

# Tugas Individu Analisis Regresi

Shalma Kaisya Candradewi, G1401221105

2024-03-06

## MENGINPUT, MEMBACA, DAN MENGOLAH DATA

```
data.anreg <- read.csv("C:/Users/ACER/Downloads/Tugas Individu Anreg.csv",  
sep=";")  
data.anreg
```

```
##      data.X data.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
```

Menginput data dari data yang sudah kita olah sebelumnya di excel.

```
Y<-data.anreg$data.Y  
X<-data.anreg$data.X  
n <- nrow(data)
```

```
model <- lm(Y~X, data.anreg)  
summary(model)
```

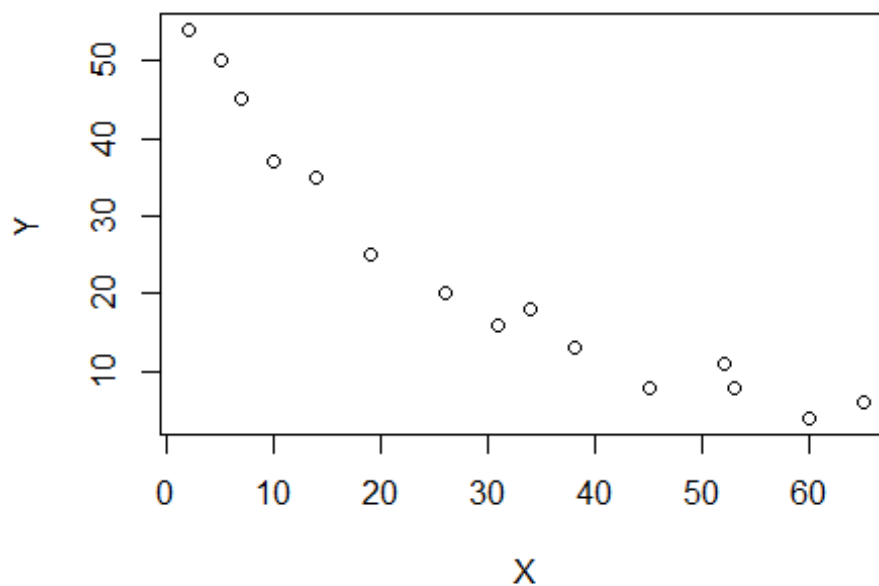
```
##  
## Call:  
## lm(formula = Y ~ X, data = data.anreg)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -7.1628 -4.7313 -0.9253  3.7386  9.0446   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (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
```

Membuat model regresi linear sederhana dan kemudian menampilkan ringkasan statistik untuk mengevaluasi kualitas model tersebut.

## EKSPLORASI DATA

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



Dilihat dari scatter plot diatas, bahwasannya terdapat hubungan anantara variabel X dan Y, dimana X dan Y tidak linear dan berbentuk pola eksponensial.

## UJI FORMAL, Normalitas:Kolmogorov-Smirnov

### *HIPOTESIS:*

H0:N (sisaan berdistribusi Normal)

H1:N (Sisaan tidak berdistribusi Normal)

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

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

Melihat residu dari model regresi linear sederhana tersebut memiliki distribusi yang mendekati distribusi normal atau tidak.

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

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

TAK TOLAK H0 karena nilai p-value > alpha(0,05) Kesimpulan: Pada taraf nyata 5% tidak terdapat cukup bukti untuk menyatakan bahwa sisaan tidak berdistribusi normal.

## HOMOGENITAS: BREUSCH-PAGAN

### *HIPOTESIS:*

H0:var[ε]=σ<sup>2</sup>I (Ragam Homogen)

H1:var[ε]≠σ<sup>2</sup>I (Ragam tidak Homogen)

```
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

(modelhomogen<- bptest(model))
```

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

Memeriksa apakah homoskedastisitas dari residu model regresi linear sederhana tersebut dapat dianggap atau tidak.

