

Asumsi Regresi

Ashari

12/8/2020

Data

Data yang digunakan adalah data marketing dari packages datarium. Install terlebih dahulu jika belum tersedia

```
install.packages('datarium')
```

Load data

```
data("marketing", package = "datarium")  
head(marketing)
```

```
##  youtube facebook newspaper sales  
## 1  276.12    45.36    83.04 26.52  
## 2   53.40    47.16    54.12 12.48  
## 3   20.64    55.08    83.16 11.16  
## 4  181.80    49.56    70.20 22.20  
## 5  216.96    12.96    70.08 15.48  
## 6   10.44    58.68    90.00  8.64
```

Data diatas merupakan hasil penjualan suatu produk dan biaya yang dikeluarkan untuk promosi pada media youtube, facebook dan newspaper.

Membuat model

```
model_reg <- lm(sales ~ youtube + facebook + newspaper, data = marketing)  
model_reg
```

```
##  
## Call:  
## lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)  
##  
## Coefficients:  
## (Intercept)      youtube      facebook      newspaper  
##    3.526667    0.045765    0.188530   -0.001037
```

```
print('#####')
```

```
## [1] "#####"
```

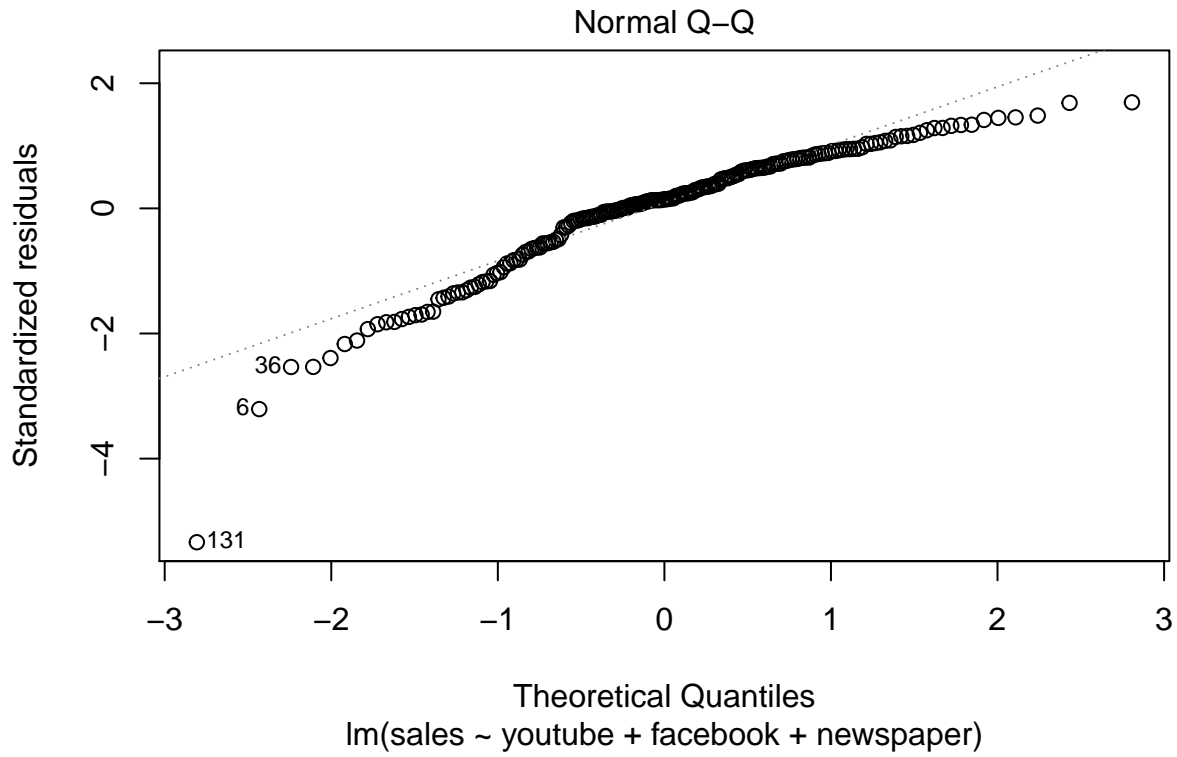
```
summary(model_reg)
```

```
##
## Call:
## lm(formula = sales ~ youtube + facebook + newspaper, data = marketing)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5932  -1.0690   0.2902   1.4272   3.3951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.526667   0.374290   9.422  <2e-16 ***
## youtube      0.045765   0.001395  32.809  <2e-16 ***
## facebook     0.188530   0.008611  21.893  <2e-16 ***
## newspaper   -0.001037   0.005871  -0.177    0.86
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.023 on 196 degrees of freedom
## Multiple R-squared:  0.8972, Adjusted R-squared:  0.8956
## F-statistic: 570.3 on 3 and 196 DF,  p-value: < 2.2e-16
```

