

Kemiskinan di Kabupaten/Kota Jawa Barat

Eka Pramudita

5/17/2021

Memilih Library yang Akan Digunakan

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --

## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.1      v dplyr   1.0.6
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(olsrr)
```

```
##
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##     rivers
```

Mengambil Data

```
data <- read.csv("Kemiskinan - Soal 7.csv")
apply(data, class) # check data types
```

```
##           kode_kabupaten_kota           nama_kabupaten_kota
##           "integer"           "character"
##           latitude           longitude
##           "character"           "character"
##           tahun   persentase_penduduk_miskin.persen.
##           "integer"           "character"
```

```
##           indeks_pembangunan_manusia           angka_harapan_hidup..tahun.
##           "character"                         "character"
##   angka_harapan_lama_sekolah.tahun.   pdrb_harga_konstan.miliar.rupiah.
##           "character"                         "integer"
##   pengeluaran_per_kapita.ribu.rupiah.   kepadatan_penduduk.jiwa.km2.
##           "integer"                         "integer"
##   tingkat_pengangguran_terbuka.persen.
##           "character"
```

Transformasi Data

Karena terdapat variabel numerik yang masih memiliki tipe data character, maka perlu dilakukan konversi ke variabel numerik. Selain itu, penulisan desimal yang ditulis dalam koma perlu diubah menjadi penulisan dalam titik.

```
data$persentase_penduduk_miskin.persen. <-
  as.double(gsub(',', '.', data$persentase_penduduk_miskin.persen.))

data$indeks_pembangunan_manusia <-
  as.double(gsub(',', '.', data$indeks_pembangunan_manusia))

data$angka_harapan_hidup..tahun. <-
  as.double(gsub(',', '.', data$angka_harapan_hidup..tahun.))

data$angka_harapan_lama_sekolah.tahun. <-
  as.double(gsub(',', '.', data$angka_harapan_lama_sekolah.tahun.))

data$tingkat_pengangguran_terbuka.persen. <-
  as.double(gsub(',', '.', data$tingkat_pengangguran_terbuka.persen.))
```

Data Slicing dan Scaling

Variabel yang diperlukan akan dipilih untuk dimasukkan ke model inisiasi. Proses data scaling juga dilakukan agar efek dari variabel dapat diukur melalui koefisien.

```
scaled.data <- scale(data[c(7:13)], center = TRUE, scale = TRUE)
scaled.data <- cbind(data[6], as.data.frame(scaled.data))
```

Multiple Linear Regression (Inisiasi)

Setelah data sudah bersih, kemudian dilakukan inisiasi model. Dilakukan pemodelan regresi linear berganda dengan memasukkan semua variabel di dalam scaled data.

Namun seperti yang terlihat pada summary, hanya variabel PDRB yang signifikan pada model menurut p-value, sehingga perlu dilakukan metode Stepwise Regression untuk memperoleh model dengan hasil yang optimal.

```
model <- lm(persentase_penduduk_miskin.persen. ~ indeks_pembangunan_manusia
  + angka_harapan_hidup..tahun. + angka_harapan_lama_sekolah.tahun.
  + pdrb_harga_konstan.miliar.rupiah. + pengeluaran_per_kapita.ribu.rupiah.
  + kepadatan_penduduk.jiwa.km2. + tingkat_pengangguran_terbuka.persen.,
```

```

data = scaled.data)
summary(model)

##
## Call:
## lm(formula = persentase_penduduk_miskin.persen. ~ indeks_pembangunan_manusia +
##     angka_harapan_hidup..tahun. + angka_harapan_lama_sekolah.tahun. +
##     pdrb_harga_konstan.miliar.rupiah. + pengeluaran_per_kapita.ribu.rupiah. +
##     kepadatan_penduduk.jiwa.km2. + tingkat_pengangguran_terbuka.persen.,
##     data = scaled.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0164 -0.7731 -0.2584  0.6982  4.4282
##
## Coefficients:
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   7.4067     0.3339   22.185  4.8e-15 ***
## indeks_pembangunan_manusia    -2.7297     2.1019   -1.299   0.2096
## angka_harapan_hidup..tahun.    0.5844     0.7301    0.800   0.4333
## angka_harapan_lama_sekolah.tahun. 0.4716     0.8970    0.526   0.6051
## pdrb_harga_konstan.miliar.rupiah. -0.8582     0.4171   -2.058   0.0536 .
## pengeluaran_per_kapita.ribu.rupiah. 0.4920     1.0920    0.451   0.6574
## kepadatan_penduduk.jiwa.km2.    -0.3001     0.9137   -0.329   0.7461
## tingkat_pengangguran_terbuka.persen. 0.5420     0.4014    1.350   0.1929
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.735 on 19 degrees of freedom
## Multiple R-squared:  0.6581, Adjusted R-squared:  0.5322
## F-statistic: 5.225 on 7 and 19 DF,  p-value: 0.001888

```

Stepwise Regression

Metode stepwise regression diimplementasikan menggunakan `ols_step_both_p` dari package `olsrr`. Model akan dipilih menurut p-value karena tujuan dari pemodelan ini adalah untuk mengetahui faktor mana saja yang memiliki pengaruh, bukan untuk melakukan prediksi dengan dataset yang lain.

pent dalam code menyatakan nilai maksimum p-value dari variabel yang akan dimasukkan ke dalam model, sedangkan prem menyatakan nilai minimum p-value dari variabel yang akan dihapus dari model. Untuk pemodelan ini diambil nilai 0.1 sebagai pent dan 0.3 sebagai prem.

