TUGAS MATA KULIAH ANALISIS REGRESI
DEPARTEMEN STATISTIKA
IPB UNIVERSITY

Nama : Alfidhia Rahman Nasa Juhanda

NRP : G1401201004

Judul : Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Barat dengan Menggunakan Regresi Berganda

---

ANALISIS STATISTIKA FAKTOR YANG MEMPENGARUHI INDEKS PEMBANGUNAN MANUSIA DI KABUPATEN/KOTA PROVINSI JAWA BARAT DENGAN MENGGUNAKAN REGRESI BERGANDA

Alfidhia Rahman Nasa Juhanda[1]

[1]Departement of Statistics, IPB University, Indonesia
[‡]corresponding author: alfidhiarahman@apps.ipb.ac.id

1. Pendahuluan

1.1 Latar Belakang

Sumber daya manusia memiliki peranan penting dalam menciptakan lingkungan yang masyarakat sehat serta dapat meningkatkan produktivitas masyarakat. Dalam arti sederhana, pembangunan dapat dimaknai sebagai proses atau usaha untuk melakukan perubahan ke arah yang lebih baik (Sangkareng *et al.* 2019). Pembangunan manusia merupakan sebuah paradigma pembangunan yang menempatkan manusia sebagai objek utama dan sasaran akhir dari seluruh kegiatan pembangunan. Indeks Pembangunan Manusia (IPM) adalah indikator untuk mengukur pembangunan kualitas hidup manusia suatu daerah atau negara. Pada tahun 1990, UNDP menetapkan tiga dimensi pembentuk IPM yang tidak mengalami perubahan hingga saat ini. Ketiga dimensi ini merupakan pendekatan yang dipilih dalam penggambaran kualitas hidup manusia. Dimensi tersebut mencakup: 1. umur panjang dan hidup sehat (*a long and healthy life*); 2. pengetahuan (*knowledge*); dan 3. standar hidup layak (*decent standard of living*). (BPS 2020)

Pandemi COVID-19 telah memberikan dampak yang luas terhadap berbagai aspek dalam kehidupan masyarakat. Seluruh indikator ekonomi dan sosial mengalami tekanan yang berat, tidak terkecuali IPM. Pertumbuhan IPM Indonesia pada tahun 2020 mengalami perlambatan yang cukup berarti dengan hanya tumbuh sebesar 0,03 persen, jauh melambat dibandingkan pertumbuhan tahun sebelumnya yang mencapai 0,74 persen. Pada tahun 2020 terdapat 24 provinsi yang mengalami peningkatan IPM, sedangkan 10 provinsi lainnya mengalami penurunan. Jawa Barat adalah salah satu provinsi yang mengalami kenaikan pada tahun 2020. IPM Jawa Barat pada tahun 2020 sebesar 72,09 persen, naik 0,06 persen jika dibandingkan tahun 2019 yaitu 72,03 persen. Meskipun masih naik, namun angka ini relatif sangat kecil jika dibandingkan kenaikan tahun sebelumnya. Selama tahun 2012 sampai dengan 2020 IPM provinsi Jawa Barat cenderung mengalami peningkatan, berturut-turut nilainya adalah 67,32; 68,25; 68,8; 69,5; 70,05; 70,69; 71,3; 72,03; 72,09. Dapat terlihat bahwa angka kenaikan di tahun 2020 merupakan angka kenaikan terkecil sepanjang delapan tahun terakhir. Pada penelitian ini akan dicari faktor yang dapat mempengaruhi IPM sehingga nantinya dapat memberikan masukan kepada pemerintah Jawa Barat sektor apa saja yang harus ditingkatkan agar IPM dapat terus meningkat. Mengingat banyak sekali faktor yang dapat mempengaruhi nilai IPM, oleh karena itu pada penelitian ini digunakan metode analisis regresi berganda.

Analisis regresi merupakan suatu kajian dari hubungan antara satu peubah respon, dengan satu atau lebih peubah lain. Regresi linier berganda merupakan pengembangan dari regresi linier sederhana, yaitu metode yang digunakan untuk mengetahui pengaruh satu atau lebih peubah bebas terhadap satu variabel terikat (Siregar 2013). Metode yang digunakan dalam analisis regresi linier adalah Metode Kuadrat Terkecil (MKT). MKT akan menghasilkan nilai dugaan parameter yang baik jika memenuhi syarat-syarat tertentu, yaitu harus bersifat BLUE (*Best Linier Unbiased Estimator*). Sifat-sifat BLUE tersebut adalah penduga tak bias dengan ragam sisaan minimum. Beberapa asumsi yang harus dipenuhi untuk menggunakan metode regresi linier berganda adalah tidak terjadi multikolinearitas, homoskedastisitas, error yang berdistribusi normal, serta tidak adanya autokorelasi. Dalam penelitian ini, analisis regresi linier berganda dilakukan untuk mengkaji pengaruh dari beberapa peubah bebas terhadap variabel terikat Indeks Pembangunan Manusia (IPM). Analisis tersebut dilakukan menggunakan bantuan perangkat lunak R dan R-studio.

1.2 Tujuan Penelitian

Penelitian ini bertujuan mencari model regresi linier terbaik dalam menggambarkan nilai Indeks Pembangunan Manusia (IPM) di Jawa Barat dengan menentukan peubah-peubah tertentu sehingga pemerintah dapat tahu serta lebih memprioritaskan sektor yang paling berpengaruh terhadap IPM.

2. Tinjauan Pustaka

2.1 Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. IPM diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 dan dipublikasikan secara berkala dalam laporan tahunan Human Development Report (HDR). IPM dibentuk oleh 3 (tiga) dimensi dasar: (1) Umur panjang dan hidup sehat; (2) Pengetahuan; dan (3) Standar hidup layak.

2.2 Analisis Regresi

Analisis regresi adalah sebuah teknik statistik untuk membuat model dan melihat hubungan antara dua variabel atau lebih. Analisis yang memiliki variabel bebas lebih dari satu disebut analisis regresi linier berganda. Teknik regresi linier berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih variabel bebas (X1, X2, X3,...., Xk) terhadap variabel terikat (Y). Model regresi linier berganda untuk populasi dapat ditunjukkan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + e$$

Model regresi linier berganda untuk populasi di atas dapat ditaksir dengan model regresi linier berganda untuk sampel, yaitu:

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_k X_k$$

dengan :

$\hat{Y}$	: Nilai penduga bagi variabel Y
b_0	: Dugaan bagi parameter konstanta
$b_1, b_2, \ldots, b_k$	: Dugaan bagi parameter konstanta $\beta_1, \beta_2, \ldots, \beta_n$
X	: Variabel bebas.

2.3 Uji Asumsi

2.3.1 Normalitas

Tujuan uji normalitas adalah untuk menguji apakah dalam sebuah model regresi, variabel terikat dan variabel bebas atau keduanya mempunyai distribusi normal ataukah tidak. Model regresi yang baik adalah distribusi data normal atau mendekati normal. Deteksi normalitas dilakukan dengan melihat grafik Normal Probability Plot ataupun uji Kolmogorov-smirnov.

2.3.2 Heteroskedastisitas

Uji heteroskedastisitas dilakukan untuk menguji apakah dalam sebuah model regresi terjadi ketidaksamaan ragam sisaan dari satu pengamatan ke pengamatan yang lain tetap, maka disebut heteroskedastisitas. Uji ini menggunakan uji Glejser.

2.3.3 Kebebasan Sisaan

Uji Kebebasan sisaan dilakukan untuk menguji apakah dalam sebuah model regresi terjadi tidak bebasan sisaan nilai residual atau tidak. Uji ini menggunakan uji Runs.

2.3.4 Multikolinearitas

Uji multikolinearitas digunakan untuk melihat ada atau tidaknya korelasi yang tinggi antara variabel-variabel bebas dalam suatu model regresi linear berganda. Jika ada korelasi yang tinggi antara peubah bebasnya, maka hubungan antara variabel bebas terhadap variabel terikatnya menjadi terganggu. Uji multikolinearitas dapat dilihat dari nilai Tolerance atau VIF (Variance Inflantion Factor) serta besaran korelasi antar peubah bebas. Suatu model regresi dikatakan dapat dikatakan bebas multikolinearitas jika mempunyai nilai VIF tidak lebih dari 10 atau mempunyai angka tolerance tidak kurang dari 0,1.