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

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

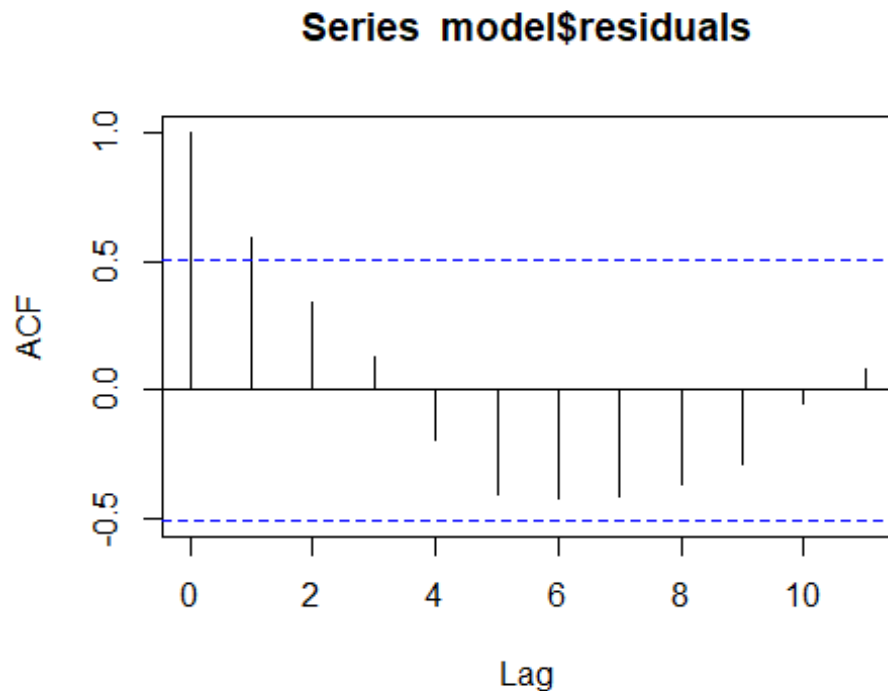
TAK TOLAK H0 karena nilai p-value > alpha(0,05) Kesimpulan: Pada taraf nyata 5% tidak terdapat cukup bukti untuk menyatakan bahwa sisaan tidak berdistribusi normal.

## 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)
```



Dilihat dari grafik diatas terdapat autokorelasi sebesar 0.5 pada lag 1 dan 0.4 pada lag 2. Kedua nilai tersebut melampaui batas kepercayaan 95%, menunjukkan bahwa autokorelasi pada lag 1 dan 2 secara signifikan berbeda dari nol. Hal ini mengisyaratkan adanya pelanggaran terhadap asumsi Gauss-Markov, terutama dalam hal non-autokorelasi. Temuan ini diperkuat oleh hasil uji p dari Uji Durbin-Watson yang juga menunjukkan nilai kurang dari 0.05.

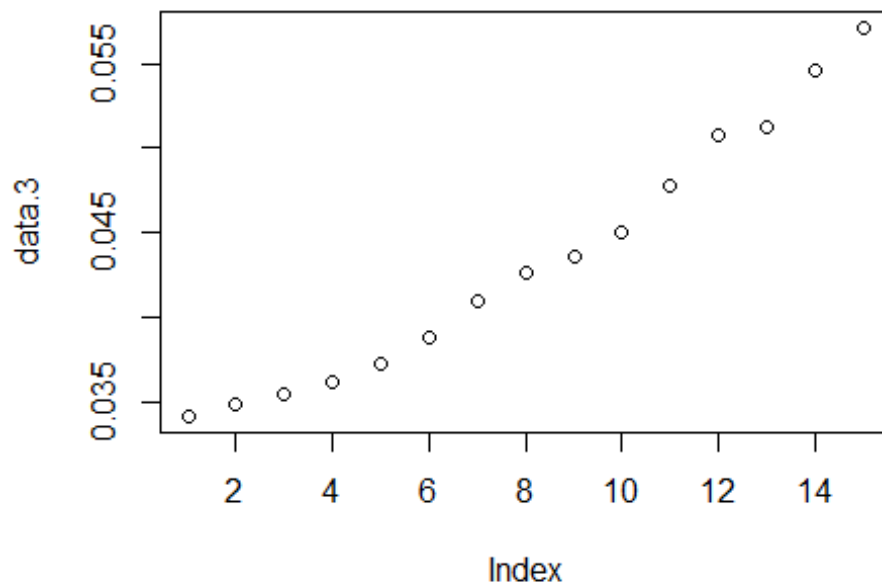
## PENANGANAN KONDISI TAK STANDAR, TRANSFORMASI WEIGHTED LEAST SQUARE

