

Analisis Regresi

FAIZ AJI MUZAKKI

2024-03-06

Baca Data

```
anreg <- read.csv("C:/Users/faiza/Downloads/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)

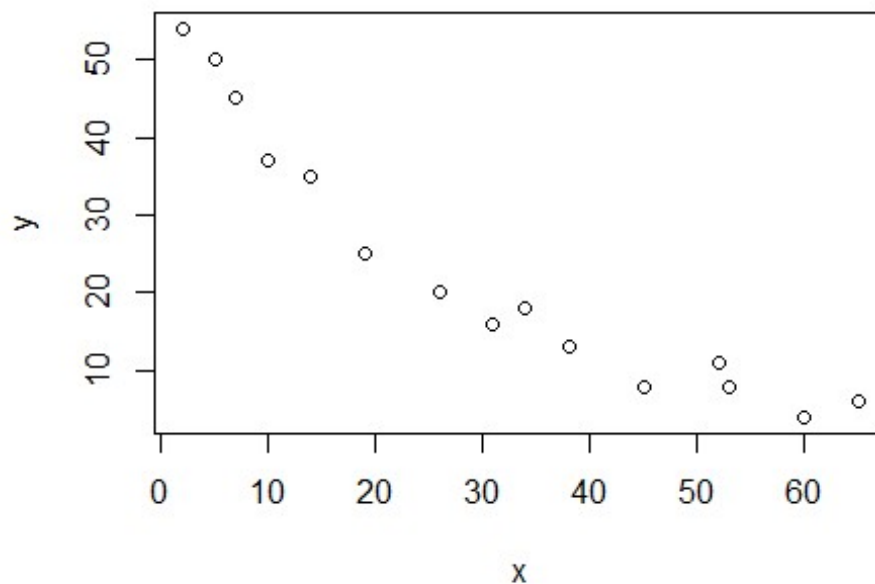
model <- lm(y~x, anreg)
summary(model)

##
## 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

```
ybar <- mean(y)
plot(x,y)
```



Berdasarkan grafik titik yang diberikan, dapat disimpulkan bahwa korelasi antara variabel X dan Y tidak bersifat linear. Sebaliknya, pola hubungan antara X dan Y terlihat mengikuti bentuk eksponensial.

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)
model_residu <- resid(model)
(norm_model <- lillie.test(model_residu))

##
##  Lilliefors (Kolmogorov-Smirnov) normality test
##
## data:  model_residu
## 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"

```

P-value yang lebih besar dari 0,05 menunjukkan bahwa kita tidak dapat menolak hipotesis nol (H_0). Artinya, tidak terdapat cukup bukti dalam taraf signifikansi 5% untuk menyimpulkan bahwa sebaran sisaan tidak normal.

Homogenitas : Breusch-Pagan Uji ini memiliki hipotesis sebagai berikut, $H_0: \text{var}[\epsilon] = \sigma^2 I$ (Ragam Homogen) $H_1: \text{var}[\epsilon] \neq \sigma^2 I$ (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 4.3.3
##
## Attaching package: 'zoo'
## The following objects are masked from 'package:base':
##
##      as.Date, as.Date.numeric

(model_homogen <- bptest(model))

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

ifelse(model_homogen$p.value < 0.05, "Ragam tidak homogen", "Ragam homogen")

##              BP
## "Ragam homogen"

```

P-value yang lebih besar dari 0,05 menandakan bahwa kita tidak dapat menolak hipotesis nol (H_0). Oleh karena itu, dalam taraf signifikansi 5%, tidak terdapat bukti yang cukup untuk menyimpulkan bahwa variasi sisaan tidak homogen.

Autokorelasi

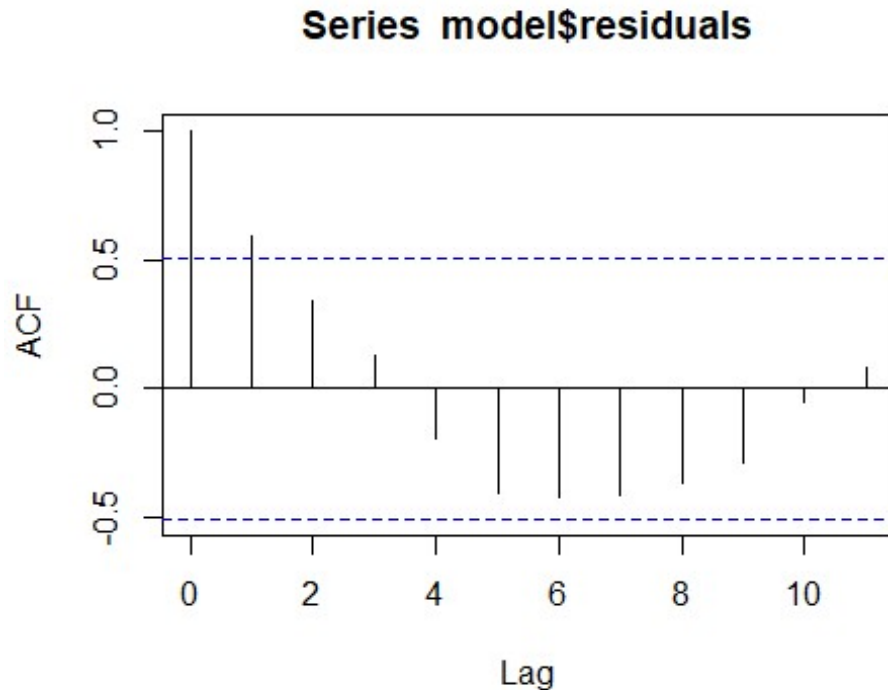
```