2.4 Pendeteksian Pencilan, Leverage, dan Amatan Berpengaruh

2.4.1 Pencilan

Pencilan adalah data amatan yang terpaut jauh dari data amatan lainnya (Maddala 1969). Kemunculan dari pencilan dapat menunjukkan kejadian yang ganjil atau kesalahan dalam transkripsi data. Keberadaan pencilan dapat berpengaruh terhadap nilai-nilai statistic ringkasan seperti R^2, bisa menyebabkan pemodelan yang tidak tepat, dan dapat mempengaruhi efisiensi dalam model. Pencilan terdapat pada arah sumbu y yang tidak mengikuti tren.

2.4.2 Leverage

Leverage adalah amatan yang memiliki nilai x yang berbeda dari nilai x lainnya. Titik ini tidak sepenuhnya berpengaruh pada estimasi koefisien regresi karena tidak semua titik leverage tidak mengikuti tren, tetapi akan masih berdampak signifikan pada statistik ringkasan model seperti R^2.

2.4.3 Amatan Berpengaruh

Amatan Berpengaruh adalah amatan yang tidak bernilai wajar pada nilai X atau Y. Amatan berpengaruh berkaitan dengan besarnya perubahan yang terjadi pada dugaan parameter regresi, R^2, uji hipotesis, jika amatan tersebut disisihkan. Titik leverage dan pencilan merupakan titik amatan yang berpotensi untuk menjadi amatan berpengaruh.

2.5 Pendugaan Model Terbaik

2.5.1 *Akaike Information Criterion* (AIC)

Akaike information criterion (AIC) merupakan penduga kesalahan prediksi dan penilaian yang diaplikasikan untuk menentukan model terbaik (Taddy 2019). Pada pemilihan model regresi terbaik, AIC digunakan untuk peramalan yang menjelaskan kecocokan model dengan data yang ada dan nilai yang akan terjadi di masa yang akan data. Syarat utama model dikatakan baik adalah ketika AIC yang didapatkan adalah nilai AIC yang terkecil.

2.5.2 Metode *Forward*

Metode *forward* adalah metode pendugaan model terbaik di mana peubah bebas dimasukkan satu per satu diurutkan berdasarkan nilai mutlak koefisien korelasi yang terbesar terhadap peubah penjelas, dan berhenti bila semua syarat telah terpenuhi (Samsosir *et al.* 2014). Pada metode ini, model awalnya adalah model tanpa peubah penjelas. Nilai AIC juga dapat digunakan dalam metode ini.

2.5.3 Metode *Backward*

Metode *backward* adalah metode pendugaan model terbaik di mana peubah dieliminasi satu per satu berdasarkan nilai mutlak koefisien korelasi yang terkecil terhadap peubah penjelas, dan berhenti bila semua syarat telah terpenuhi (Samsosir *et al.* 2014). Pada metode ini, model awalnya adalah model dengan semua peubah penjelas. Nilai AIC juga dapat digunakan dalam metode ini.

2.5.4 Metode *Stepwise*

Metode *stepwise* adalah metode pemilihan peubah penjelas dengan langkah memasukkan satu per satu peubah bebas berdasarkan koefisien korelasi parsial yang terbesar terhadap peubah respons (Pujilestari *et al.* 2014). Pada metode ini, digunakan *Akaike Information Criterion* (AIC) untuk mengetahui persamaan mana yang terbaik.

2.5.5 *Best Subset Regression*

Prosedur *best subset regression* atau dikenal juga sebagai *all possible subset regression* dapat menghasilkan model dari semua kemungkinan fitur atau kombinasi peubah penjelas. Model terbaik dipilih berdasarkan penilaian atau beberapa kriteria statistik. Model dengan k peubah memiliki 2^k kemungkinan kombinasi peubah yang dipasangkan. Regresi subset menggunakan fungsi *regsubsets* dari *package leaps* di R.

2.5.6 *Adjusted R-square*

Model terbaik adalah model yang memiliki nilai *adjusted r-squared* yang paling tinggi. Nilai kuadrat tengah sisaan yang kecil akan menaikan nilai adjusted r-squared. Nilai dari a*djusted R-square* dinilai lebih efektif daripada nilai *R-square* pada model dengan lebih dari satu peubah karena kecenderungan *R-square* yang akan terus menaik jika peubah bertambah, sehingga hasilnya akan berbias. *Adjusted R-square* hadir untuk mengatasi hal tersebut.

2.5.7 *Cp Mallow's*

Cp Mallow's digunakan untuk menilai kecocokan model regresi. Kita dapat mengidentifikasi model regresi "terbaik" dengan mengidentifikasi model dengan nilai Cp terendah yang mendekati k+1, di mana k adalah jumlah peubah dalam model. Jika dalam model tersedia nilai Cp yang rendah, pilihlah nilai yang paling kecil di antaranya.

2.5.8 *PRESS*

Predicted Residual Error Sum of Squares (PRESS) adalah bentuk validasi silang yang digunakan dalam analisis regresi untuk memberikan ukuran kecocokan model dengan sampel pengamatan yang tidak digunakan untuk memperkirakan model itu sendiri (Tarpey 2000). Nilai ini dihitung dengan menkuadratkan sisaan prediksi model.

3. Metodologi

3.1 Data

Penelitian ini menggunakan data sekunder yang diperoleh dari hasil publikasi Kementerian Pendidikan dan Kebudayaan, Badan Pusat Statistik Jawa Barat dan Dinas Kesehatan Provinsi Jawa Barat. Data terdiri dari sepuluh peubah penjelas dan satu peubah respons dari masing-masing kabupaten/kota di Jawa Barat pada tahun 2020. Pemilihan peubah pada penelitian berdasarkan penelitian terdahulu yang telah dilakukan. Keterangan peubah-peubah yang digunakan tercantum pada Tabel 1.

Tabel 1 Daftar peubah yang digunakan

Kode	Peubah	Satuan
Y	Indeks Pembangunan Manusia	Poin
X1	Kepadatan Penduduk	Jiwa/km^2
X2	Angka Harapan Hidup	Tahun
X3	PDRB per kapita	Juta rupiah/orang/tahun
X4	Rata-rata lama Sekolah	Tahun
X5	Rasio Penduduk dan Dokter	Penduduk/dokter
X6	Persentase Jumlah Penduduk Miskin	Persen
X7	Jumlah Rumah Sakit	Unit
X8	Rasio Siswa dan Guru	Siswa/guru
X9	Anggaran Pemerintah untuk Pendidikan	Milyar Rupiah
X10	Tingkat Partisipasi Angkatan Kerja (TPAK)	Persen

3.2 Metode Penelitian

Prosedur analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Melakukan eksplorasi data pada peubah Indeks Pembangunan Manusia dengan menggunakan Histogram, Boxplot, dan Statistik lima serangkai. Serta membuat matriks korelasi antar peubah bebas dan peubah respons. Pada tahap ini akan dilakukan reduksi peubah pada peubah yang memiliki korelasi yang sangat rendah dengan respons.
2. Menduga model regresi linier berganda untuk semua peubah yang memiliki korelasi dengan peubah respons.
3. Menguji adanya Multikolinearitas antar peubah bebas dengan menggunakan nilai VIF.
4. Melakukan uji asumsi sisaan berupa uji Normalitas, heteroskedastisitas, dan kebebasan sisaan.
5. Mendeteksi pencilan, leverage, dan amatan berpengaruh.
6. Menduga model terbaik dengan menggunakan metode *Backward, Forward, Stepwise,* serta metode *best subset* yang memanfaatkan perbandingan nilai *Adjusted R-square* dan *Cp Mallow's*.
7. Menentukan model terbaik dengan prosedur *PRESS*.
8. Menguji asumsi sisaan kembali terhadap model terbaik yang telah dipilih.
9. Menginterpretasikan model terbaik.

