

Tugas Analisis Regresi K1

Elke Frida Rahmawati - G1401221025

2024-03-05

Data

```
bookdown::pdf_document2
```

```
## function (...)  
## {  
##   pdf_book(..., base_format = rmarkdown::pdf_document)  
## }  
## <bytecode: 0x0000023b870837b0>  
## <environment: namespace:bookdown>
```

```
library(car)
```

```
## Loading required package: carData
```

```
library(dplyr)
```

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

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##   recode
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##   filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##   intersect, setdiff, setequal, union
```

```
library(readxl)
```

```
data_ar<-read_xlsx("C:/Users/user/OneDrive/Dokumen/semester 4/Data.xlsx")
```

```
data_ar
```

```
## # A tibble: 15 × 3
```

```
##       no      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
```

```
View(data_ar)
```

Eksplorasi Data

```
library(GGally)
```

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

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
```

```
##   method from
```

```
##   +.gg      ggplot2
```

```
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.3.3
```

```
##
```

```
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
```

```
##
```

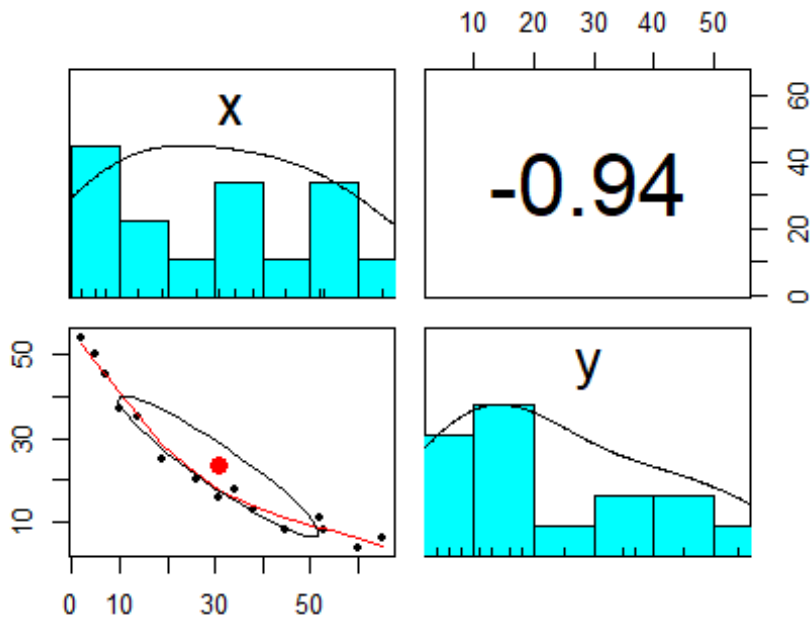
```
##   %+%, alpha
```

```
## The following object is masked from 'package:car':
```

```
##
```

```
##   logit
```

```
pairs.panels(data_ar<-data_ar[, -1])
```



Pemodelan Regresi Linear

Pemodelan Regresi

```
reg_ar<-lm(y~x, data=data_ar)
summary(reg_ar)
```

```
##
## Call:
## lm(formula = y ~ x, data = data_ar)
##
## 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 metode regresi awal sebelum pemilihan metode regresi terbaik, yaitu sebagai berikut.

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

Uji Asumsi Model

Uji Normalitas

Hipotesis:

\$\$ H_0 : \text{\textbackslash Data\ menyebar\ secara\ normal\ (p-value > \{\alpha\}) } \text{\textbackslash\textbackslash H}_1 : \text{\textbackslash Data\ menyebar\ secara\ tidak\ normal\ (p-value \leq \{\alpha\}) } \text{\textbackslash } \$\$

```
shapiro.test(data_ar$y)

##
##  Shapiro-Wilk normality test
##
## data:  data_ar$y
## W = 0.89636, p-value = 0.08374
```

Dengan α
= 0.05, dapat diputuskan bahwa H_0 diterima karena ($p - \text{value} = 0.08374 > \alpha = 0.05$), sehingga dapat diputuskan bahwa data menyebar secara normal.

Uji Multikolinieritas

Uji multikolinearitas tidak dapat dilakukan karena variabel independen yang ada dalam data hanya satu yaitu variabel x. Artinya tidak ada multikolinearitas antarpeubah bebas. Oleh karena itu, di bawah hanya dicantumkan perhitungan koefisien korelasi Pearson. Koefisien korelasi Pearson mengukur seberapa kuat dan searah hubungan linear antara dua variabel.