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)
```



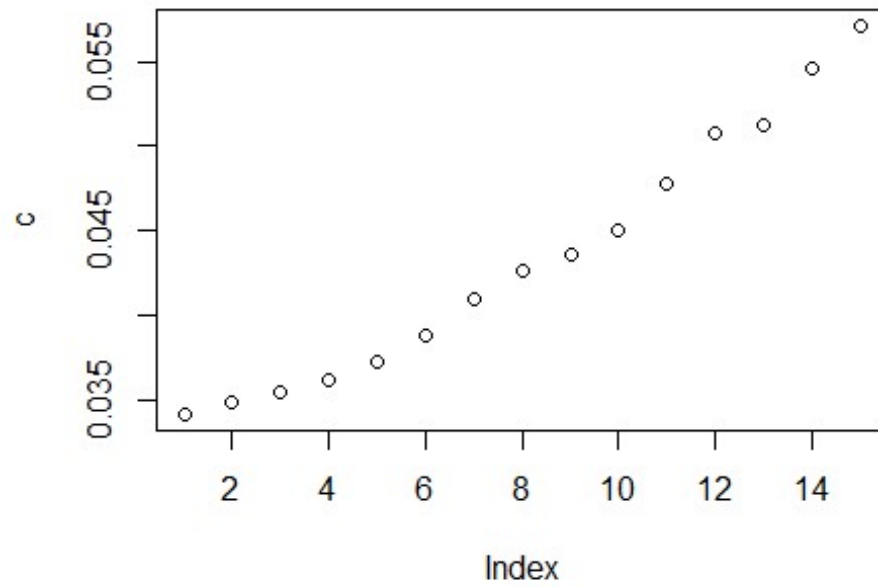
Berdasarkan grafik yang disajikan, terlihat bahwa autokorelasi pada lag 1 mencapai 0,5, sedangkan pada lag 2 mencapai 0,4. Kedua nilai tersebut melewati batas kepercayaan 95%, menunjukkan bahwa autokorelasi pada lag 1 dan 2 memiliki signifikansi statistik. Hal ini menyarankan adanya ketidaksesuaian dengan asumsi Gauss-Markov, terutama dalam konteks asumsi non-autokorelasi. Temuan ini diperkuat oleh hasil p-test dari Uji Durbin-Watson, yang menunjukkan nilai kurang dari 0,05.