4. Hasil dan Pembahasan

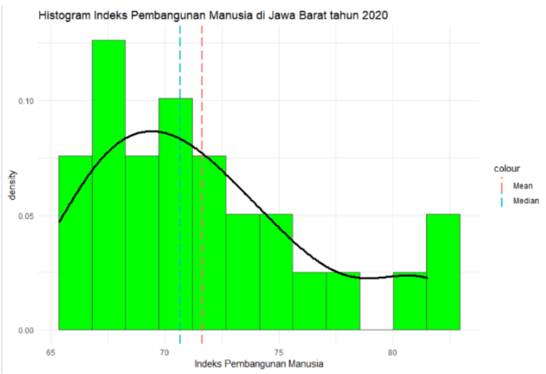
4.1 Eksplorasi Data

Data respons yang digunakan dalam penelitian ini adalah Indeks Pembangunan Manusia di Provinsi Jawa Barat pada tahun 2020. Data ini digunakan untuk mendapatkan model terbaik dengan regresi linier berganda. Tabel berikut menunjukkan ringkasan dari peubah respons:

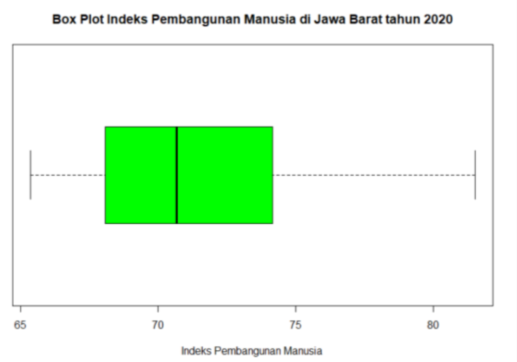
Tabel 2 Deskriptif Statistik Indeks Pembangunan Manusia

Minimum	65,36 (Kab. Cianjur)
Maksimum	81,51 (Kota Bandung)
1st Quartil	68,07
Median	70,66
Rata-rata	71,64
3rd Quartil	74,14
Ragam	22,3
Total Kota/Kabupaten	27

Dari tabel di atas kita tahu bahwa nilai IPM Jabar berentang antara 65.36 hingga 81.51. Dengan nilai terendah di kabupaten Cianjur, dan nilai tertinggi di kota Bandung. Selain itu, dapat diketahui juga nilai Quartil, rata-rata, ragam, dan total amatan.



Gambar 1 Histogram IPM Jawa Barat 2020

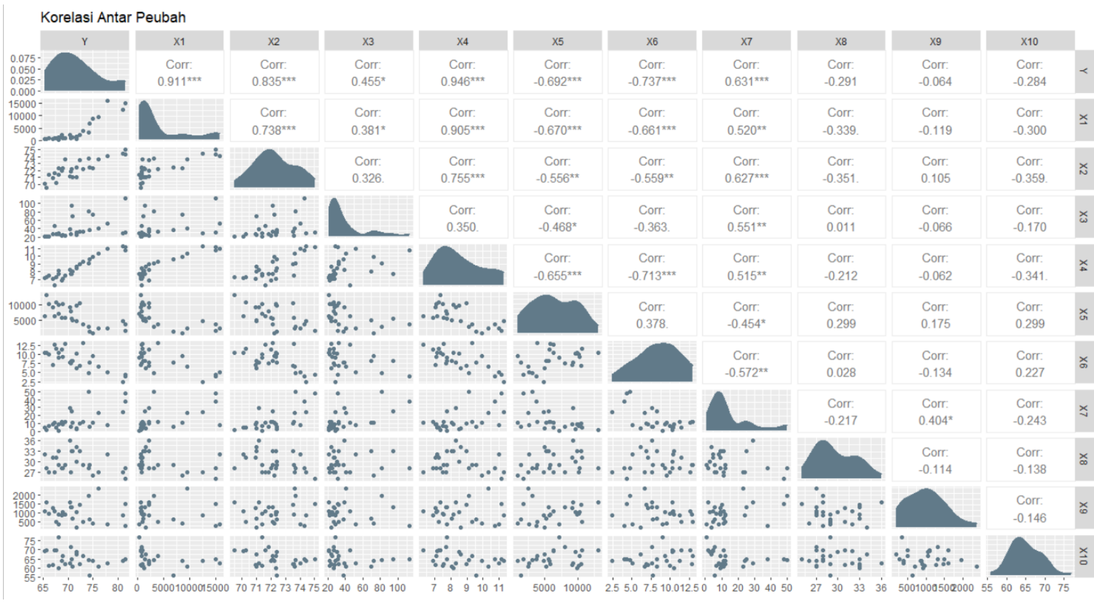


Gambar 2 Box Plot IPM Jawa Barat 2020

Dari gambar 1, dapat terlihat sebaran data Indeks Pembangunan Manusia di Jawa Barat pada Tahun 2020. Sebaran datanya lebih banyak menyebar di sebelah kiri yang berarti lebih banyak daerah di Jawa Barat dengan IPM rendah daripada IPM tinggi. Kemudian dari gambar 2, dapat terlihat bahwa tidak terdapat pencilan di sebaran data IPM.

- Matriks Korelasi

  Hipotesis :
  H_0 : Tidak ada korelasi antar dua peubah
  H_1 : Ada korelasi antar dua peubah



Gambar 3 Matriks korelasi antar peubah penuh

Dalam matriks korelasi di atas, terdapat tanda bintang (*) yang menandakan bahwa H_0 ditolak pada taraf nyata 5%. Terlihat bahwa peubah X8, X9, dan X10 memiliki korelasi yang tidak signifikan terhadap peubah respons. Maka ketiga peubah ini dapat terwakili oleh peubah lain.

4.2 Pendugaan Model Regresi

Dengan pereduksian peubah di atas, maka dugaan model regresi yang didapat adalah :

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + \beta_5 X5 + \beta_6 X6 + \beta_7 X7 + e$$

4.3 Pendeteksian Multikolinearitas

$VIF > 10$: Ada Multikolinearitas

$VIF \leq 10$: Tidak ada Multikolinearitas

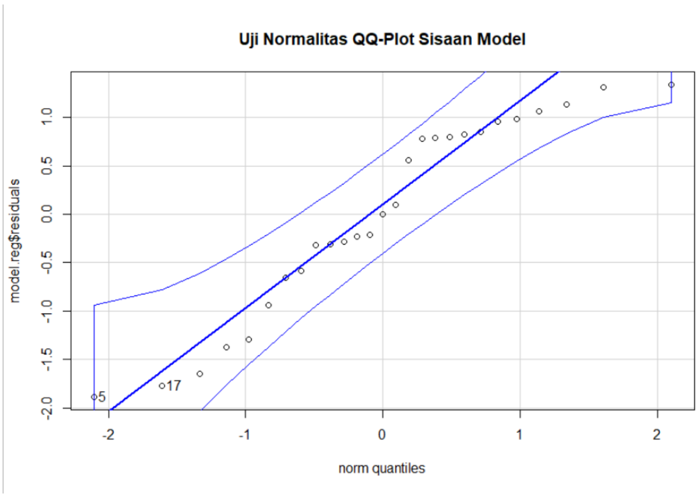
Tabel 3 Tabel Pendeteksian Multikolinearitas

Kode	Peubah	VIF
X1	Kepadatan Penduduk	6,093412
X2	Angka Harapan Hidup	2,993927
X3	PDRB per kapita	1,631268
X4	Rata-rata lama Sekolah	7,327609
X5	Rasio Penduduk dan Dokter	2,223493
X6	Persentase Jumlah Penduduk Miskin	2,514569
X7	Jumlah Rumah Sakit	2,366373

Tabel di atas menunjukan bahwa semua peubah tidak memiliki VIF yang lebih besar dari 10. Maka dari itu, dapat disimpulkan bahwa ketujuh peubah di atas **tidak memiliki multikolinearias antar peubah**.

4.4 Pengujian Asumsi Sisaan

4.4.1 Normalitas



Gambar 4 Q-Q Plot Sisaan Model

Dari Plot di atas, dapat terlihat bahwa titik titik sebaran berada di sekitar garis QQ-plot yang mengindikasikan sebaran **sisaan relatif normal**.

