Latihan Pertemuan 7 Analisis Regresi

Adinda Shabrina Putri Salsabila (G1401221081)

2024-03-05

Import Data

```
datalatihan <- read_xlsx("C:/Users/hp/Documents/KULIAH/SEMESTER 
4/ANREG/KULIAH/DataAnreg.xlsx")
datalatihan
```

```
## # A tibble: 15 × 2
##        X     Y
##    <dbl> <dbl>
##  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
```

Model Awal

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

```
## 
## Call:
## lm(formula = Y ~ X, data = datalatihan)
## 
## 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
```

Model persamaan regresi linear sederhana yang diperoleh :

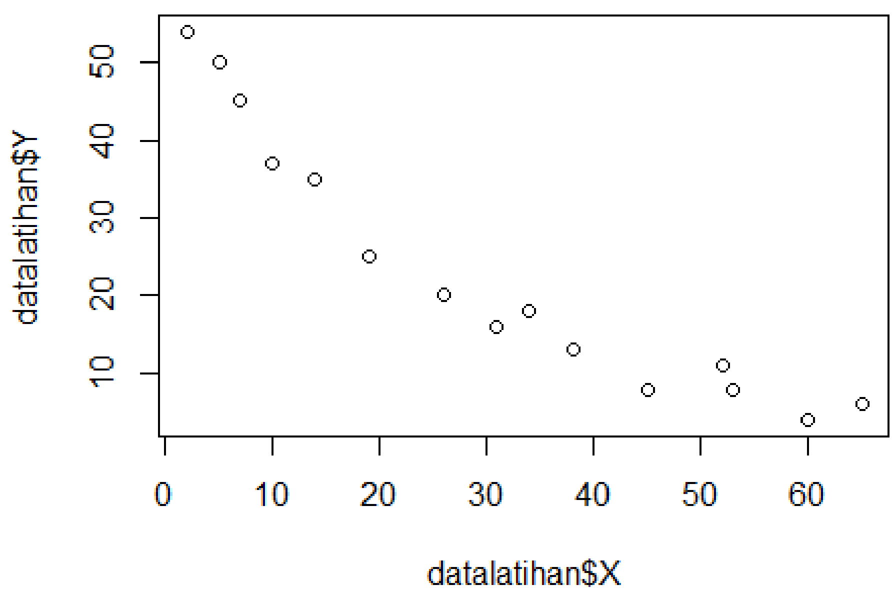
$$\hat{Y} = 46.46041 - 0.75251X + \varepsilon$$

Perlu dilakukan serangkaian uji asumsi untuk memastikan apakah model tersebut merupakan model terbaik. Hal ini bisa dilakukan dengan eksplorasi, pengujian asumsi Gauss-Marcov dan Normalitas.

Eksplorasi

Plot Hubungan X dan Y

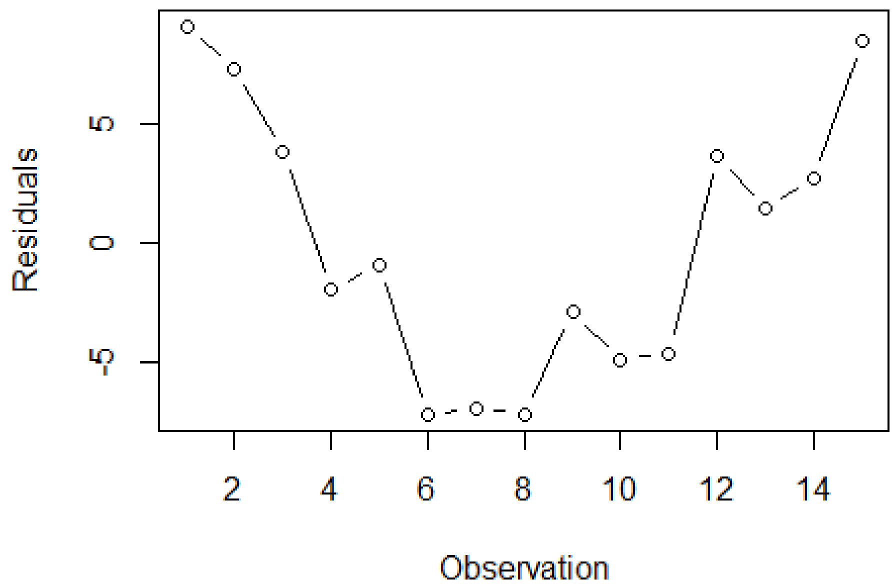
```
plot(x = datalatihan$X,y = datalatihan$Y)
```



Hasil dari plot di atas menggambarkan hubungan X dan Y yang tidak linier, menyerupai parabola.