Penanganan Kondisi Tak Standar Transformasi Weighted Least Square

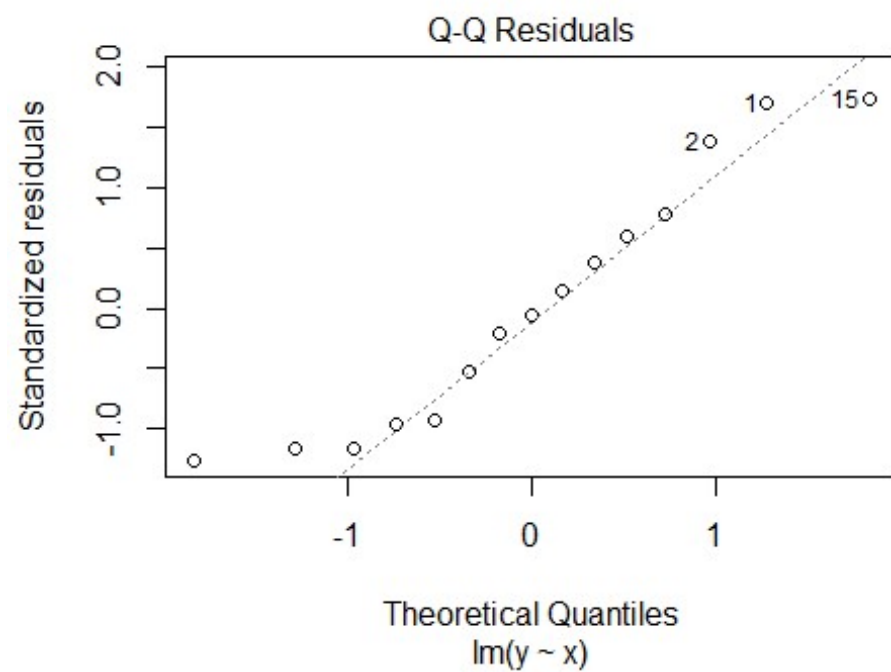
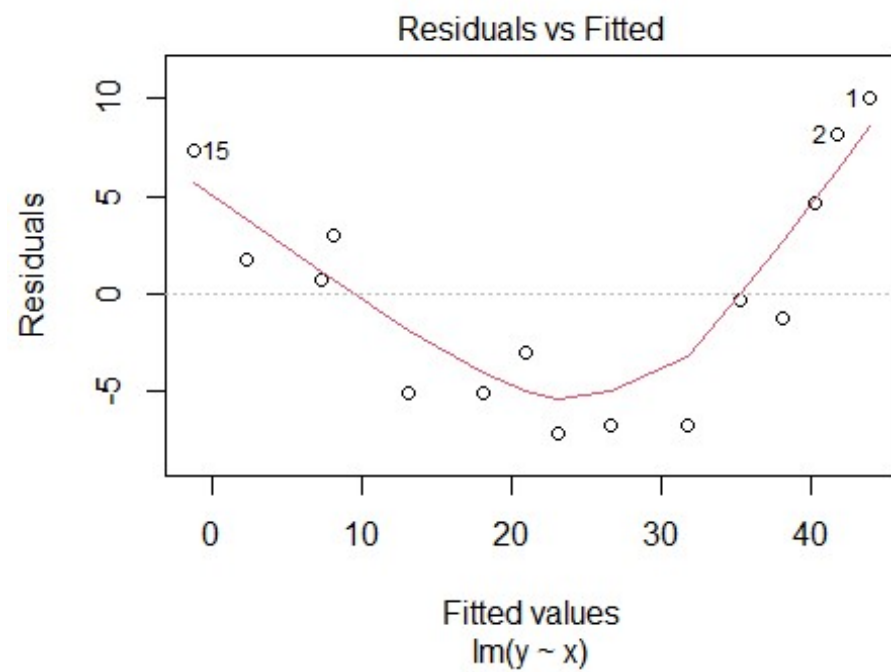
```
a <- abs(model$residuals)
b <- model$fitted.values
fit <- lm(a~b, anreg)
c <- 1 / fit$fitted.values^2
c
```

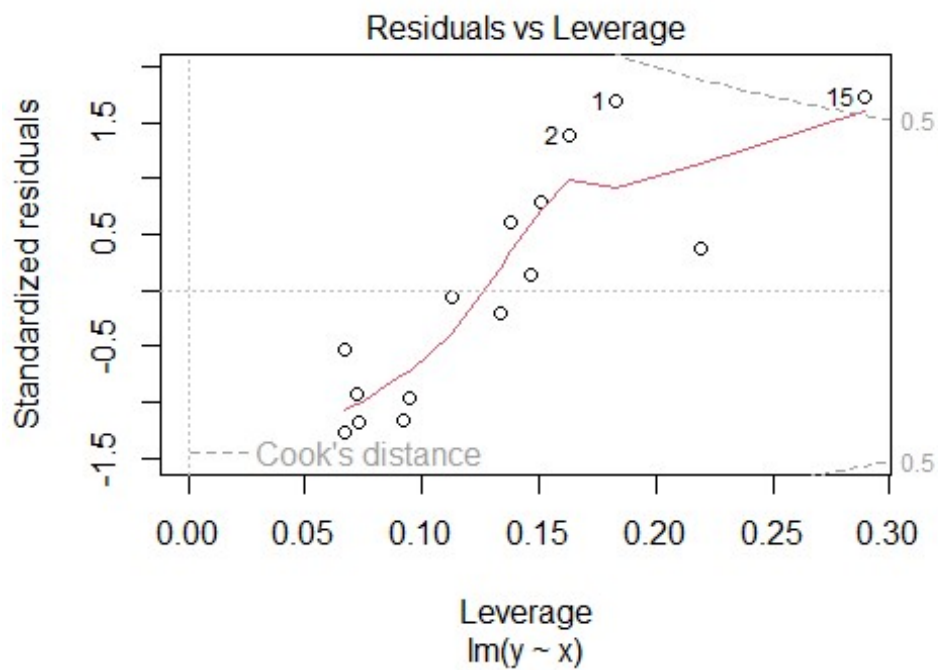
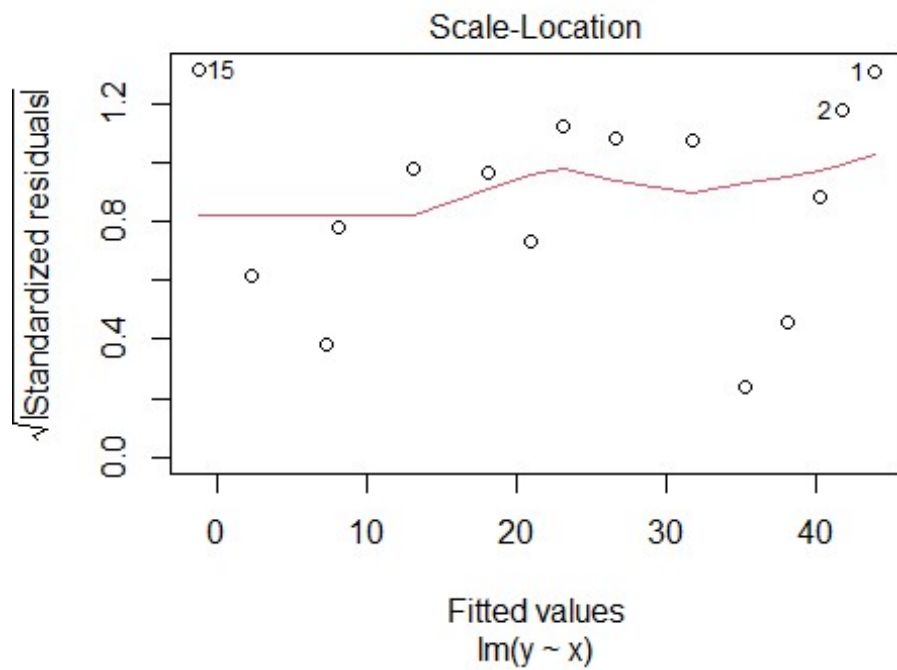
	1	2	3	4	5	6
##	0.03414849	0.03489798	0.03541143	0.03620311	0.03730067	0.03874425
7	0.04091034					
	8	9	10	11	12	13
##	0.04257072	0.04361593	0.04507050	0.04779711	0.05077885	0.05122749
14	0.05454132					