- Uji Formal

  Uji formal ini akan dilakukan berdasarkan taraf nyata 5% dengan uji kolmogorov-smirnov, dengan hipotesis sebagai berikut :

  H_0 : Sisaan menyebar Normal

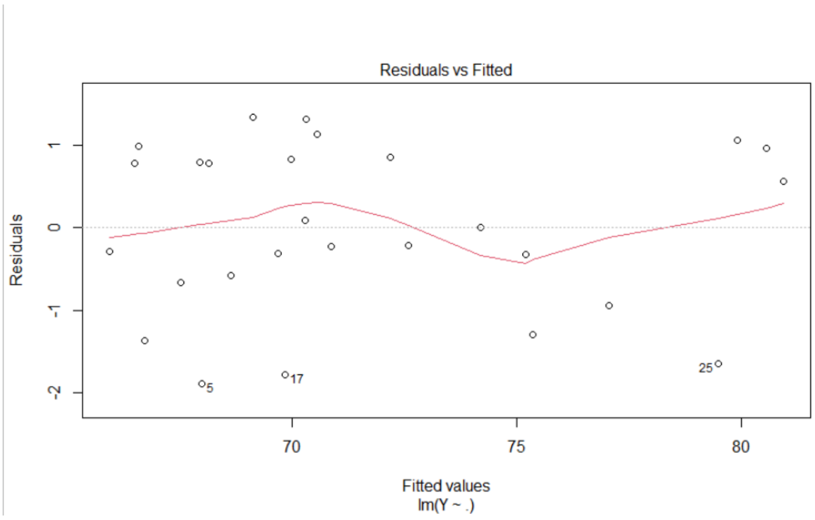
  H_1 : Sisaan tidak menyebar Normal

Tabel 4 Uji Normalitas - Kolmogorov Smirnov

Data	***P-Value***	**Keputusan**
Sisaan Model	0,2557	Sisaan menyebar normal

Karena *P-Value* (0,2257) lebih besar dari taraf nyata (0,05), maka H_0 tidak ditolak atau sisaan menyebar normal.

4.4.2 Heteroskedastisitas



Gambar 5 Plot uji kehomogenan ragam

Dari Plot di atas, dapat terlihat bahwa sebaran sisaan masih berada di sekitar nilai 0, kemudian tidak ada pola di sebaran tersebut. Jadi, dapat dikatakan bahwa **sisaannya relatif homogen**.

- Uji Formal

  Uji formal ini akan dilakukan berdasarkan taraf nyata 5% dengan uji glejser, dengan hipotesis sebagai berikut :

  H_0 : Sisaan homogen

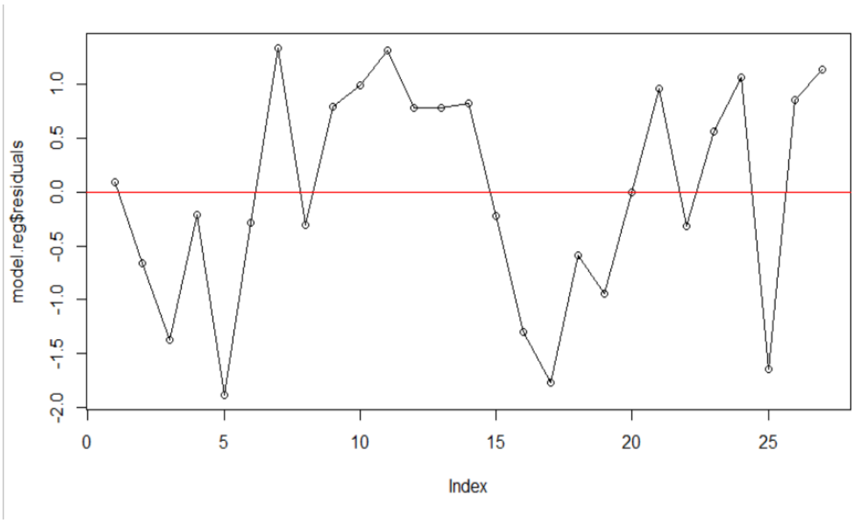
  H_1 : Sisaan tidak homogen

Tabel 5 Uji Heteroskedastisitas - glejser

Data	*P-Value*	Keputusan
Sisaan Model	0,993	Sisaan homogen

Karena *P-Value* (0,993) lebih besar dari taraf nyata (0,05), maka H_0 tidak ditolak atau sisaan homogen.

4.4.3 Kebebasan Sisaan



Gambar 6 Plot uji kebebasan sisaan

Dari Plot di atas, dapat terlihat bahwa sebaran sisaan masih berada di sekitar nilai 0, kemudian tidak ada pola di sebaran tersebut. Jadi, dapat dikatakan bahwa **sisaannya relatif saling bebas**.

- Uji Formal

  Uji formal ini akan dilakukan berdasarkan taraf nyata 5% dengan uji runs, dengan hipotesis sebagai berikut :

  H_0 : Sisaan saling bebas

  H_1 : Sisaan tidak saling bebas

Tabel 6 Uji kebebasan sisaan - runs

Data	*P-Value*	Keputusan
Sisaan Model	0,2377	Sisaan saling bebas

Karena *P-Value* (0,2377) lebih besar dari taraf nyata (0,05), maka H_0 tidak ditolak atau sisaan saling bebas.

4.5 Pendeteksian Leverage, Pencilan, dan Amatan berpengaruh

4.5.1 Mendeteksi Leverage

Rumus yang digunakan untuk mendeteksi leverage adalah sebagai berikut:

$$H = X(X'X)^{-1}X'$$

h_{ii} : unsur diagonal dari matriks H
p : banyaknya parameter
n : banyaknya amatan

Jika nilai $h_{ii} > 2\,p/n$, amatan tersebut dapat dikatakan sebagai titik *leverage*.

Dengan $n = 27;\ p = 8,$ maka $2\frac{p}{n} = \frac{2(8)}{27} = 0{,}5925.$ Maka, jika nilai $h_{ii} >$ 0.5925, amatan tersebut dapat dikatakan sebagai titik leverage.

Tabel 7 Nilai hii

obs	hii
1	0,299
2	0,219
3	0,2105
4	0,3625
5	0,0924
6	0,403
7	0,1349
8	0,3748
9	0,2657
10	0,208
11	0,2489
12	0,3268
13	0,192
14	0,2107
15	0,3214
16	0,5306
17	0,1561
18	0,1136
19	0,1267
20	0,2296
21	0,5532
22	0,3209
23	0,5434
24	0,3146
25	0,4462
26	0,4562
27	0,3392

Dari semua nilai h_{ii} yang didapat pada tabel di samping, terlihat bahwa tidak ada nilai h_{ii} yang lebih dari 0.5925. Maka dapat disimpulkan bahwa **tidak ada titik leverage** pada model regresi penuh antara Y dan X1 hingga X7.

4.5.2 Mendeteksi Pencilan

Rumus yang digunakan untuk mendeteksi pencilan adalah sebagai berikut:

$$r_i = \frac{e_i}{s\sqrt{1 - h_{ii}}};\ H = X(X'X)^{-1}X'$$

e_i : sisaan ke-i
r_i : sisaan terbakukan ke-i
s : Standar error sisaan
h_{ii} : unsur diagonal dari matriks H

Jika nilai $|r_i| > 2$ atau 3, amatan tersebut dapat dikatakan sebagai pencilan.

Tabel 8 nilai ei, hii, dan ri

obs	ei	hii	ri
1	0,095	0,299	0,096
2	-0,66	0,219	-0,633
3	-1,37	0,2105	-1,306
4	-0,214	0,3625	-0,227
5	-1,888	0,0924	-1,68
6	-0,282	0,403	-0,309
7	1,341	0,1349	1,222
8	-0,306	0,3748	-0,328
9	0,796	0,2657	0,787
10	0,987	0,208	0,94
11	1,317	0,2489	1,288
12	0,784	0,3268	0,81
13	0,787	0,192	0,742
14	0,825	0,2107	0,787
15	-0,227	0,3214	-0,233
16	-1,296	0,5306	-1,603
17	-1,774	0,1561	-1,636
18	-0,584	0,1136	-0,525
19	-0,94	0,1267	-0,852
20	0,001	0,2296	0,001
21	0,958	0,5532	1,214
22	-0,317	0,3209	-0,326
23	0,562	0,5434	0,705
24	1,064	0,3146	1,089
25	-1,647	0,4462	-1,875
26	0,854	0,4562	0,982
27	1,135	0,3392	1,183

Dari semua nilai r_i yang didapat pada tabel di samping, terlihat bahwa tidak ada nilai mutlak r_i yang lebih dari 2 atau 3. Maka dapat disimpulkan bahwa **tidak ada pencilan** pada model regresi penuh antara Y dan X1 hingga X7.