```
data.1<- abs(model$residuals)
data.2<- model$fitted.values
fit <- lm(data.1 ~ data.2, data.anreg)
data.3 <- 1 / fit$fitted.values^2
data.3
```

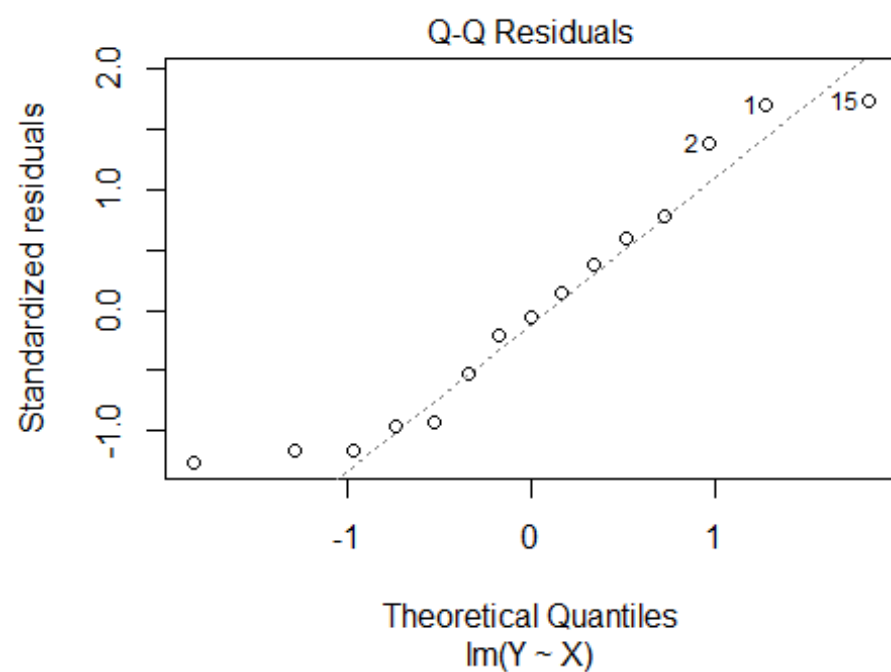
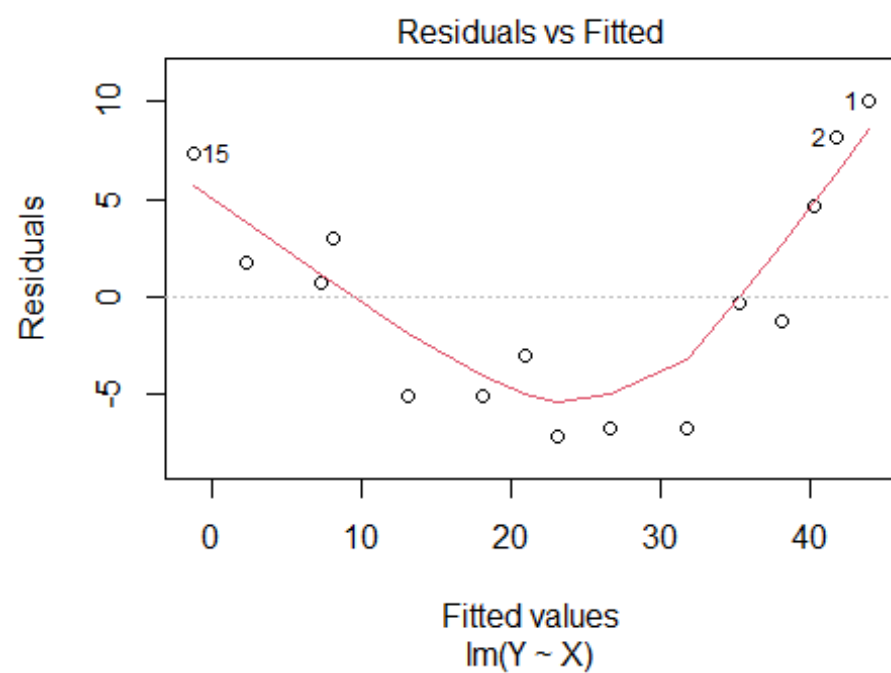
	1	2	3	4	5	6
##	0.03414849	0.03489798	0.03541143	0.03620311	0.03730067	0.03874425
7	0.04091034					
	8	9	10	11	12	13
