Tugas Individu Analisis Regresi Pertemuan 7

Ghonniyu Hiban Saputra

2024-03-05

Library

```
library(readxl)
```

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

```
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

## Warning: package 'ggplot2' was built under R version 4.3.2

##
## 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)
```

Data

```
data <- read_xlsx("C:/Users/Ghonniyu/Downloads/Anreg Individu.xlsx")
```

```
## New names:
## * `` -> `...1`
```

```
data
```

```
## # A tibble: 15 x 3
##     ...1     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
##  7     7    26    20
##  8     8    31    16
##  9     9    34    18
## 10    10    38    13
## 11    11    45     8
## 12    12    52    11
## 13    13    53     8
## 14    14    60     4
## 15    15    65     6
```

Model Awal

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

```
##
## Call:
## lm(formula = Y ~ X, 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 regresi sebagai berikut

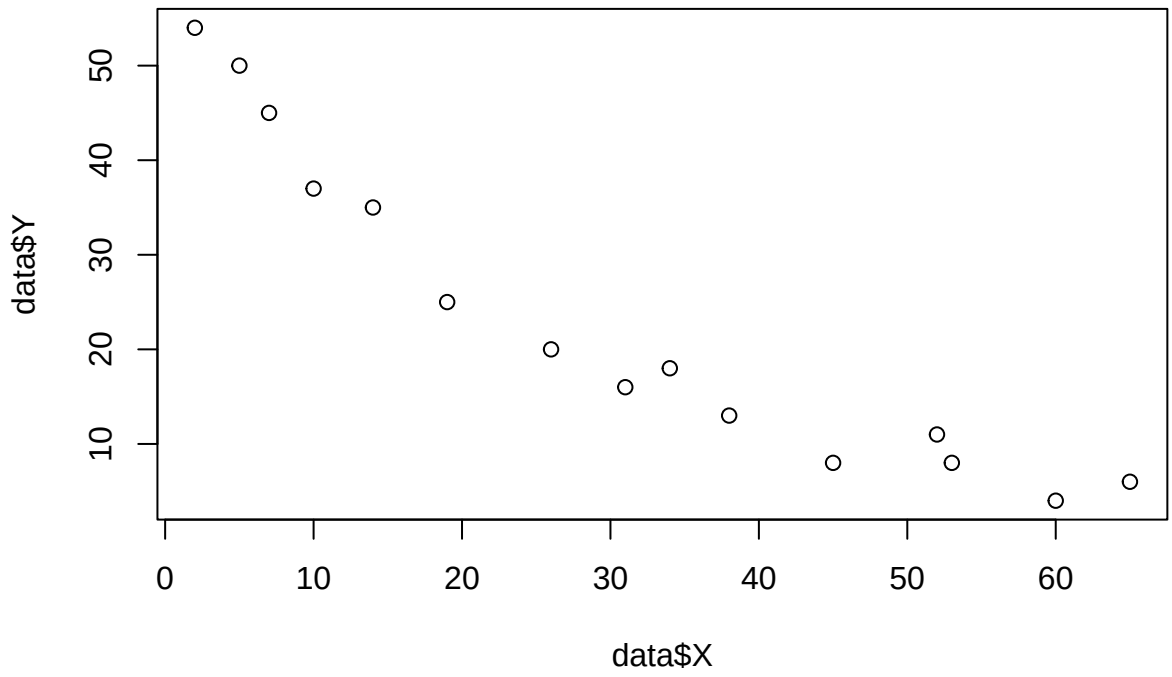
$$\hat{Y} = 46.46041 - 0.75251X + e$$

Model diatas belum bisa dikatakan menjadi model terbaik karena belum melalui serangkaian uji asumsi, sehingga diperlukan eksplorasi kondisi dan pengujian asumsi Gauss Markov dan normalitas untuk menghasilkan model terbaik.

Eksplorasi

Plot Hubungan X dan Y

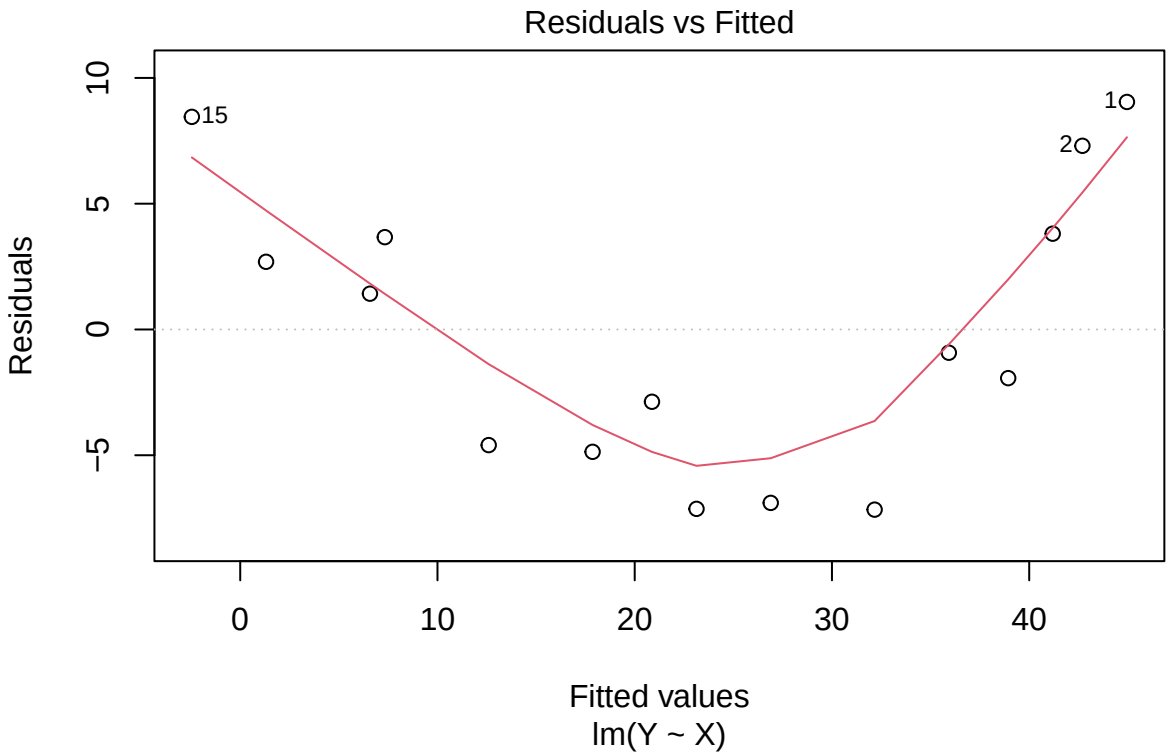
```
plot(x=data$X,y=data$Y)
```



Dari plot diatas dapat ditarik kesimpulan hubungan X dan Y memiliki hubungan yang tidak linear dan cenderung membentuk parabola

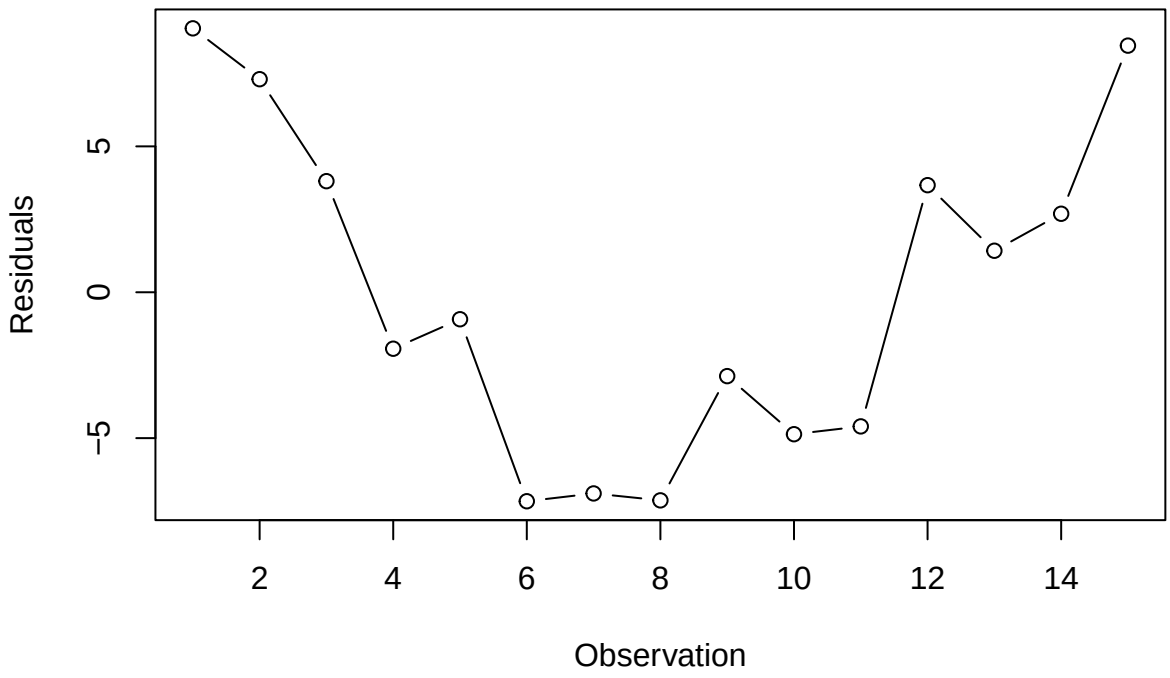
Plot Sisaan vs Y duga