```
##          15  
## 0.05710924  
  
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
```

Penggunaan Weighted Least Squares (WLS) belum dianggap efektif karena masih terdapat ketidakpenuhan terhadap asumsi Gauss-Markov.

TRANSFORMASI AKAR PADA x, y, ATAU X DAN Y

```
library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.3.3
## Warning: package 'ggplot2' was built under R version 4.3.2
## Warning: package 'readr' was built under R version 4.3.3
## 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 4.3.3
## Warning: package 'lubridate' was built under R version 4.3.3

## — Attaching core tidyverse packages ————— tidyverse
2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ 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 4.3.3

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

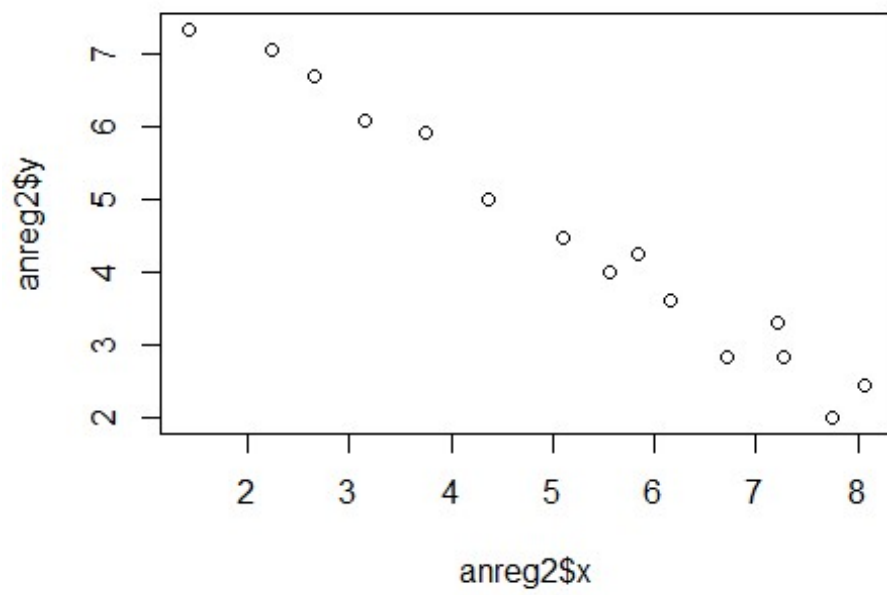
## The following object is masked from 'package:stats':
##
##   filter