Plot Sisaan Berdasarkan Urutan Data

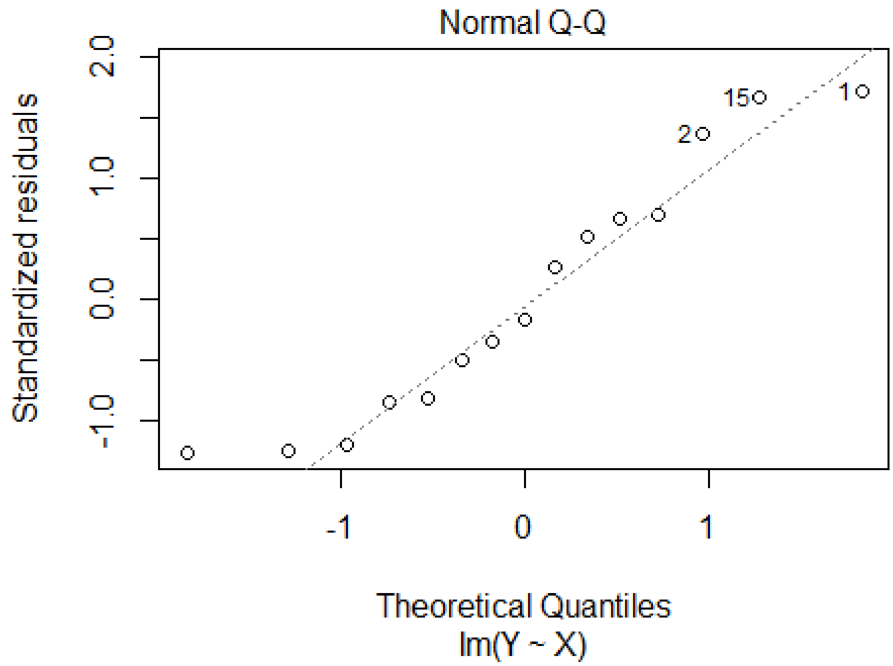
```
plot(x = 1:dim(datalatihan)[1],
     y = model$residuals,
     type = 'b',
     ylab = "Residuals",
     xlab = "Observation")
```



Hasil dari plot di atas membentuk sebuah pola kurva yang mana sisaan tidak saling bebas dan model tidak pas

Normalitas Sisaan dengan QQ-Plot

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



Uji Formal Asumsi

Kondisi Gauss Marcov

1. Nilai Harapan Sisaan sama dengan Nol

$$ H_0: \text{Nilai harapan sisaan sama dengan nol}\\ H_1: \text{Nilai harapan sisaan tidak sama dengan nol }\ $$

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

## 
##  One Sample t-test
## 
## data:  model$residuals
## t = -4.9493e-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 
## -7.254614e-16
```

P-value = 1 > alpha = 0.05, maka tak tolak H0 yang berarti nilai harapan sisaan sama dengan nol

2. Ragam Sisaan Homogen

Uji formal untuk mendeteksi homogenitas ragam sisaan dapat dilakukan dengan uji Breusch-Pagan menggunakan fungsi bptest. Uji ini memiliki hipotesis sebagai berikut

$$ H_0: var [ϵ] = σ^2I \text {(ragam sisaan homogen)} \\ H_1: var[ϵ] ≠ σ^2I \text {(ragam tidak homogen)}\ $$

```
kehomogenan = lm(formula = abs(model$residuals) ~ X, # y: abs residual
    data = datalatihan)
summary(kehomogenan)
```

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

```
ncvTest(model)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.1962841, Df = 1, p = 0.65774
```

P-value = 0.4674 > alpha = 0.05, maka tak tolak H0 yang berarti ragam sisaan homogen

3. Sisaan Saling Bebas

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