```
ols_step_both_p(model, pent = 0.1, prem = 0.3, details = FALSE)
```

```

##
##                               Stepwise Selection Summary
## -----
##                               Added/          Adj.
## Step          Variable      Removed    R-Square  R-Square    C(p)    AIC
## -----
## 1      indeks_pembangunan_manusia    addition    0.553    0.535    1.8430   110.1238
## 2      pdrb_harga_konstan.miliar.rupiah.    addition    0.605    0.572    0.9740   108.8105
## -----

```

Seperti yang ditunjukkan pada summary bahwa variabel yang signifikan adalah Indeks Pembangunan Manusia dan PDRB Harga Konstan, maka model yang akan dibentuk hanya akan membawa dua variabel tersebut.

```
model1 <- lm(persentase_penduduk_miskin.persen. ~ indeks_pembangunan_manusia + pdrb_harga_konstan.miliar.rupiah, data = scaled.data)
summary(model1)
```

```
##
## Call:
## lm(formula = persentase_penduduk_miskin.persen. ~ indeks_pembangunan_manusia +
##     pdrb_harga_konstan.miliar.rupiah., data = scaled.data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.9849 -0.8361 -0.2303  0.8834  4.2403
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.4067     0.3195  23.185 < 2e-16 ***
## indeks_pembangunan_manusia    -1.7074     0.3408  -5.009 4.06e-05 ***
## pdrb_harga_konstan.miliar.rupiah. -0.6033     0.3408  -1.770  0.0894 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.66 on 24 degrees of freedom
## Multiple R-squared:  0.6046, Adjusted R-squared:  0.5716
## F-statistic: 18.35 on 2 and 24 DF,  p-value: 1.461e-05
```

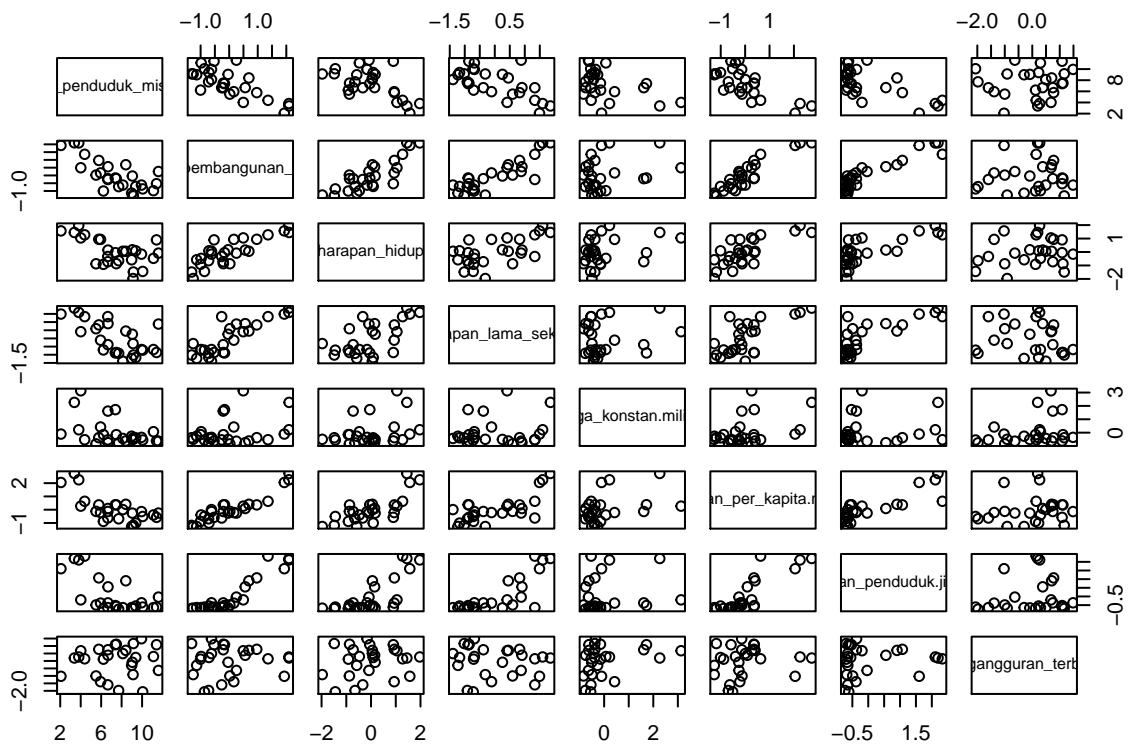
```
AIC(model1)
```