##	0.04257072	0.04361593	0.04507050	0.04779711	0.05077885	0.05122749
14	0.05454132					
	15					
##	0.05710924					

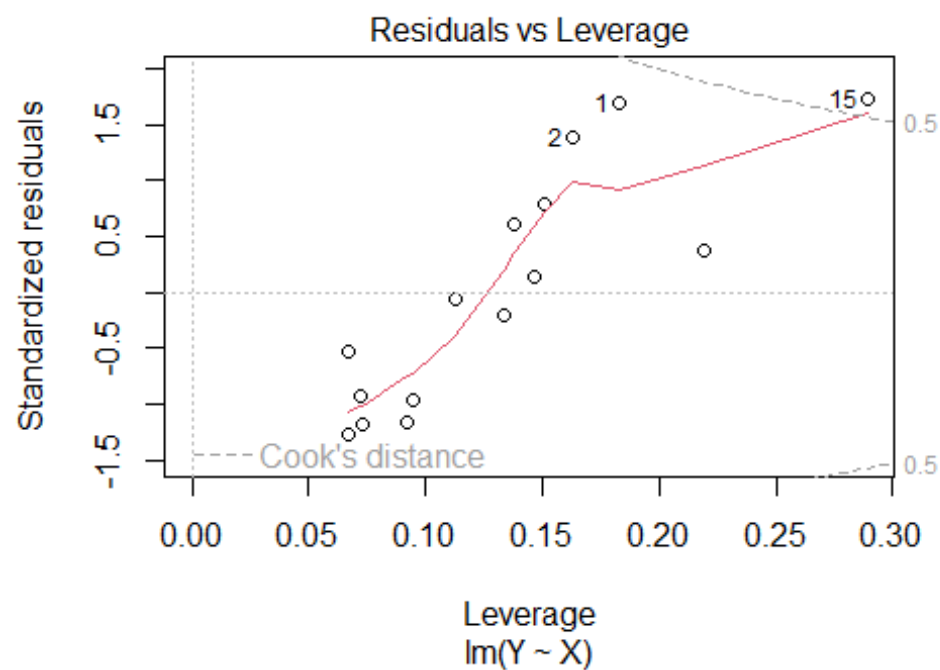
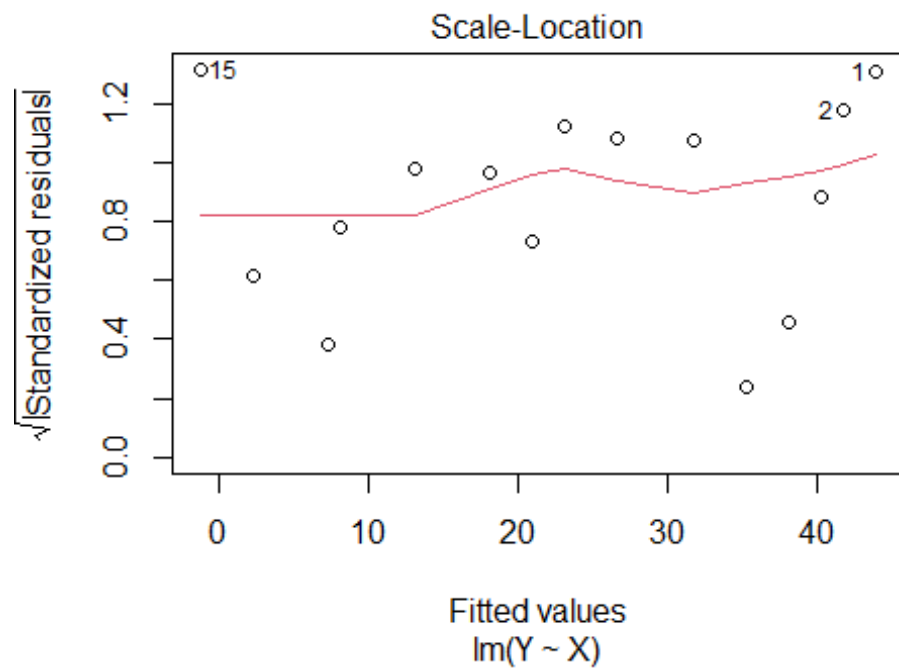
Meenghasilkan beberapa vektor yang dapat digunakan untuk analisis lebih lanjut terhadap residu dan nilai prediksi dari model regresi linear.