4.5.3 Mendeteksi Amatan berpengaruh

Rumus yang digunakan untuk mendeteksi amatan berpengaruh adalah sebagai berikut:

$$D_i = \frac{r_i^2}{p} \frac{h_{ii}}{1 - h_{ii}} \; ; \; H = X(X'X)^{-1}X'$$

D_i : Nilai Jarak Cook ke-i
r_i : sisaan terbakukan ke-i
h_{ii} : unsur diagonal dari matriks H
p : banyaknya parameter

Jika nilai $D_i > F_{(p,n-p;1-\alpha)}$, maka amatan tersebut dapat dikatakan sebagai amatan berpengaruh

Dengan $n = 27$; $p = 8$; $F_{(8,19;0.95)} = 2.4768$. Maka jika nilai $D_i > 2.4768$, amatan tersebut dapat dikatakan sebagai amatan berpengaruh.

Tabel 9 nilai hii, ri, dan Di

obs	hii	ri	Di
1	0,299	0,096	0,0005
2	0,219	-0,633	0,014
3	0,2105	-1,306	0,0569
4	0,3625	-0,227	0,0037

Dari semua nilai D_i yang didapat pada tabel di samping, terlihat bahwa tidak ada nilai D_i yang lebih dari 2.4768. Maka dapat disimpulkan bahwa **tidak ada amatan berpengaruh** pada model regresi penuh antara Y dan X1 hingga X7.

5	0,0924	-1,68	0,0359
6	0,403	-0,309	0,0081
7	0,1349	1,222	0,0291
8	0,3748	-0,328	0,0081
9	0,2657	0,787	0,028
10	0,208	0,94	0,029
11	0,2489	1,288	0,0687
12	0,3268	0,81	0,0398
13	0,192	0,742	0,0163
14	0,2107	0,787	0,0206
15	0,3214	-0,233	0,0032
16	0,5306	-1,603	0,3632
17	0,1561	-1,636	0,0619
18	0,1136	-0,525	0,0044
19	0,1267	-0,852	0,0132
20	0,2296	0,001	0
21	0,5532	1,214	0,2282
22	0,3209	-0,326	0,0063
23	0,5434	0,705	0,0739
24	0,3146	1,089	0,0681
25	0,4462	-1,875	0,354
26	0,4562	0,982	0,1011
27	0,3392	1,183	0,0898

4.6 Menduga Model Terbaik

4.6.1 Metode *Backward*

Tahapan yang perlu dilakukan untuk metode *backward* adalah:

- Membentuk persamaan regresi berganda lengkap;
- Menentukan nilai $F_{parsial}$ dari masing variabel X (diuji sesuai urutan dari variabel X yang memiliki nilai mutlak korelasi terendah terhadap Y), atau menentukan nilai AIC dari masing-masing variabel X;
- Mengeliminasi variabel yang nilai $F_{parsial}$-nya tidak lebih dari F_{tabel}, atau mengeliminasi variabel yang nilai AIC-nya terkecil;
- Proses ini terus dilakukan hingga suatu persamaan terbaik diperoleh.

Dengan menggunakan bantuan *software* R, diperoleh persamaan regresi terbaik menurut metode *Backward* yaitu:

$$\boldsymbol{Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e}$$

**tahapan perhitungan AIC menggunakan R terdapat di lampiran.*

4.6.2 Metode *Forward*

Tahapan yang perlu dilakukan untuk metode *forward* adalah:

- Membentuk persamaan regresi tanpa peubah;
- Menentukan nilai $F_{sekuensial}$ dari masing variabel X (diuji sesuai urutan dari variabel X yang nilai mutlak korelasi tertinggi terhadap Y), atau menentukan nilai AIC dari masing-masing variabel X;

- Menambahkan variabel yang nilai $F_{sekuensial}$-nya lebih dari F_{tabel}, atau menambahkan variabel yang nilai AIC-nya terkecil;
- Proses ini terus dilakukan hingga suatu persamaan terbaik diperoleh.

Dengan menggunakan bantuan *software* R, diperoleh persamaan regresi terbaik menurut metode *Stepwise* yaitu:

$$\boldsymbol{Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e}$$

**tahapan perhitungan AIC menggunakan R terdapat di lampiran*.

4.6.3 Metode *Stepwise*

Metode *stepwise* merupakan gabungan dari metode *forward* dan *backward*. Metode ini dimulai dengan membentuk model regresi tanpa peubah, kemudian bekerja dengan menyeleksi variabel seperti pada metode *forward*. Namun, variabel yang telah masuk dalam model dapat dieliminasi kembali dari model seperti pada *backward*. Proses ini terus dilakukan hingga suatu persamaan terbaik diperoleh.

Dengan menggunakan bantuan *software* R, diperoleh persamaan regresi terbaik menurut metode *Backward* yaitu:

$$\boldsymbol{Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e}$$

**tahapan perhitungan AIC menggunakan R terdapat di lampiran*.

4.6.4 Metode *Best Subset*

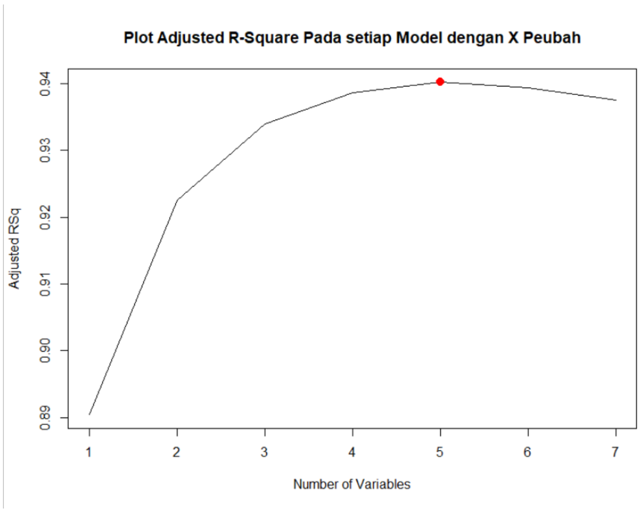
Model yang diuji pada metode ini adalah model dengan 7 peubah, maka banyaknya model yang mungkin dibentuk untuk dibandingkan ada sebanyak 2^7 model. Metode ini memanfaatkan fungsi *regsubsets* dari *package leaps* di R sehingga menghasilkan *output* sebagai berikut :

Tabel 10 Pendugaan Model Terbaik Metode *Best Subset*

Banyak Peubah pada Model	Peubah yang digunakan
1 Peubah	X4
2 Peubah	X2, X4
3 Peubah	X2, X3, X4
4 Peubah	X1, X2, X3, X4
5 Peubah	X1, X2, X3, X4, X6
6 Peubah	X1, X2, X3, X4, X5, X6
7 Peubah	X1, X2, X3, X4, X5, X6, X7

7 Model di atas dipilih dari masing-masing banyak peubah berdasarkan metode Best Subset. Lalu, pendugaan model terbaik dari 7 model di atas dilakukan dengan mencari nilai Adjusted R-squared dan Cp Mallow’s.