## The following object is masked from 'package:graphics':
##
##   layout

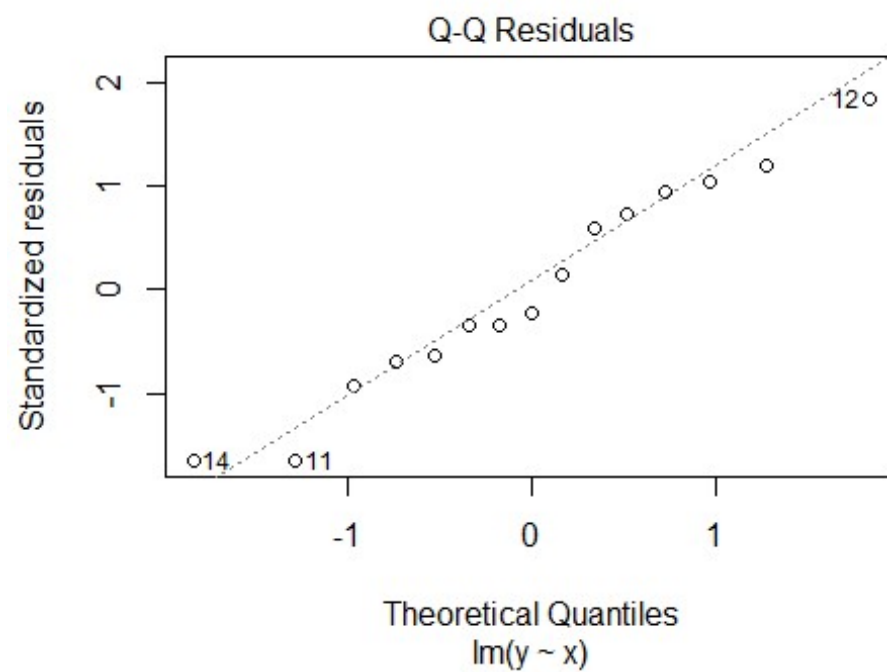
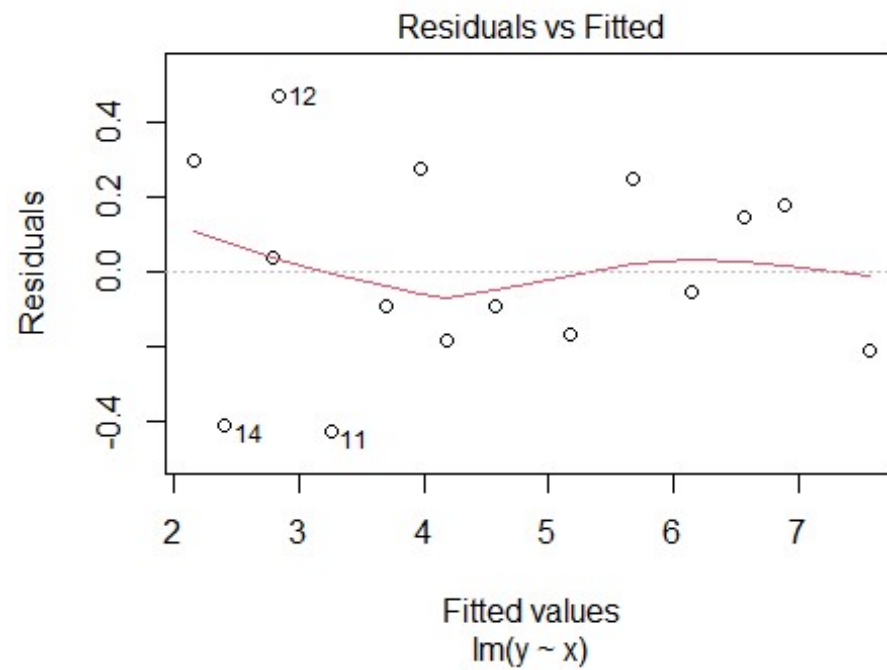
library(dplyr)
library(lmtest)
library(stats)