```
plot(data.3)
```



```
data.model<- lm(Y~X, data=data.anreg, weights = data.3)  
plot(data.model)
```





```
summary(data.model)
```

```
##
```

```
## Call:
```



```
## lm(formula = Y ~ X, data = data.anreg, weights = data.3)
##
## 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
```

Memberikan ringkasan statistik atau informasi penting. Dari data yang diperoleh bahwasannya WEIGHTED LEAST SQUARE belum cukup efisien 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 'tibble' was built under R version 4.3.2
## Warning: package 'tidyr' was built under R version 4.3.2
## Warning: package 'readr' was built under R version 4.3.2
## Warning: package 'purrr' was built under R version 4.3.2
## Warning: package 'dplyr' 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      1.1.4      ✓ 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()
## i 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 4.3.3

library(GGally)

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

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

library(plotly)

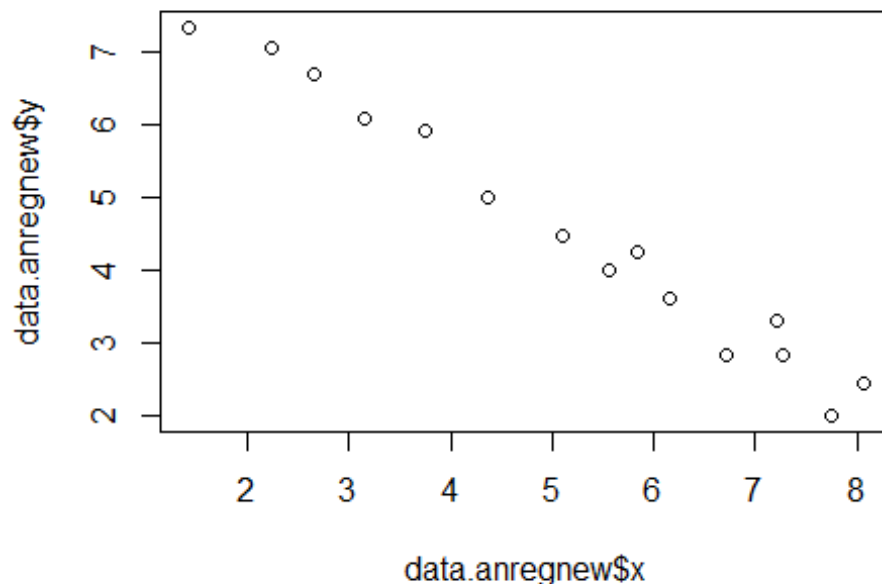
## Warning: package 'plotly' was built under R version 4.3.3

##
## 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)

data.anregnew <- data.anreg %>%
  mutate(y = sqrt(Y)) %>%
  mutate(x = sqrt(X))
model.new <- lm(y ~ X, data = data.anregnew)
plot(x = data.anregnew$x, y = data.anregnew$y)

```



Melihat visualisasi hubungan antara variabel dependen dan independen yang telah dimodifikasi serta melakukan analisis regresi linear sederhana dengan menggunakan transformasi akar kuadrat.

```
summary(model.new)
```

```
##
## Call:
## lm(formula = y ~ X, data = data.anregnew)
##
## 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
```