```
dwtest(model)
```

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

Karena p-value = 1.333e-05 < alpha = 0.05, maka tolak H0 yang berarti sisaan tidak saling bebas, asumsi tidak terpenuhi

Uji Formal Normalitas Sisaan

H0: Sisaan menyebar normal
H1: Sisaan tidak menyebar normal

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

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

Karena p-value = 0.226 > alpha = 0.05, maka tak tolak H0, sisaan menyebar normal

Autokorelasi

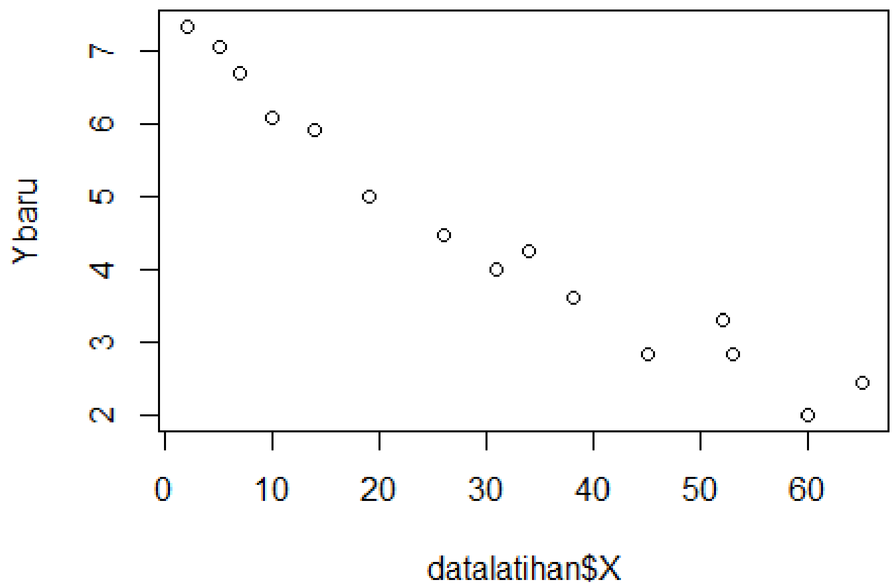
```
(uji_autokol <- durbinWatsonTest(model,
                                 alternative="two.sided"))
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1        0.587863      0.484617       0
##  Alternative hypothesis: rho != 0
```

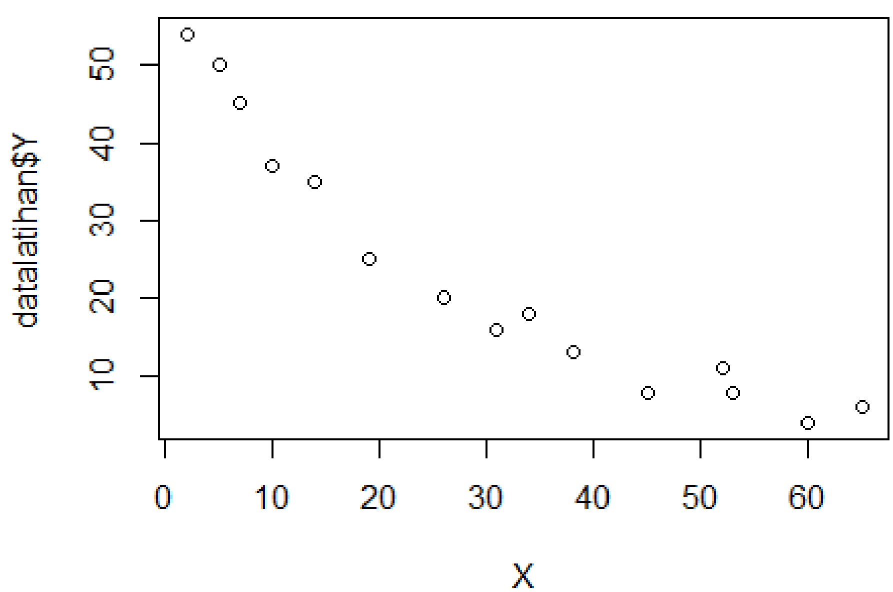
Nilai p-value < 0,05 bermakna tolak H0. Hal ini mengindikasikan bahwa dalam taraf nyata 5%, ada cukup bukti untuk menyatakan bahwa terdapat autokorelasi dalam sisaan.

Transformasi Data

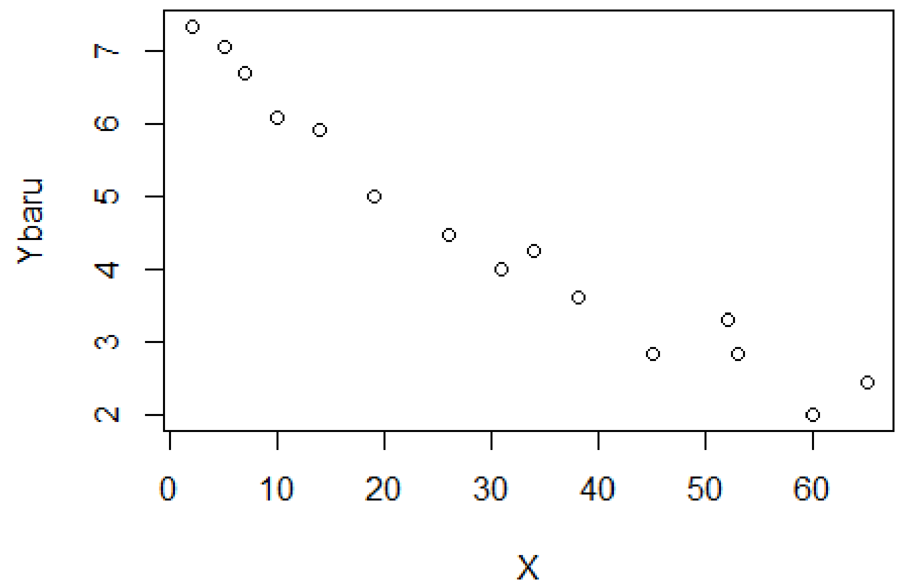
```
Ybaru = sqrt(datalatihan$Y)
X = datalatihan$X