4.6.4.1 *Adjusted R-square*

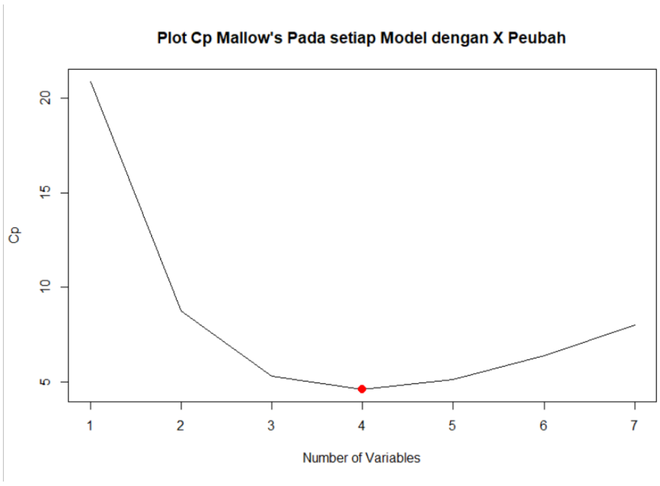


Gambar 7 Plot *Adjusted R-square*

Dari plot di atas, dapat terlihat bahwa dari 7 Model yang ada, model yang memiliki nilai Adjusted R-square terbesar adalah model dengan 5 Peubah. Maka dari itu, model dugaan terbaik berdasarkan nilai Adjusted R-square adalah :

$$\boldsymbol{Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + \beta_6 X6 + e}$$

4.6.4.2 *Cp Mallow's*



Gambar 8 Plot *Cp Mallow's*

Dari plot di atas, dapat terlihat bahwa dari 7 Model yang ada, model yang memiliki nilai Cp Mallow's terkecil adalah model dengan 4 Peubah. Cp pada model ini bernilai di sekitar 5 yang sesuai dengan syarat ketakbiasan Cp yaitu k+1. Maka dari itu, model dugaan terbaik berdasarkan nilai Cp Mallow's adalah:

$$\boldsymbol{Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e}$$

Dari 5 metode pendugaan model terbaik, didapatkan hasil sebagai berikut :

Tabel 11 Hasil Pendugaan Model Terbaik

Metode Pendugaan	Peubah	Model Regresi
Backward	X1, X2, X3, X4	$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$
Forward	X1, X2, X3, X4	$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$

Stepwise	X1, X2, X3, X4	$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$
Adjusted R-square	X1, X2, X3, X4, X6	$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + \beta_6 X6 + e$
Cp Mallow's	X1, X2, X3, X4	$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$

Dari tabel di atas didapat dua dugaan model terbaik, yaitu model 1 dengan peubah X1, X2, X3, X4 dan model 2 dengan tambahan peubah X6. Untuk menentukan mana model terbaik, akan dilakukan prosedur PRESS di tahap selanjutnya.

4.7 Memilih Model Terbaik

4.7.1 Prosedur PRESS

4.7.1.1 Model 1

Dengan Menggunakan perhitungan PRESS, didapatkan tabel di bawah:

Tabel 12 Perhitungan PRESS Model 1

Row	OBS	DELFITS	DELRES
1	70,4	69,93	0,47
2	66,88	67,05	-0,17
3	65,36	66,61	-1,25
4	72,39	73,04	-0,65
5	66,12	68,29	-2,17
6	65,67	66,40	-0,73
7	70,49	68,79	1,70
8	69,38	70,29	-0,91
9	68,75	67,52	1,23
10	67,59	66,78	0,81
11	71,64	70,81	0,84
12	67,29	66,33	0,96
13	68,95	68,05	0,90
14	70,82	69,68	1,14
15	70,66	70,93	-0,27
16	74,07	74,76	-0,69
17	68,08	70,49	-2,41
18	68,06	68,62	-0,56
19	76,11	77,08	-0,97
20	74,21	74,12	0,09
21	81,51	80,21	1,30
22	74,89	75,80	-0,91
23	81,50	79,91	1,59
24	80,97	79,20	1,77
25	77,83	80,67	-2,84
26	73,04	72,75	0,29
27	71,7	69,66	2,04

OBS : Observasi ke-i (y_i)
DELFITS : y duga setelah menyisihkan amatan ke-i ($\hat{y}_{i,-1}$)
DELRES : $y_i - \hat{y}_{i,-1}$

Nilai PRESS Model 1 didapatkan dengan menjumlahkan kuadrat DELRES. Sehingga didapat PRESS model 1 yaitu **45,1521**.

4.7.1.2 Model 2

Dengan Menggunakan perhitungan PRESS, didapatkan tabel di bawah:

Tabel 13 Perhitungan PRESS Model 2

Row	OBS	DELFITS	DELRES
1	70,4	70,11	0,29
2	66,88	67,71	-0,83
3	65,36	66,64	-1,28
4	72,39	73,26	-0,87
5	66,12	68,29	-2,17
6	65,67	66,40	-0,73
7	70,49	69,05	1,44
8	69,38	69,73	-0,35
9	68,75	67,41	1,34
10	67,59	66,57	1,02
11	71,64	70,50	1,14
12	67,29	65,97	1,32
13	68,95	68,24	0,71
14	70,82	69,75	1,07
15	70,66	71,01	-0,35
16	74,07	75,34	-1,27
17	68,08	70,31	-2,23
18	68,06	68,73	-0,67
19	76,11	76,98	-0,87
20	74,21	74,02	0,19
21	81,51	80,31	1,20
22	74,89	75,27	-0,38
23	81,50	80,04	1,46
24	80,97	79,58	1,39
25	77,83	80,65	-2,82
26	73,04	71,23	1,81
27	71,7	69,97	1,73

OBS : Observasi ke-i (y_i)
DELFITS : y duga setelah menyisihkan amatan ke-i ($\hat{y}_{i,-1}$)
DELRES : $y_i - \hat{y}_{i,-1}$

Nilai PRESS Model 2 didapatkan dengan menjumlahkan kuadrat DELRES. Sehingga didapat PRESS model 2 yaitu **45,9859**.

Karena PRESS Model 1 (45,1521) < PRESS Model 2 (45,9859), dapat disimpulkan bahwa Model 1 lebih baik menurut Prosedur PRESS. Maka, model terbaik menurut prosedur PRESS adalah :

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$$

Telah ditentukan bahwa peubah X1, X2, X3, dan X4 merupakan peubah untuk model terbaik. Hal terakhir yang perlu dilakukan untuk memilih model terbaik adalah dengan menguji intersepnya. Akan dilakukan uji $F_{parsial}$ terhadap intersep untuk melihat apakah intersep berpengaruh atau tidak.

Hipotesis:

$H_0 : \beta_0 = 0$ (intersep tak berpengaruh)

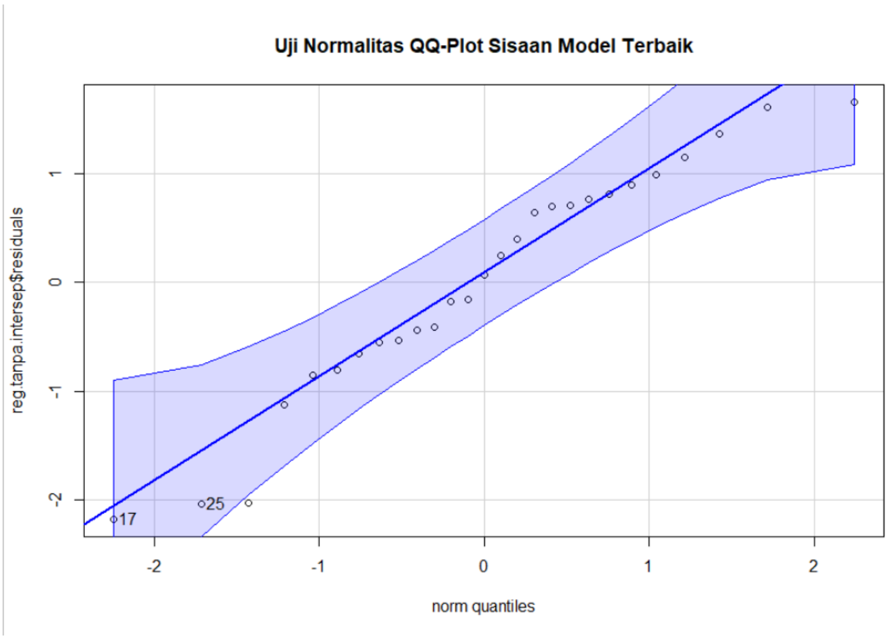
$H_1 : \beta_0 \neq 0$ (intersep berpengaruh)

Setelah melakukan uji di R, didapatkan nilai *P-Value* dari hasil uji sebesar 0.8336 < 0.05, artinya intersep bernilai nol atau nilai intersep tidak berpengaruh terhadap model. Maka dari itu, model terbaiknya adalah:

$$\boldsymbol{Y = \beta_1 X1 + \beta_2 X2 + \beta_3 X3 + \beta_4 X4 + e}$$

4.8 Pengujian Asumsi Sisaan kembali terhadap Model Terbaik

4.8.1 Normalitas



Gambar 9 Q-Q Plot Sisaan Model Terbaik

Dari Plot di atas, dapat terlihat bahwa titik titik sebaran berada di sekitar garis QQ-plot yang mengindikasikan sebaran **sisaan relatif normal**.