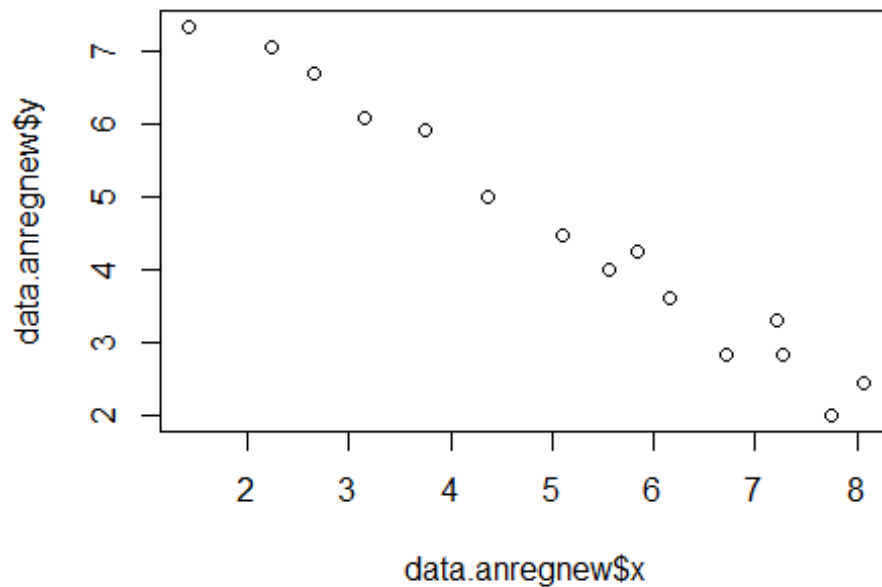
Dari data diatas F-statistic: 218.9 dengan p-value sangat rendah (1.634e-09), menunjukkan bahwa model secara keseluruhan signifikan.

## UJI AUTOKORELASI MODEL REGRESI TRANSFORMASI

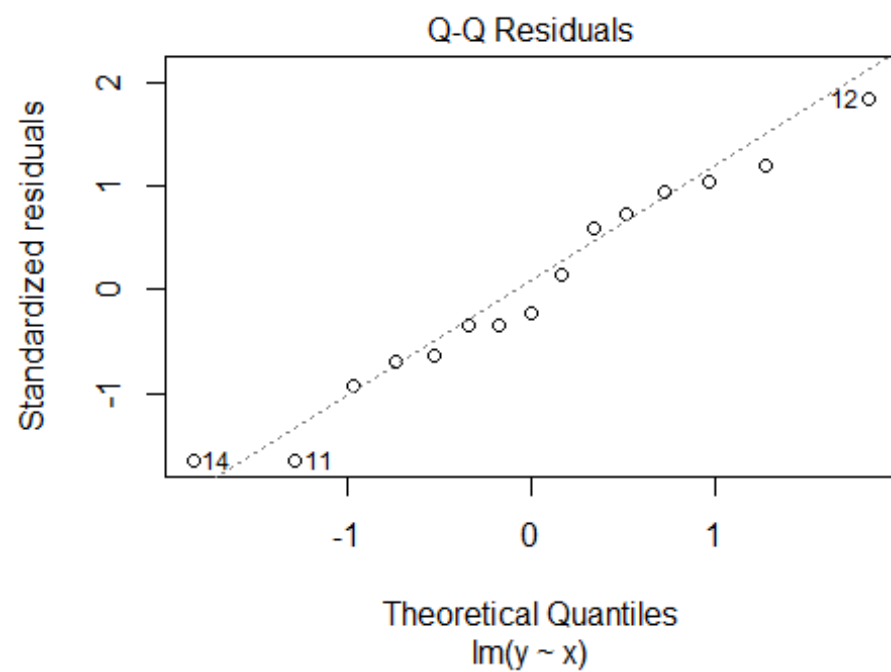
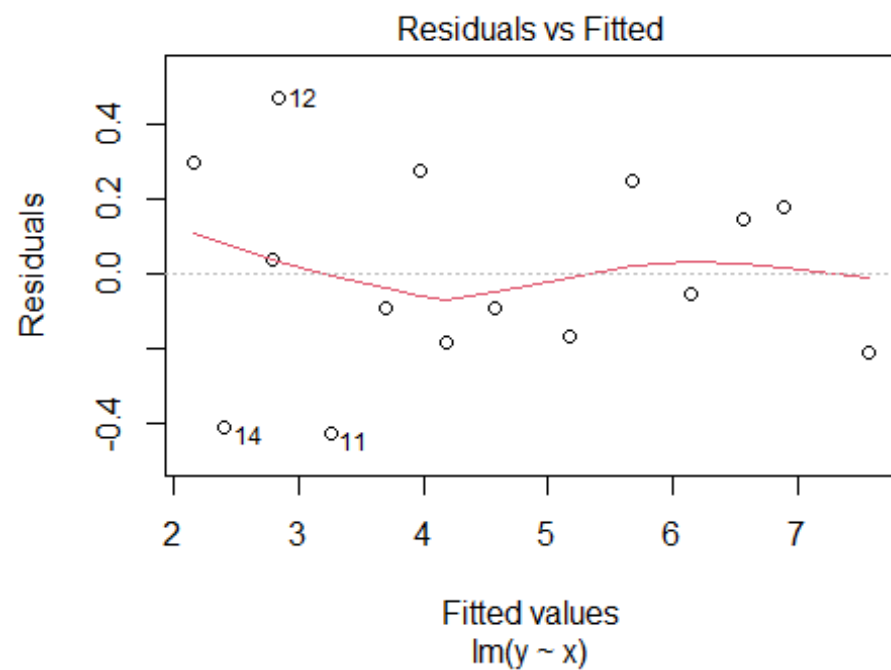
```
dwtest(model.new)
```

