

Regresi Linear

Stevanus Sembiring/162112133099

2022-10-05

```
library(readxl)
df = read_excel("Tingkat Pengangguran.xlsx")
```

1. Model regresi dan menentukan variabel prediktor dan respon

```
# Menentukan variabel prediktor dan respon
x1 = df$`Kepadatan (Jiwa/m2)`
x2 = df$`Proporsi remaja dengan TIK (%)`
y = df$`Tingkat Pengangguran Terbuka (%)`
```

Variabel Tingkat Pengangguran Terbuka adalah sebagai Variabel respon (y) sedangkan variabel Proporsi remaja dengan TIK dan Kepadatan adalah variabel prediktor (x).

```
# Model regresi
model1 = lm(y~x1+x2, data = df)
model1
```

```
##
## Call:
## lm(formula = y ~ x1 + x2, data = df)
##
## Coefficients:
## (Intercept)          x1          x2
##  1.7346537    0.0002008    0.0407785
```

$$Y(Kadar) = 1.73456537 + 0.0002008X1(kepadatan) + 0.0407785X2(TIK)$$

2. Estimasi model menggunakan excel

Hasil estimasi model menggunakan excel ada di dalam folder

3. Uji Asumsi Klasik

Uji Variansi error konstan

1. Menguji ketidaksamaan variansi dari residual dengan menggunakan **uji glejser**

H_0 : Data bersifat Homogen
 H_1 : Data bersifat Heterogen
 α : 5%

```

# Uji Glejser
library(skedastic)
glejser(model1)

```

```

## # A tibble: 1 x 4
##   statistic p.value parameter alternative
##       <dbl>   <dbl>       <dbl> <chr>
## 1      2.11   0.348         2 greater

```

Dari hasil uji glejser didapatkan P-Value = 0.3477001, sehingga kesimpulannya **Gagal Tolak H_0** , karena P-Value(0.3477001) > $\alpha(0,05)$. Kesimpulannya, data ini **tidak** melanggar asumsi Homoskedastisitas (data bersifat homogen).

Uji Independensi Error (autokorelasi)

menguji apakah dalam model regresi linier ada korelasi antara error dari observasi satu dan lainnya. Uji yang digunakan adalah **Uji Durbin-Watson**.

H_0 : Tidak ada autokorelasi

H_1 : Terdapat autokorelasi

α : 5

```

# Uji Durbin Watson
library(lmtest)
dwtest(model1)

```

```

##
## Durbin-Watson test
##
## data: model1
## DW = 2.0095, p-value = 0.4524
## alternative hypothesis: true autocorrelation is greater than 0

```

Dari hasil pengujian Durbin-Watson, P-Value yang didapat adalah 0.4524. Sehingga kesimpulan adalah **Gagal Tolak H_0** , karena P-Value(0.4524) > (0,05). Sehingga dapat disimpulkan bahwa data tersebut **Tidak ada autokorelasi**.

Uji Normalitas

Menguji error berdistribusi normal.

H_0 : Error berdistribusi normal

H_1 : Error tidak berdistribusi normal

α : 5%

```
# Uji Kolmogorov-Smirnov (KS)
```

```
error = model1$residuals
```

```
library(stats)
```

```
ks.test(error, "pnorm")
```

```
##
```

```
## One-sample Kolmogorov-Smirnov test
```

```
##
```

```
## data: error
```

```
## D = 0.17509, p-value = 0.2207
```

```
## alternative hypothesis: two-sided
```

Dari pengujian Kolmogorov-Smirnov (KS) di atas, P-Value yang didapat sama dengan 0.2207, maka **Gagal Tolak H_0** karena $P\text{-value}(0.2207) > \alpha(0,05)$ dan dapat disimpulkan bahwa dalam data ini error telah berdistribusi normal.

Uji Multikolinieritas

Menguji hubungan antara variabel prediktor dengan melihat nilai **Variance Inflation Factor (VIF)**, Nilai VIF ketika tidak terjadi Multikolinieritas adalah < 5 .

```
# Nilai VIF
```

```
library(regclass)
```

```
VIF(model1)
```

```
##          x1          x2
```

```
## 1.041007 1.041007
```

Dari hasil Uji Multikolinieritas di dapatkan hasil VIF dari variabel X1 dan X2 < 5 sehingga dapat disimpulkan bahwa **Tidak terjadi Multikolinieritas**

Dari hasil pengujian asumsi klasik yang telah dilakukan didapatkan **TIDAK ADANYA** pelanggaran pada Uji Asumsi Klasik (Semua Asumsi Terpenuhi)