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



Plot Sisaan vs Urutan

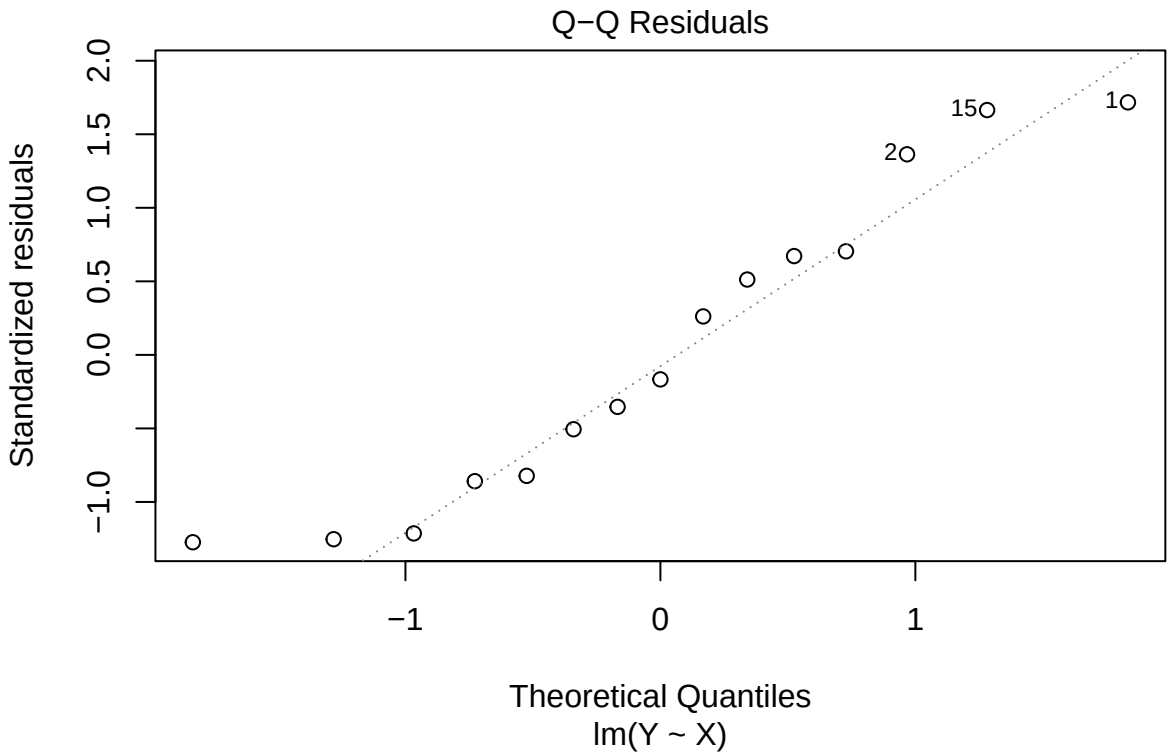
```
plot(x = 1:dim(data)[1],
y = model$residuals,
type = 'b',
ylab = "Residuals",
xlab = "Observation")
```



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

Sisaan dengan QQ-Plot

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



Uji Normalitas Sisaan

H_0 : Sisaan menyebar normalH_1 : Sisaan tidak menyebar normal

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

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

Dapat dilihat p-value > alpha, maka tak tolak H_0, sisaan menyebar normal

Uji Asumsi Gauss-Markov

1. Nilai Harapan Sisaan sama dengan nol

H_0 : Nilai harapan sama dengan 0H_1 : Nilai harapan tidak sama dengan 0

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