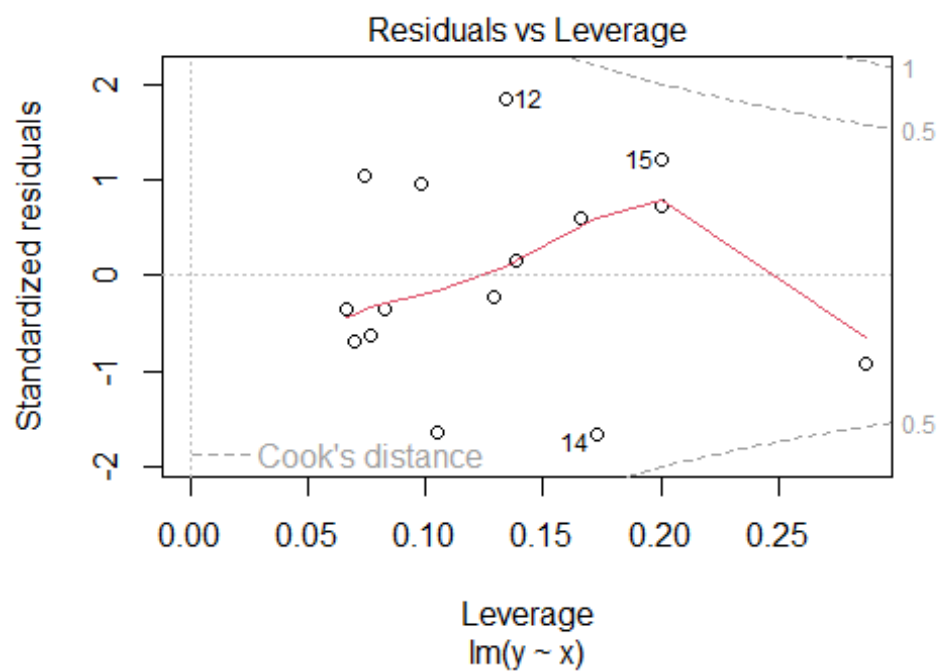
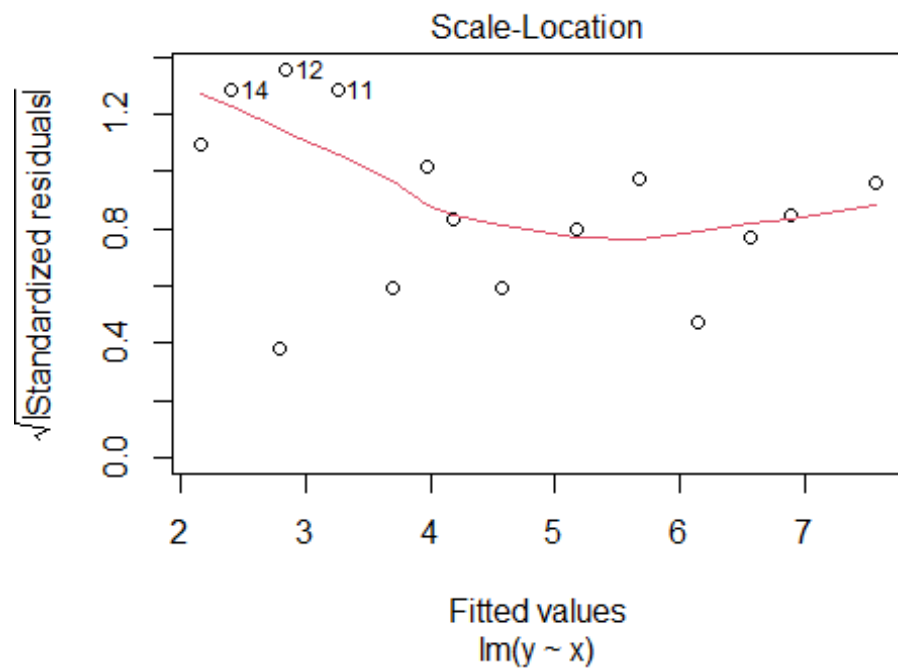
```
##  
## Durbin-Watson test  
##  
## data: model.new  
## DW = 1.2206, p-value = 0.02493  
## alternative hypothesis: true autocorrelation is greater than 0
```

```
model.baru3<- lm(y ~ x, data = data.anregnew)  
plot(x = data.anregnew$x, y = data.anregnew$y)
```



```
plot(model.baru3)
```





```
summary(model.baru3)
```

```
##
```

```
## Call:
```