```
## [1] 108.8105
```

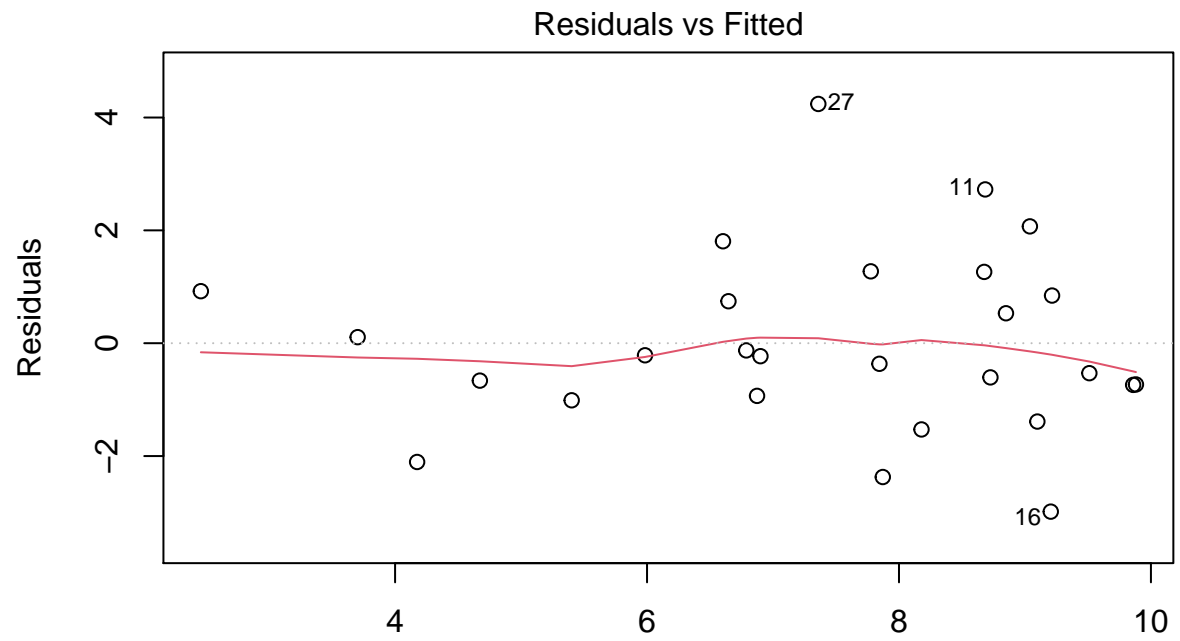
p-value dari model yang telah dibuat memiliki nilai di bawah 0.1, sesuai dengan nilai pent yang ditentukan sebelumnya.

Pengecekan asumsi

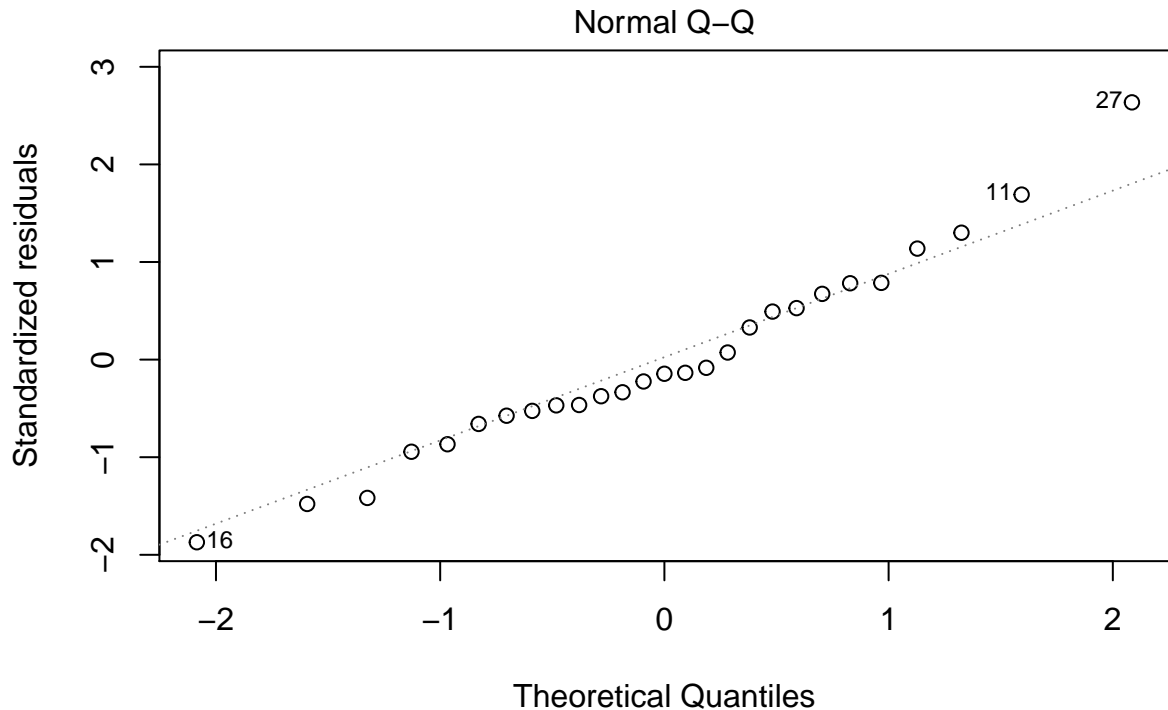
```
# Linearity
plot(scaled.data)
```