```
##
##  One Sample t-test
```

```
##
## data:  model$residuals
## t = -3.5338e-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
## -5.179884e-16
```

Dapat dilihat p-value > alpha , maka tak tolak H_0, nilai harapan sisaan sama dengan nol

2. Ragam Sisaan Homogen

H_0 : Ragam sisaan homogenH_1 : 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
```

Dapat dilihat p-value > alpha, maka tak tolak H_0, ragam sisaan homogen

3. Sisaan Saling Bebas

$$H_0 : \text{Sisaan saling bebas} H_1 : \text{Sisaan tidak saling bebas}$$

```
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)
```

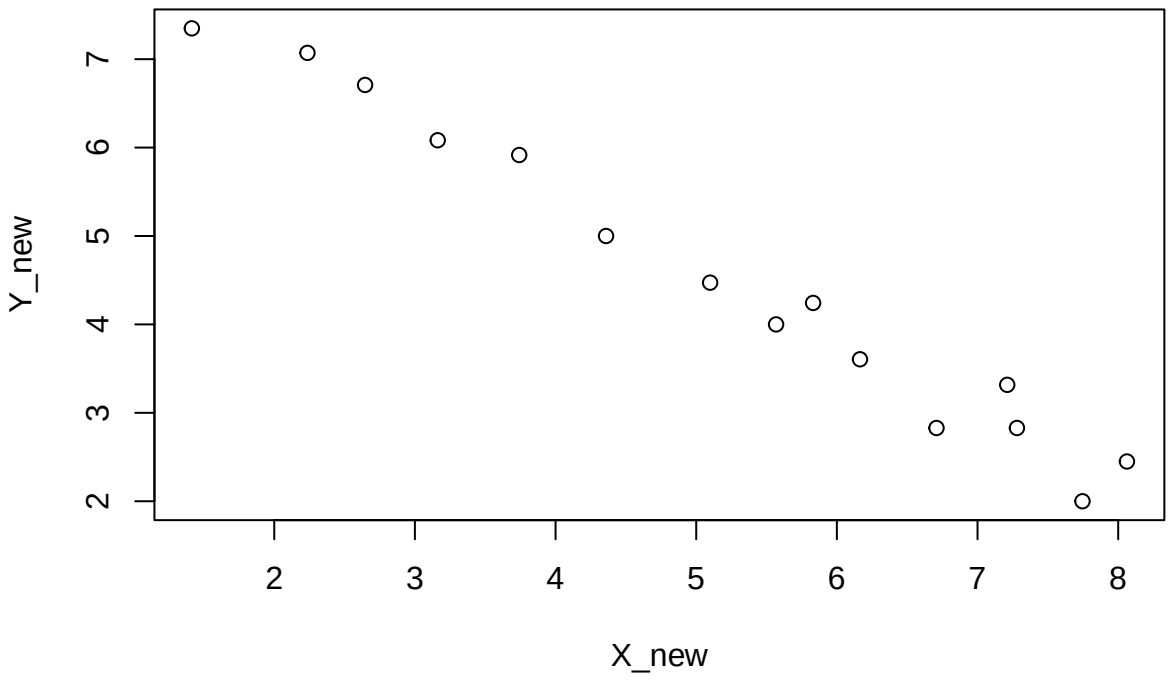
```
##
##  Durbin-Watson test
##
## data:  model
## DW = 0.48462, p-value = 1.333e-05
## alternative hypothesis: true autocorrelation is greater than 0
```

Dapat dilihat p-value (pada DW test) $<$ alpha , maka tolak H_0, sisaan tidak saling bebas.

Dari ketiga asumsi diatas menunjukkan adanya asumsi Gauss-Markov yang tidak terpenuhi yaitu pelanggaran asumsi tidak ada autokorelasi. Dapat dilihat dari Durbin Watson Test yang dilakukan.

Transformasi Data

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




```
model_new=lm(formula=Y_new~X_new)
summary(model_new)
```

```
##
## Call:
## lm(formula = Y_new ~ X_new)
##
## 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_new       -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

$$Y = 8.71245 - 0.81339X + e$$

```
dwtest(model_new)
```

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

Karena p-value>alpha, maka tak tolak H_0, sisaan saling bebas.

Kesimpulan

Berdasarkan transformasi yang dilakukan yaitu transformasi $Y^{\frac{1}{2}}$dan$X^{\frac{1}{2}}$ akan menghasilkan model regresi yang lebih efektif disertai dengan semua asumsi telah terpenuhi dalam analisis regresi linear sederhana.

Model regresi setelah di transformasi adalah sebagai berikut :

$$Y^* = 8.71245 - 0.81339X^* + eY^* = \sqrt{Y}X^* = \sqrt{X}$$

Sehingga model terbaik untuk data ini adalah:

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