plot(x = datalatihan$X,y = Ybaru)
```




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




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




```
model2 = lm(formula = Ybaru ~ X)
summary(model2)

##
## Call:
## lm(formula = Ybaru ~ X)
##
## 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
```

Diperoleh model persamaan regresi sebagai berikut

$$\hat{Y} = 7.015455 - 0.081045 + \varepsilon$$

Uji Asumsi Setelah Transformasi

```
shapiro.test(model2$residuals)

##
##  Shapiro-Wilk normality test
##
## data:  model2$residuals
## W = 0.93128, p-value = 0.2852
```

Karena p-value = 0.2852 > alpha = 0.05, maka tak tolak H0, sisaan menyebar normal

```
(uji_autokol <- durbinWatsonTest(model2,
                                 alternative="two.sided"))

##  lag Autocorrelation D-W Statistic p-value
##    1       0.2362986      1.220617    0.05
##  Alternative hypothesis: rho != 0
```

P-value = 0.074 > alpha = 0.05, maka tak tolak H0. Hal ini mengindikasikan bahwa dalam taraf nyata 5%, tidak cukup bukti untuk menyatakan bahwa terdapat autokorelasi dalam sisaan.

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

##
##  One Sample t-test
##
## data:  model$residuals
## t = -4.9493e-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
## -7.254614e-16
```

P-value = 1 > alpha = 0.05, maka tak tolak H0 yang berarti nilai harapan sisaan sama dengan nol

```
library(lmtest)
(model2 <- bptest(model2))

##
##  studentized Breusch-Pagan test
##
## data:  model2
## BP = 0.93605, df = 1, p-value = 0.3333
```

Karena p-value = 0.578 > alpha = 0.05, maka tak tolak H0 yang berarti ragam sisaan homogen

```r
runs.test(as.numeric(model2$residuals))

##
##  Runs Test
##
## data:  as.numeric(model2$residuals)
## statistic = NaN, runs = 0, n1 = 0, n2 = 0, n = 0, p-value = NA
## alternative hypothesis: nonrandomness
```

Karena p-value = 0.578 > alpha = 0.05, maka tak tolak H0 yang berarti sisaan saling bebas dan semua asumsi terpenuhi

Kesimpulan

Berdasarkan transformasi $Y^{\frac{1}{2}}$ akan menghasilkan model regresi linear sederhana yang lebih efektif degan semua asumsi yang sudah terpenuhi. Model regresi setelah di transformasi adalah sebagai berikut :

$$Y^* = 7.015455 - 0.081045 + \varepsilon$$

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

$$\hat{Y} = (7.015455 - 0.081045)^2 + \varepsilon$$

Jadi, secara keseluruhan, interpretasi model regresi linear sederhana ini adalah bahwa terdapat hubungan linier negatif antara variabel independen X dan variabel dependen Y, dengan setiap peningkatan satu satuan dalam X, diharapkan nilai Y menurun sebesar 0.081045, dengan intersep yang menunjukkan nilai Y ketika ketika X=0 (meskipun mungkin tidak selalu memiliki interpretasi praktis).