```
# Multivariate Normal and Homoscedasticity
plot(model1, which=1:2)
```



Fitted values
 $\text{lm}(\text{persentase_penduduk_miskin.persen.} \sim \text{indeks_pembangunan_manusia} + \text{pdrb_h})$



lm(persentase_penduduk_miskin.persen. ~ indeks_pembangunan_manusia + pdrb_h

```
# Multicollinearity
car::vif(model1)
```

```
##           indeks_pembangunan_manusia pdrb_harga_konstan.miliar.rupiah.
##                               1.096113                               1.096113
```

- Asumsi Linearitas Seperti yang ditunjukkan pada plot, dapat terlihat hubungan linear dari tiap pasang variabel di data.
- Homoskedastisitas Seperti yang terlihat di grafik Residual vs Fitted, pola scatterplot cenderung tidak terbentuk sehingga dapat disimpulkan bahwa asumsi homoskedastisitas terpenuhi.
- Multivariat Normal Seperti yang terlihat di Normal Q-Q plot, standardized residuals mengikuti normal plot, maka asumsi kenormalan terpenuhi.
- Tidak ada multikolinearitas Nilai VIF dari kedua variabel adalah 1.09, maka dapat disimpulkan bahwa tidak ada multikolinearitas dalam model karena nilai VIF di bawah 10.

Koefisien

Koefisien model menunjukkan seberapa besar pengaruh variabel terhadap model.

```
model1$coefficients
```

```
##           (Intercept)           indeks_pembangunan_manusia
##           7.4066667           -1.7073865
## pdrb_harga_konstan.miliar.rupiah.
##           -0.6033423
```

Jika diambil nilai mutlak dari koefisien tersebut, maka Indeks Pembangunan Manusia memberikan efek yang paling besar dalam model. Semua variabel memiliki koefisien negatif, artinya setiap penambahan nilai pada variabel- variabel tersebut akan menurunkan Persentase Penduduk Miskin.

Kesimpulan dan Rekomendasi

Dari penjelasan pada bagian Koefisien, dapat disimpulkan bahwa:

1. Peningkatan Indeks Pembangunan Manusia (IPM) harus diperhatikan dengan serius oleh pemerintah. Program-program pemerintah provinsi harus fokus pada menambah faktor- faktor yang ikut mendorong (IPM) yaitu angka harapan hidup, literasi, pendidikan dan standar hidup. Peningkatan IPM akan secara efektif menurunkan Persentase Penduduk Miskin di Kabupaten/Kota di Jawa Barat.
2. PDRB atau Produk Domestik Regional Bruto dapat dilihat sebagai kegiatan ekonomi. Program pemerintah provinsi harus berfokus pada peningkatan aktivitas ekonomi dengan mengelola sumber daya alam dengan baik, meningkatkan kualitas manusia, mengembangkan inovasi kewirausahaan, serta mengembangkan teknologi.