Uji kenormalan

QQPLOT

```
plot(model_reg, 2)
```



Shapiro Wilk

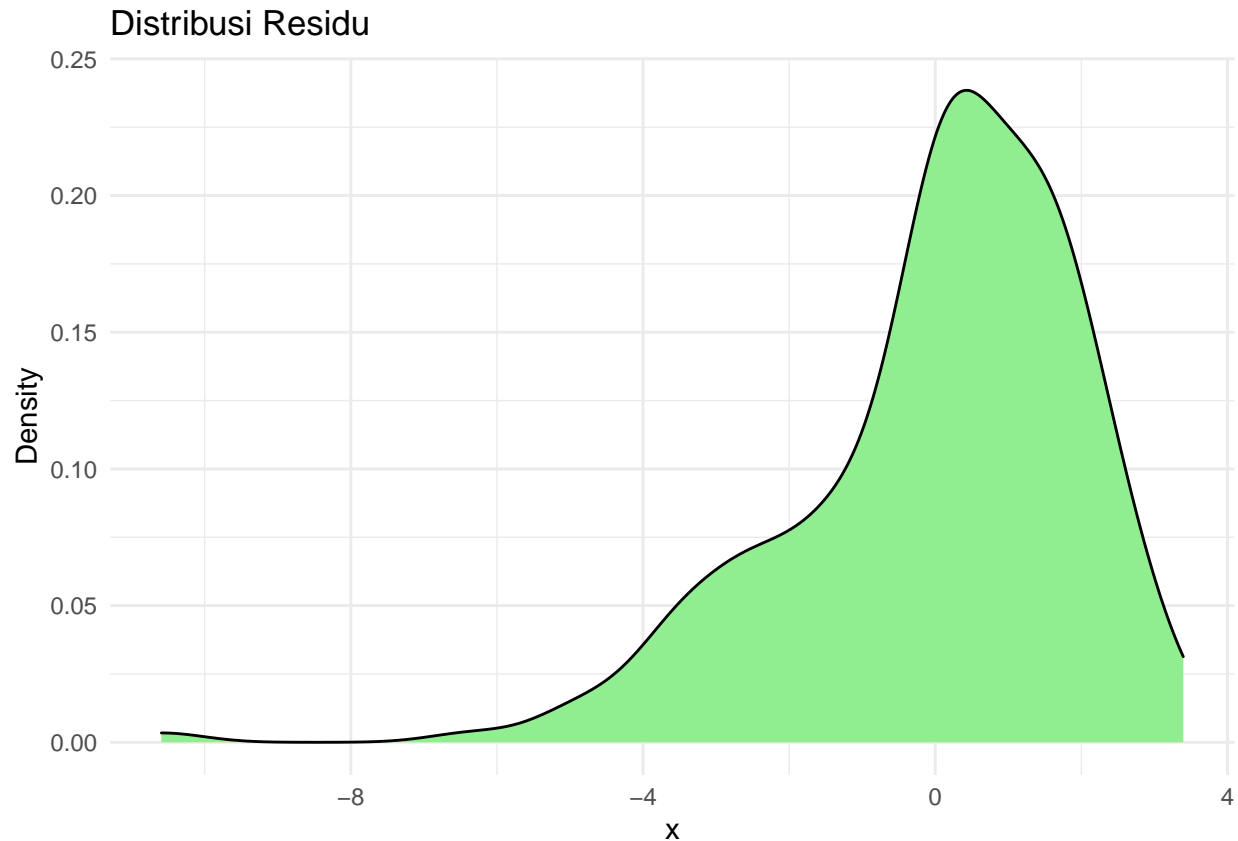
Mengambil nilai residual pada model

```
residual_model <- resid(model_reg)
head(residual_model) # 6 residual pertama
```

```
##           1           2           3           4           5           6
##  1.8912307 -2.3254258 -3.6092049  1.0826046 -0.3464062 -6.3340172
```