- Uji Formal

  Uji formal ini akan dilakukan berdasarkan taraf nyata 5% dengan uji kolmogorov-smirnov, dengan hipotesis sebagai berikut :

  H_0 : Sisaan menyebar Normal

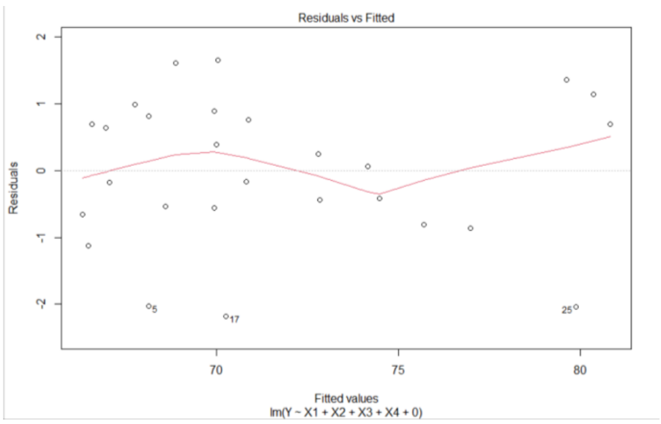
  H_1 : Sisaan tidak menyebar Normal

Tabel 14 Uji Normalitas - Kolmogorov Smirnov

Data	***P-Value***	**Keputusan**
Sisaan Model	0,6974	Sisaan menyebar normal

Karena *P-Value* (0, 6974) lebih besar dari taraf nyata (0,05), maka H_0 tidak ditolak atau sisaan menyebar normal.

4.8.2 Heteroskedastisitas



Gambar 10 Plot uji kehomogenan ragam model terbaik

Dari Plot di atas, dapat terlihat bahwa sebaran sisaan masih berada di sekitar nilai 0, kemudian tidak ada pola di sebaran tersebut. Jadi, dapat dikatakan bahwa **sisaannya relatif homogen**.

- Uji Formal

  Uji formal ini akan dilakukan berdasarkan taraf nyata 5% dengan uji glejser, dengan hipotesis sebagai berikut :

  H_0 : Sisaan homogen

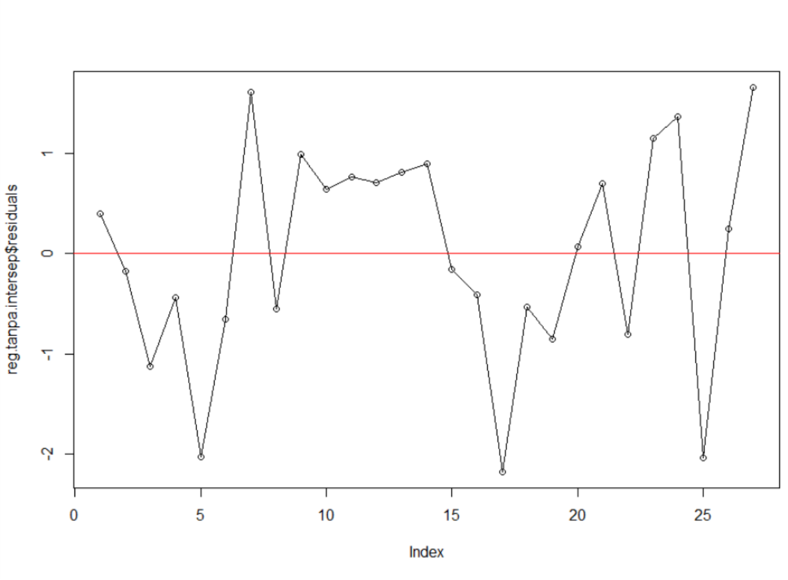
  H_1 : Sisaan tidak homogen

Tabel 15 Uji Heteroskedastisitas - glejser

Data	***P-Value***	**Keputusan**
Sisaan Model	0,439	Sisaan homogen

Karena *P-Value* (0,439) lebih besar dari taraf nyata (0,05), maka H_0 tidak ditolak atau sisaan homogen.

4.8.3 Kebebasan Sisaan



Gambar 11 Plot uji kebebasan sisaan model terbaik

Dari Plot di atas, dapat terlihat bahwa sebaran sisaan masih berada di sekitar nilai 0, kemudian tidak ada pola di sebaran tersebut. Jadi, dapat dikatakan bahwa **sisaannya relatif saling bebas**.

- Uji Formal

  Uji formal ini akan dilakukan berdasarkan taraf nyata 5% dengan uji runs, dengan hipotesis sebagai berikut :

  H_0 : Sisaan saling bebas

  H_1 : Sisaan tidak saling bebas

Tabel 16 Uji kebebasan sisaan - runs

Data	***P-Value***	**Keputusan**
Sisaan Model	0,2298	Sisaan saling bebas

Karena *P-Value* (0,2298) lebih besar dari taraf nyata (0,05), maka H_0 tidak ditolak atau sisaan saling bebas.

4.9 Interpretasi Model Terbaik

Berdasarkan hasil analisis yang telah dilakukan, didapat uji-T dan uji-F pada model regresi terbaik sebagai berikut:

Tabel 17 Pengujian Model Terbaik

Kode	Peubah	Koefisien	Sig.t
X1	Kepadatan Penduduk	0,0002	0,075
X2	Angka Harapan Hidup	0,7535	0,000
X3	PDRB per kapita	0,0214	0,043
X4	Rata-rata lama Sekolah	1,8221	0,000

$R^2 = 0.9998;\ R^2_{adj} = 0.9997;\ RSE = 1.145$

Hipotesis uji-t :
$H_0: \beta_i = 0$ (Peubah tak berpengaruh)
$H_1: \beta_i \neq 0$ (intersep berpengaruh)

Berdasarkan tabel di atas, masing-masing peubah akan signifikan pada taraf nyata 10%. Sehingga didapat model terbaiknya adalah :

$$Y = 0{,}0002X1 + 0{,}7535X2 + 0{,}0214X3 + 1{,}18221X4 + e$$

Dengan keragaman yang dapat dijelaskan oleh model (R^2) sebesar 99,98% yang mana hampir sempurna. Hal ini mengindikasikan bahwa model sudah sangat baik dalam menggambarkan peubah-peubah yang memengaruhi Indeks Pembangunan Manusia di Jawa Barat.

Koefisien pada setiap peubah menginterpretasikan bahwa setiap penambahan satu satuan pada peubah X_i, akan meningkatkan Indeks Pembangunan Manusia (IPM) sebanyak besarnya koefisien dari X_i. Berdasarkan hal tersebut, dapat terlihat bahwa peubah rata-rata sekolah memiliki pengaruh paling tinggi terhadap IPM karena memiliki koefisien terbesar yaitu 1,18221. Kemudian, disusul oleh Angka Harapan Hidup, PDRB per kapita, lalu koefisien terkecil yaitu Kepadatan Penduduk.

5. Kesimpulan

Dari analisis yang telah dilakukan, dapat disimpulkan bahwa Indeks Pembangunan Manusia (IPM) dapat dipengaruhi oleh:

1. Rata-rata lama sekolah;
2. Angka Harapan Hidup;
3. PDRB per kapita; dan
4. Kepadatan Penduduk.

Keempat faktor di atas diurutkan dari pengaruh tertinggi-terendah. Terlihat bahwa pendidikan merupakan faktor utama yang paling berpengaruh dalam meningkatkan Indeks Pembangunan Manusia (IPM) di Jawa Barat. Keempat peubah ini merupakan faktor utama yang dapat mewakilkan enam peubah lainnya.