```
correlation <- cor(data_ar$x, data_ar$y)
correlation

## [1] -0.9410528
```

Nilai yang diperoleh adalah -0.94 dimana dapat diartikan bahwa kedua variabel memiliki hubungan negatif yang sangat kuat antara dua variabel.

Uji Homoskedastisitas

Hipotesis:

\$\$ H_0 : \text{\textbackslash Ragam\ galat\ bersifat\ homoskedastisitas\ (p-value > \{\alpha\}) } \text{\textbackslash\textbackslash H}_1 : \text{\textbackslash Ragam\ galat\ bersifat\ heteroskedastisitas\ (p-value \leq \{\alpha\}) } \text{\textbackslash } \$\$

```
lmtest::bptest(reg_ar)
```

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

Dengan $\alpha = 0.05$, dapat diputuskan bahwa H_0 diterima karena ($p - value = 0.4674 > \alpha = 0.05$), sehingga dapat diputuskan bahwa ragam galat bersifat homoskedastisitas.

Uji Autokorelasi

Hipotesis:

$H_0: E\{\{\varepsilon_i, \varepsilon_j\}\} = 0$ (tidak ada autokorelasi) $H_1: E\{\{\varepsilon_i, \varepsilon_j\}\} \neq 0$ (ada autokorelasi)

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

Dengan $\alpha = 0.05$, dapat diputuskan bahwa H_0 ditolak karena ($p - value = 0.48462 > \alpha = 0.05$), sehingga dapat diputuskan bahwa terdapat autokorelasi dalam sisaan.

Seleksi Peubah/Pemilihan Model Regresi Terbaik

Metode Backward

```
olsrr::ols_step_backward_p(reg_ar)
## [1] "No variables have been removed from the model."
```

Tidak ada variabel yang dihilangkan karena hanya ada satu variabel independen sehingga tidak ada variabel lain yang dapat dihapus.

Metode Forward

```
olsrr::ols_step_forward_p(reg_ar)
##
##
## Stepwise Summary
## -----
## Step Variable AIC SBC SBIC R2 Adj. R2
## -----
```

```

## 0      Base Model      130.144      131.560      84.333      0.00000      0.00000
## 1      x                99.626      101.750      57.626      0.88558      0.87678
## -----
##
## Final Model Output
## -----
##
##                               Model Summary
## -----
## R                0.941          RMSE                5.484
## R-Squared        0.886          MSE                34.707
## Adj. R-Squared   0.877          Coef. Var          25.248
## Pred R-Squared   0.839          AIC                99.626
## MAE              4.852          SBC                101.750
## -----
## RMSE: Root Mean Square Error
## MSE: Mean Square Error
## MAE: Mean Absolute Error
## AIC: Akaike Information Criteria
## SBC: Schwarz Bayesian Criteria
##
##                               ANOVA
## -----
##                               Sum of
##                               Squares      DF      Mean Square      F      Sig.
## -----
## Regression      3492.139          1      3492.139      100.617      0.0000
## Residual        451.195          13      34.707
## Total           3943.333          14
## -----
##
##                               Parameter Estimates
## -----
##
##                               Beta      Std. Error      Std. Beta      t      Sig
## lower      upper
## -----
## (Intercept)    46.460          2.762          16.820      0.000
## 40.493      52.428
## x              -0.753          0.075          -0.941     -10.031      0.000      -
## 0.915      -0.590
## -----
## -----

```

Selain karena alasan hanya ada variabel independen, nilai-t hitung yang diperoleh < tingkat signifikansi, maka variabel x dapat dianggap signifikan dalam model. Oleh karena itu, tidak ada proses eliminasi variabel yang dapat dilakukan.

Metode Stepwise