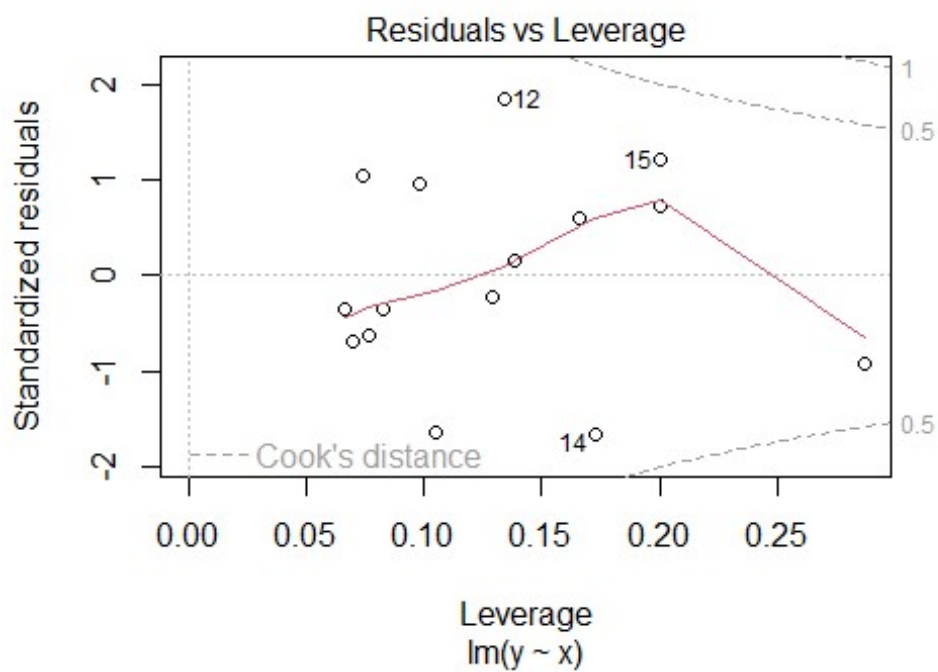
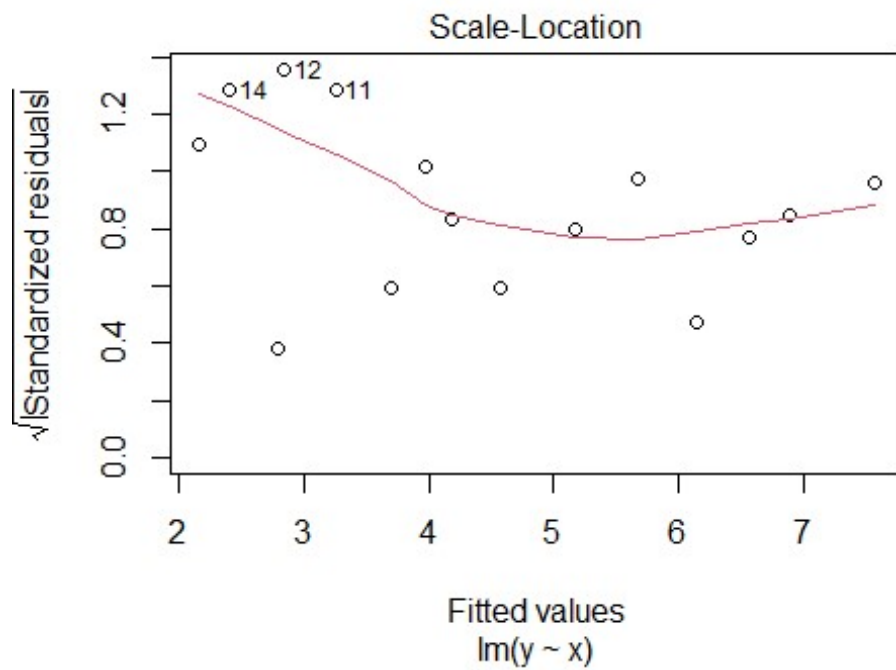
anreg2 <- anreg %>%
  mutate(y = sqrt(y)) %>%
  mutate(x = sqrt(x))
model3 <- lm(y~x, data = anreg2)
plot(x=anreg2$x, y=anreg2$y)

```