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

F-statistic (557.3) dengan p-value sangat rendah (4.643e-12), menunjukkan bahwa model secara keseluruhan sangat signifikan dalam menjelaskan variasi dalam data.G14

```
dwtest(model.baru3)
##
## Durbin-Watson test
##
## data: model.baru3
## DW = 2.6803, p-value = 0.8629
## alternative hypothesis: true autocorrelation is greater than 0
```

Simpulan: TAK TOLAK H0 karena nilai p-value > alpha (0.05), dimana tidak memiliki cukup bukti untuk menolak hipotesis nol, yang menyatakan bahwa tidak ada autokorelasi dalam data. Setelah melakukan transformasi akar pada variabel dependen dan independen, ditemukan bahwa persamaan regresi menjadi lebih efektif, diwakili oleh persamaan:

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

$$Y = \left(8.7124535 - 0.8133888\sqrt{X}\right)^2 + \epsilon$$

Dengan Y\* sebagai akar kuadrat dari Y dan X1 sebagai akar kuadrat dari X. Dalam proses transformasi balik, kita mendapatkan persamaan:

[Y =

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

$$Y = \left(8.7124535 - 0.8133888\sqrt{X}\right)^2 + \epsilon$$

Penafsiran model mengungkap bahwa terdapat korelasi negatif antara Y dan akar kuadrat dari X, menunjukkan adanya hubungan kuadratik. Seiring dengan meningkatnya nilai akar

kuadrat dari X, rata-rata Y cenderung mengalami penurunan yang semakin signifikan. Semakin besar nilai akar kuadrat dari X, semakin kecil rata-rata nilai Y, dengan tingkat penurunan yang semakin meningkat.

Titik puncak pada kurva mencerminkan nilai maksimum rata-rata Y untuk suatu nilai tertentu dari X. Konstanta koefisien sebesar 8.71245 mewakili nilai Y ketika X sama dengan 0, sedangkan koefisien regresi X sebesar -0.81339 menunjukkan hubungan terbalik antara Y dan akar kuadrat dari X, menandakan sifat kuadratik dari keterkaitan keduanya. Pemangkatan dua pada koefisien regresi menggambarkan bahwa perubahan Y tidak sebanding dengan perubahan X, melainkan berubah dengan tingkat peningkatan yang semakin tinggi.