```
step(lm(y~x,data=data_ar),direction="both")
```

```
## Start:  AIC=55.06
## y ~ x
##
##           Df Sum of Sq    RSS    AIC
## <none>                 451.2 55.058
## - x           1    3492.1 3943.3 85.576

##
## Call:
## lm(formula = y ~ x, data = data_ar)
##
## Coefficients:
## (Intercept)                x
##      46.4604         -0.7525

reg_ar <- lm(y ~ x, data = data_ar)
reg_ar

##
## Call:
## lm(formula = y ~ x, data = data_ar)
##
## Coefficients:
## (Intercept)                x
##      46.4604         -0.7525
```

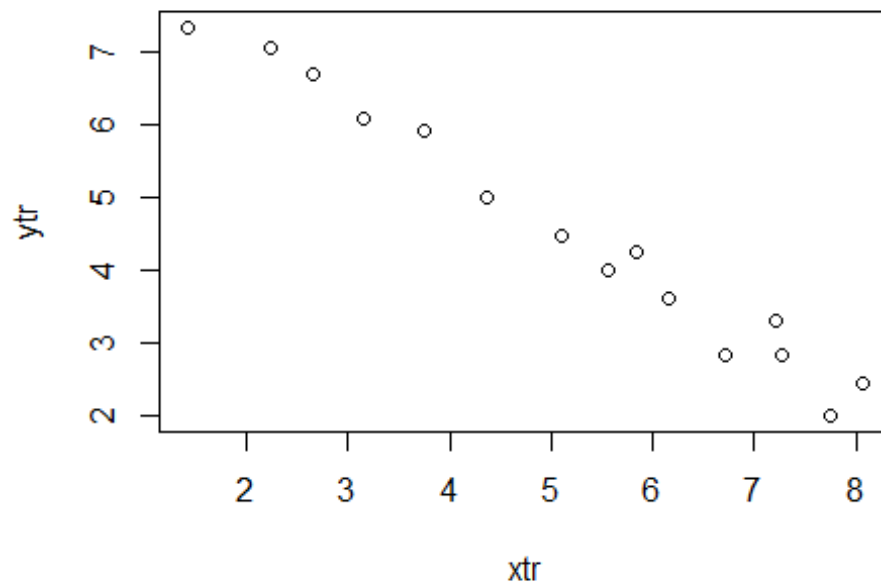
Jika hanya ada satu variabel independen (x) dan satu variabel dependen (y), maka metode stepwise tidak akan bermanfaat karena tidak ada pilihan untuk menambah atau menghapus variabel independen dari model. Metode stepwise biasanya digunakan untuk memilih subset terbaik dari sekumpulan variabel independen yang tersedia.

Dari hasil perhitungan dengan metode stepwise, nilai variabel dependen (y) diprediksi akan sekitar 46.4604 ketika variabel independen (x) adalah nol, dan setiap penambahan satu unit dalam variabel independen (x) akan menghasilkan penurunan sebesar -0.7525 unit dalam nilai variabel dependen (y). Berdasarkan prosedur seleksi peubah dengan metode stepwise, model terbaik yang terbentuk adalah model dengan empat peubah penjelas.

Penyesuaian Data

Transformasi Data

```
ytr <- sqrt(data_ar$y)
xtr <- sqrt(data_ar$x)
plot(x=xtr, y=ytr)
```



Model dan Pemeriksaan Asumsi

```
model <- lm(ytr~xtr)
summary(model)
```

```
##
## Call:
## lm(formula = ytr ~ xtr)
##
## 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 ***
## xtr          -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

lmtest::dwtest(model)
```



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

Dengan $\alpha = 0.05$, dapat diputuskan bahwa H_0 diterima karena ($p - value = 0.8629 > \alpha =$

0.05), sehingga dapat diputuskan bahwa tidak ada autokorelasi dalam sisaan atau galat saling be

Dengan demikian, diperoleh model terbaik diperoleh setelah melakukan transformasi pada variabel x dan y ke dalam bentuk akar yaitu sebagai berikut.

$$\hat{Y} = 8.71245 - 0.81339X$$

Model ini telah memenuhi asumsi-asumsi dalam regresi linear