4. Hasil Estimasi

```
# hasil estimasi
```

```
summary(model1)
```

```
##
```

```
## Call:
```

```
## lm(formula = y ~ x1 + x2, data = df)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -2.4484 -1.0422 -0.4126  0.9772  4.1546
```

```
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.7346537  2.2664993   0.765   0.4499
## x1          0.0002008  0.0001102   1.822   0.0781 .
## x2          0.0407785  0.0255746   1.594   0.1210
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.688 on 31 degrees of freedom
## Multiple R-squared:  0.1907, Adjusted R-squared:  0.1385
## F-statistic: 3.652 on 2 and 31 DF,  p-value: 0.03767
```

$$\beta_0 = 1.734653685$$

$$\beta_1 = 0.000200775 \quad \beta_2 = 0.040778518$$

$$R - squared = 0.1907$$

5. pemodelan

Pemodelan dengan polinom derajat 2

```
# Model Regresi Kuadrat (Polinom derajat dua)
polinom_model = lm(y~x1+I(x1^2)+x2+I(x2^2), data=df)
summary(polinom_model)

##
## Call:
## lm(formula = y ~ x1 + I(x1^2) + x2 + I(x2^2), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5863 -1.0651 -0.1261  1.0701  4.4363
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.313e+00  5.715e+00   0.230   0.8198
## x1           1.957e-03  8.925e-04   2.193   0.0365 *
## I(x1^2)      -1.087e-07  5.441e-08  -1.998   0.0552 .
## x2           7.264e-02  1.696e-01   0.428   0.6716
## I(x2^2)      -3.597e-04  1.216e-03  -0.296   0.7694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.632 on 29 degrees of freedom
## Multiple R-squared:  0.2925, Adjusted R-squared:  0.1949
## F-statistic: 2.998 on 4 and 29 DF,  p-value: 0.0347
```

Pemodelan dengan interaksi

```
# Model Regresi dengan Interaksi
linmod_Interaksi = lm(y~x1+x2+x1*x2, data = df)
summary(linmod_Interaksi)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2 + x1 * x2, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0974 -0.9409 -0.2114  0.8772  4.5854
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.3077642   2.1190801   0.617   0.5418
## x1           0.0430352   0.0179153   2.402   0.0227 *
## x2           0.0395024   0.0238321   1.658   0.1078
## x1:x2        -0.0004353   0.0001821  -2.391   0.0233 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.573 on 30 degrees of freedom
## Multiple R-squared:  0.3202, Adjusted R-squared:  0.2522
## F-statistic:  4.71 on 3 and 30 DF,  p-value: 0.00824
```

Dari 2 pemodelan yang dilakukan di dapatkan bahwa model regresi dengan interaksi memiliki R-Squared lebih tinggi.

6. Membandingkan hasil no.4 dan 5

Dari hasil kedua Model Regresi di atas, ketika data pengamatan yang diregresikan dengan Regresi dengan interaksi mendapatkan nilai kebaikan model (R-Squared) **lebih baik atau lebih tinggi** jika dibandingkan dengan Model Regresi Linier dan Model regresi polinomial 2 derajat. model yang didapatkan adalah **0,3202 atau 32,20%**. Sehingga model ini lebih baik untuk dimodelkan ke dalam **model regresi dengan interaksi**.

7. Interpretasi model terbaik dari no.6

```
linmod_Interaksi

##
## Call:
## lm(formula = y ~ x1 + x2 + x1 * x2, data = df)
##
## Coefficients:
## (Intercept)          x1          x2          x1:x2
##   1.3077642   0.0430352   0.0395024  -0.0004353
```

$$y(\text{Tingkat Pengangguran}) = 1.3077642 - 0.0430352x_1 + 0.0395024x_2 + -0.0004353x_1 : x_2$$

- Ketika seluruh variabel prediktor konstan maka Tingkat Pengangguran Terbuka bernilai 1.3077642%.
- Ketika variabel prediktor lainnya konstan, maka Kepadatan akan menambah Tingkat Pengangguran Terbuka sebesar 0,0430352 setiap Jiwa/m².
- Ketika variabel prediktor lainnya konstan, maka Proporsi Remaja dengan TIK akan akan menambah Tingkat Pengangguran Terbuka sebesar 0,0395024 setiap Persen(%).
- Ketika variabel prediktor lainnya konstan, maka Interaksi antara Kepadatan dan Proporsi Remaja dengan TIK akan mengurangi Tingkat Pengangguran Terbuka sebesar 0,0004353.