```
plot(model13)
```





```
summary(model3)
```

```
##
```

```
## Call:
```

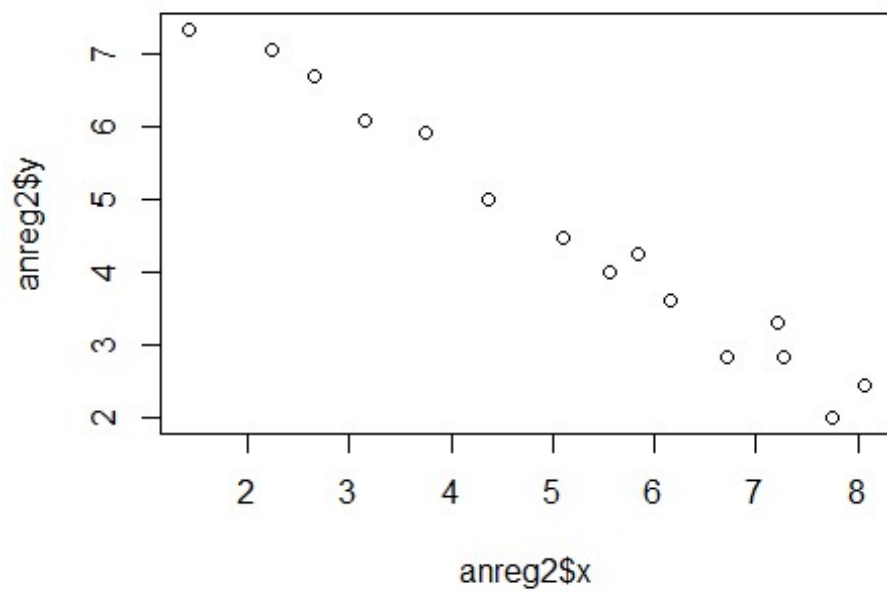
```
## lm(formula = y ~ x, data = anreg2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.42765 -0.17534 -0.05753  0.21223  0.46960
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   8.71245     0.19101   45.61 9.83e-16 ***
## x             -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
```

UJI AUTOKORELASI MODEL REGRESI TRANSFORMASI

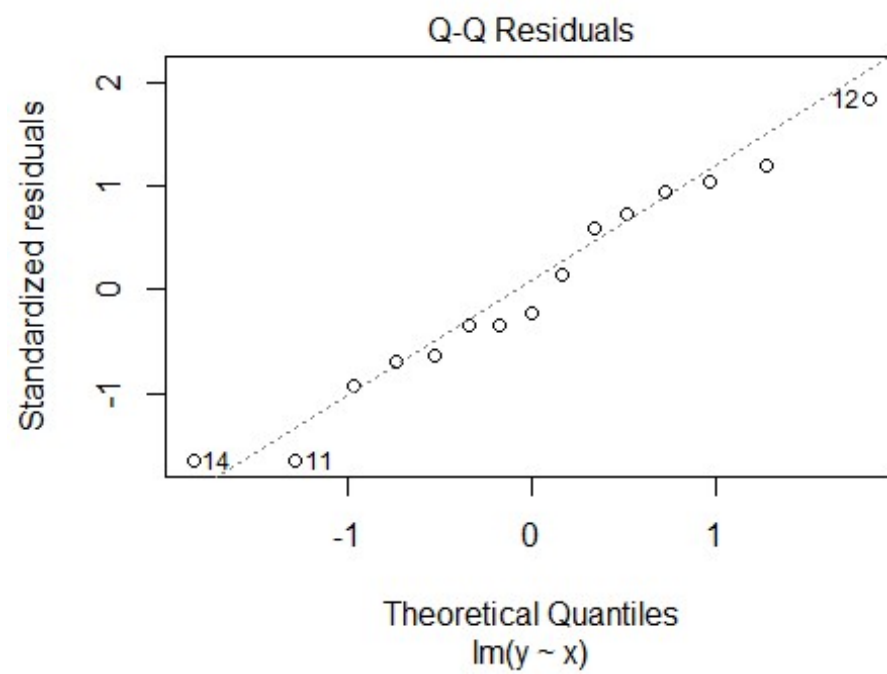
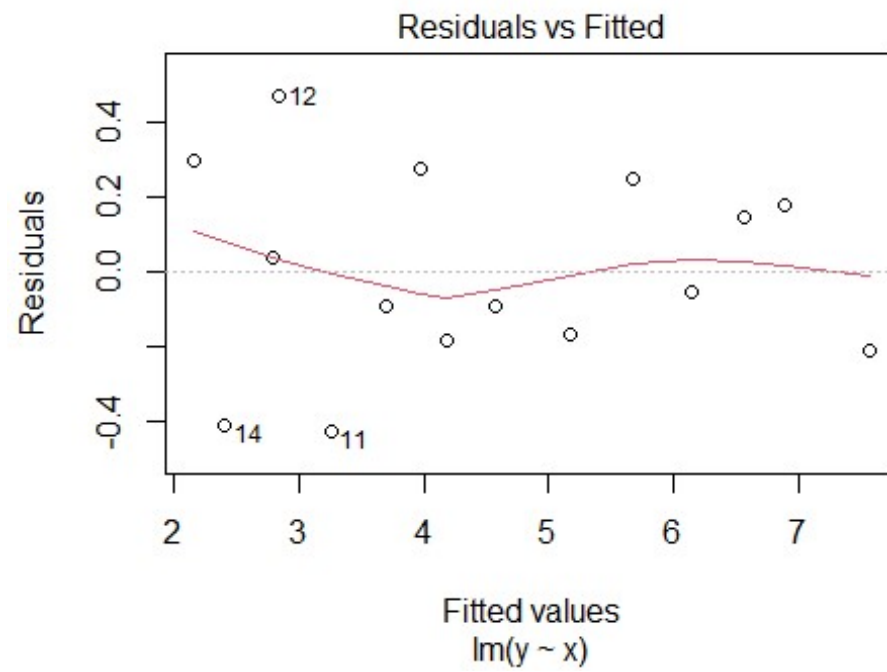
```
dwtest(model3)

