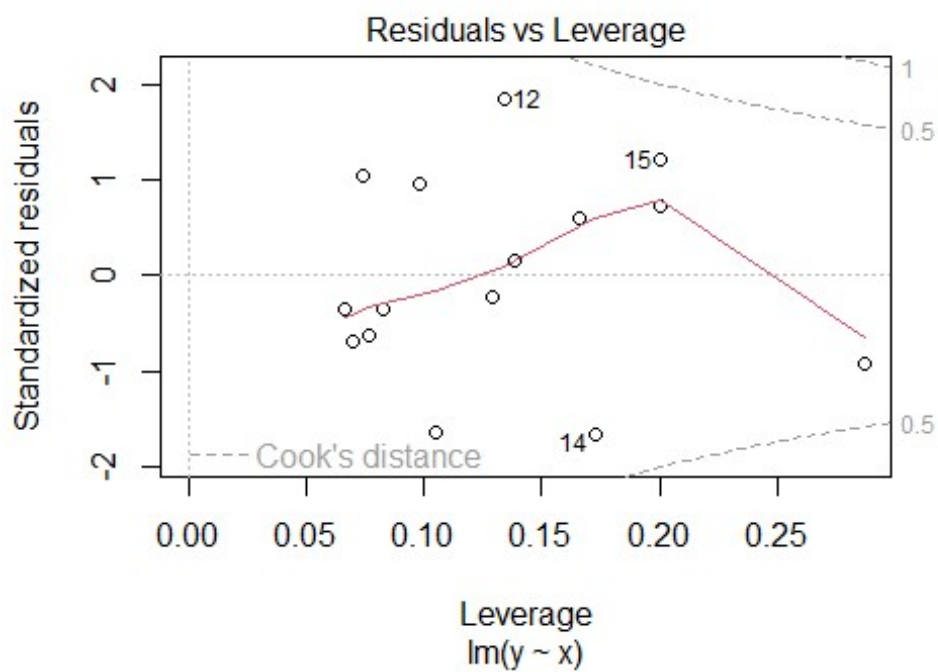
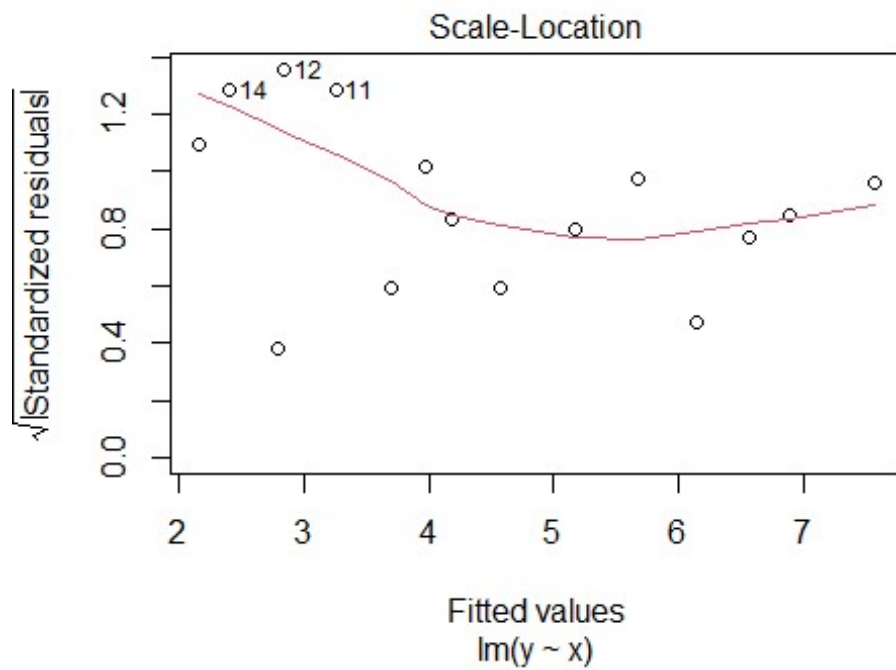
##
## Durbin-Watson test
##
## data:  model3
## DW = 1.2206, p-value = 0.02493
## alternative hypothesis: true autocorrelation is greater than 0

model3<- lm(y ~ x, data = Anreg2)
plot(x = Anreg2$x, y = Anreg2$y)
```



```
plot(model3)
```





```
summary(model3)
```

```
##
```

```
## Call:
```



```
## lm(formula = y ~ x, data = Anreg2)
##
## 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

dwtest(model3)

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

### Kesimpulan

Nilai p yang lebih besar dari 0.05 menunjukkan bahwa tidak ada bukti yang cukup untuk menolak hipotesis nol, yang menyatakan bahwa tidak ada autokorelasi. Dari hasil transformasi tersebut, dapat disimpulkan bahwa transformasi akar Y membuat persamaan regresi menjadi lebih efektif. Model regresi setelah transformasi dinyatakan sebagai:

$$Y^* = 8.71245 - 0.81339X_1 + \epsilon$$

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

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

Dengan melakukan transformasi balik, kita mendapatkan:

$$Y = \left(8.71245 - 0.81339X^{\frac{1}{2}}\right)^2 + \epsilon$$

Interpretasi model menunjukkan bahwa Y berkorelasi terbalik dengan akar kuadrat dari X, dengan hubungan yang bersifat kuadratik. Semakin besar nilai akar kuadrat dari X, semakin kecil rata-rata nilai Y, dengan tingkat penurunan yang semakin meningkat. Puncak kurva menunjukkan nilai rata-rata maksimum Y untuk nilai tertentu dari X. Konstanta 8.71245 mewakili nilai Y ketika X sama dengan 0. Koefisien -0.81339 merupakan koefisien regresi untuk variabel X. Nilai negatif menunjukkan hubungan terbalik antara Y dan akar kuadrat dari X. Dengan kata lain, semakin besar akar kuadrat dari X, semakin kecil nilai Y.

Pangkat dua pada koefisien regresi menunjukkan bahwa hubungan antara Y dan X bersifat kuadratik. Ini berarti perubahan Y tidak proporsional dengan perubahan X, melainkan berubah dengan tingkat peningkatan yang semakin tinggi.