Plot Residu

```
library(ggplot2)
ggplot(as.data.frame(residual_model), aes(residual_model))+
  geom_density(color="black", fill="lightgreen") +
  labs(title="Distribusi Residu", x="x", y = "Density") +
  theme_minimal()
```



Uji Shapiro Wilk

Hipotesis

- H_0 = Data tidak berdistribusi normal
- H_1 = Data berdistribusi normal

Tolak H_0 jika p-value ≤ 0.05

```
shapiro.test(residual_model)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  residual_model  
## W = 0.91767, p-value = 3.939e-09
```

Dari hasil di atas diperoleh p-value $3.939e-09$ lebih kecil dari 0.05, gagal menolak H_0 , sehingga data tidak berdistribusi normal.

Multikolinieritas

Cara ini sangat mudah, hanya melihat apakah nilai VIF untuk masing-masing variabel lebih besar dari 10 atau tidak. Bila nilai VIF lebih besar dari 10 maka diindikasikan model tersebut memiliki gejala Multikolinieritas.

```
library(car)
```

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

```
vif(model_reg)
```

```
##   youtube  facebook newspaper  
## 1.004611  1.144952  1.145187
```

VIF semua variabel independen dibawah 10 berarti tidak terjadi multikolinieritas

Autokorelasi

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

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

```
##
```

```
##      as.Date, as.Date.numeric
```

```
dwtest(model_reg)
```

```
##
```

```
## Durbin-Watson test
```

```
##
```

```
## data: model_reg
```

```
## DW = 2.0836, p-value = 0.7236
```

```
## alternative hypothesis: true autocorrelation is greater than 0
```

Deteksi Autokorelasi Positif:

- Jika $dw < dL$ maka terdapat autokorelasi positif,
- Jika $dw > dU$ maka tidak terdapat autokorelasi positif,
- Jika $dL < dw < dU$ maka pengujian tidak meyakinkan atau tidak dapat disimpulkan.

Deteksi Autokorelasi Negatif:

- Jika $(4 - dw) < dL$ maka terdapat autokorelasi negatif,
- Jika $(4 - dw) > dU$ maka tidak terdapat autokorelasi negatif,
- Jika $dL < (4 - dw) < dU$ maka pengujian tidak meyakinkan atau tidak dapat disimpulkan.

Lihat tabel disini <http://lkeb.umm.ac.id/files/file/tabel-dw.pdf>

- $dw = 2.0836$
- $dl = 1.7279$
- $du = 1.8094$

Dapat dilihat bahwa $+dw > dU$ maka tidak terdapat autokorelasi positif $+(4 - dw) > dU$ maka tidak terdapat autokorelasi negatif

Sehingga dapat disimpulkan tidak terjadi autokorelasi

```
durbinWatsonTest(model_reg)
```

```
## lag Autocorrelation D-W Statistic p-value
## 1 -0.04687792 2.083648 0.508
## Alternative hypothesis: rho != 0
```

Hipotesis

- H_0 = Tidak terjadi autokorelasi
- H_1 = Terjadi autokorelasi

Tolak H_0 jika $p\text{-value} \leq 0.05$ $p\text{-value} 0.0582$ dan > 0.05 maka gagal menolak H_0 . Sehingga tidak terjadi autokorelasi.

Heterokedastisitas

Menggunakan Breusch-Pagan Test

```
library(lmtest)
bptest(model_reg)
```

```
##
## studentized Breusch-Pagan test
##
## data: model_reg
## BP = 5.1329, df = 3, p-value = 0.1623
```

- H_0 = Tidak terjadi heterokedastisitas
- H_1 = Terjadi heterokedastisitas

Tolak H_0 jika $p\text{-value} \leq 0.05$

$p\text{-value} 0.1623$ dan > 0.05 maka gagal menolak H_0 . Sehingga tidak terjadi heterokedastisitas

Referensi

<https://rpubs.com/aryn999/LinearRegressionAssumptionsAndDiagnosticsInR>