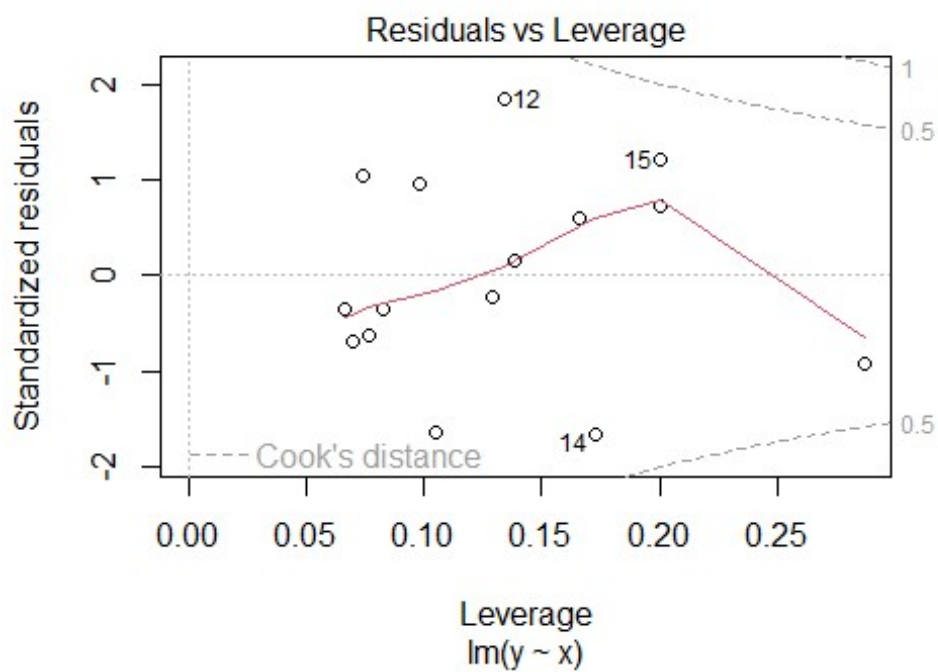
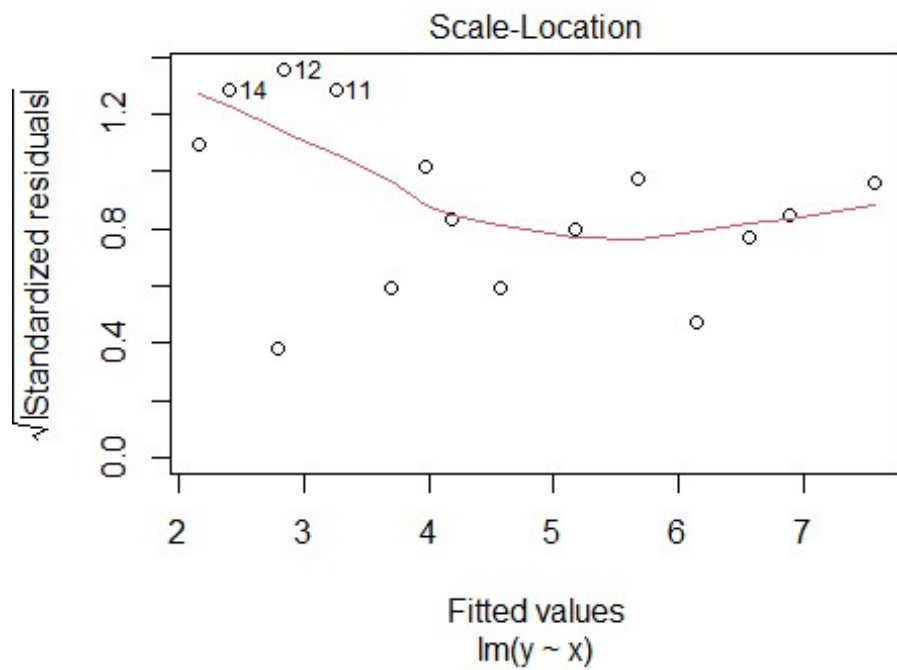
##
## Durbin-Watson test
##
## data: model3
## DW = 2.6803, p-value = 0.8629
## alternative hypothesis: true autocorrelation is greater than 0

model <- lm(y~x, data = anreg2)
plot(x = anreg2$x, y = anreg2$y)
```



```
plot(model13)
```





```
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 cukup bukti untuk menolak hipotesis nol, yang menyatakan bahwa tidak ada autokorelasi. Dari hasil transformasi tersebut, dapat disimpulkan bahwa transformasi akar Y berhasil meningkatkan efektivitas persamaan regresi. Model regresi setelah transformasi dapat dirumuskan sebagai berikut:

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

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

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

Dengan melakukan transformasi balik, kita mendapatkan:

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

Interpretasi model menunjukkan bahwa terdapat korelasi terbalik antara Y dan akar kuadrat dari X, dengan hubungan bersifat kuadratik. Semakin besar nilai akar kuadrat dari X, semakin kecil rata-rata nilai Y, dan tingkat penurunan ini semakin meningkat. Puncak kurva menunjukkan nilai rata-rata maksimum Y yang dapat dicapai 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 menandakan 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. Artinya, perubahan Y tidak proporsional dengan perubahan X, melainkan berubah dengan tingkat peningkatan yang semakin tinggi.