Daftar Pustaka

[BPS] Badan Pusat Statistik. 2022. Indeks Pembangunan Manusia 2021. Jakarta: BPS.

Kacaribu RD. 2013. Analisis indeks pembangunan manusia dan faktor-faktor yang memengaruhi di Provinsi Papua [skripsi]. Bogor: Institut Pertanian Bogor.

Maddala GS. 1992. "Outliers". Introduction to Econometrics (edisi ke-2). New York: MacMillan.

Pujilestari, S. Dwidayati, N. Sugiman, S. 2017. Pemilihan Model Regresi Linier Berganda Terbaik pada Kasus Multikolinieritas berdasarkan Metode Principal Component Analysis (PCA) dan Metode Stepwise. UNNES Journal of Mathematics. 6(1) : 70-81.

Samosir, N. Siagian, P. Bangun, P. 2014. Analisa Metode Backward dan Metode Forward untuk Menentukan Persamaan Regresi Linier Berganda. Saintia Matematika. 2(4) : 345-360.

Sangkereng W, Engka DSM, Sumual JI. 2019. Faktor-faktor yang mempengaruhi indeks pembangunan manusia di Provinsi Sulawesi Utara. Jurnal Berkala Ilmiah Efisiensi. 19(4): 60-71.

Si'lang ILS, Hasid Z, Priyagus. 2019. Analisis faktor-faktor yang berpengaruh terhadap indeks pembangunan manusia. Jurnal Manajemen. 11(2): 159-169.

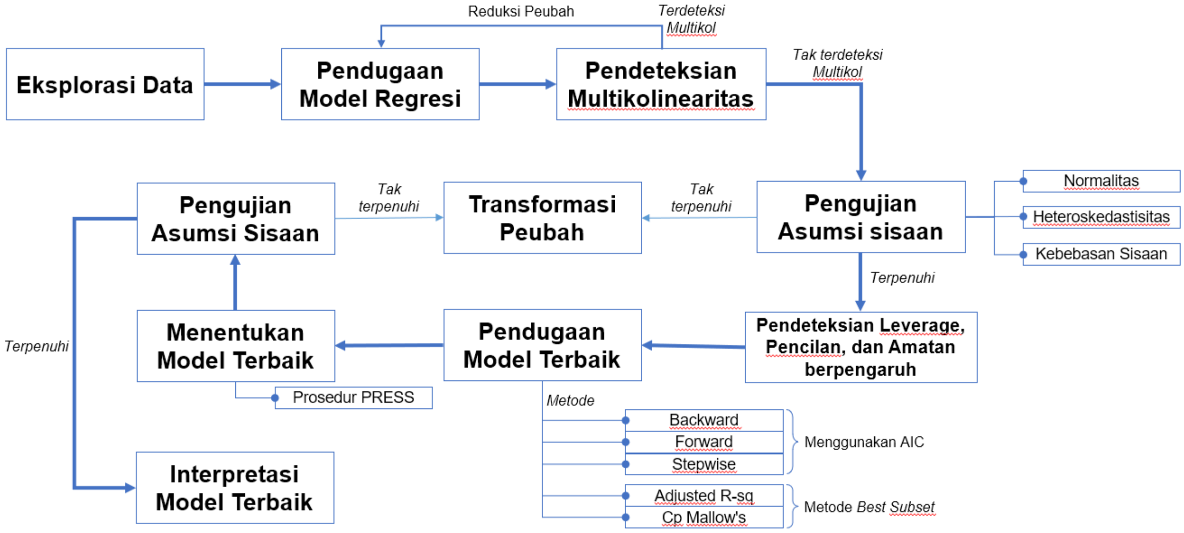
Siregar S. 2013. Metode Penelitian Kuantitatif Dilengkapi dengan Perbandingan Perhitungan Manual dan SPSS Edisi ke-1. Jakarta (ID) : Kencana Prenadamedia Group.

Taddy M. 2019. Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions. New York: McGraw-Hill.

Tarpey T. 2000. A Note on the Prediction Sum of Squares Statistic for Restricted Least Squares. The American Statistician. 54(2): 116–118.

Yamin, Sofyan dan Heri Kurniawan. 2011. SPSS Complete: Teknik Analisis Statistik Terlengkap dengan Software SPSS. Jakarta (ID): Salemba Infotek.

Lampiran 1 : Diagram Prosedur Metodologi Penelitian



Lampiran 2 : Tahapan Metode *Backward*

Metode Backward → *Eliminasi Peubah dengan AIC terkecil*

```
Start:  AIC=15.45
Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7

       Df Sum of Sq    RSS    AIC
- X7    1    0.5736 27.032 14.032
- X5    1    0.8575 27.316 14.314
- X3    1    1.5636 28.022 15.004
- X6    1    1.7383 28.197 15.171
<none>              26.459 15.453
- X1    1    2.8581 29.317 16.223
- X2    1    8.5321 34.991 21.000
- X4    1   18.4580 44.917 27.742
```

```
Step:  AIC=14.03
Y ~ X1 + X2 + X3 + X4 + X5 + X6

       Df Sum of Sq    RSS    AIC
- X5    1    0.9679 28.000 12.982
<none>              27.032 14.032
- X6    1    2.7373 29.770 14.637
- X1    1    2.8213 29.854 14.713
- X3    1    2.8461 29.878 14.735
- X2    1   12.9111 39.943 22.574
- X4    1   17.9256 44.958 25.767
```

```
Step:  AIC=12.98
Y ~ X1 + X2 + X3 + X4 + X6

       Df Sum of Sq    RSS    AIC
- X6    1    2.1032 30.103 12.938
<none>              28.000 12.982
- X1    1    3.6523 31.653 14.293
- X3    1    4.6979 32.698 15.170
- X2    1   13.4864 41.487 21.598
- X4    1   20.7502 48.750 25.954
```

```
Step:  AIC=12.94
Y ~ X1 + X2 + X3 + X4

       Df Sum of Sq    RSS    AIC
<none>              30.103 12.938
- X1    1    3.7679 33.871 14.122
- X3    1    5.9407 36.044 15.801
- X2    1   13.7515 43.855 21.096
- X4    1   28.9189 59.022 29.116
```

Tidak ada model yang lebih baik

Model Dugaan Terbaik Metode Backward :

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$$

Lampiran 3 : Tahapan Metode *Forward*

Metode Forward → *Menambahkan Peubah dengan AIC terkecil*

```
Start:  AIC=84.8
Y ~ 1

       Df Sum of Sq    RSS    AIC
+ X4    1    518.70  61.10 26.049
+ X1    1    481.62  98.17 38.854
+ X2    1    403.89 175.90 54.601
+ X6    1    314.58 265.21 65.686
+ X5    1    277.87 301.92 69.187
+ X7    1    231.07 348.72 73.078
+ X3    1    120.03 459.76 80.541
<none>              579.79 84.804
```

```
Step:  AIC=26.05
Y ~ X4

       Df Sum of Sq    RSS    AIC
+ X2    1   19.6752 41.421 17.555
+ X7    1   16.3424 44.754 19.644
+ X3    1   10.0910 51.005 23.174
+ X1    1    9.9277 51.169 23.261
+ X5    1    5.3750 55.721 25.562
+ X6    1    4.5242 56.572 25.971
<none>              61.096 26.049
```

```
Step:  AIC=17.55
Y ~ X4 + X2

       Df Sum of Sq    RSS    AIC
+ X3    1    7.5498 33.871 14.122
+ X7    1    5.7165 35.705 15.545
+ X1    1    5.3769 36.044 15.801
+ X6    1    3.7251 37.696 17.010
+ X5    1    3.1657 38.255 17.408
<none>              41.421 17.555
```

```
Step:  AIC=14.12
Y ~ X4 + X2 + X3

       Df Sum of Sq    RSS    AIC
+ X1    1    3.7679 30.103 12.938
<none>              33.871 14.122
+ X6    1    2.2188 31.653 14.293
+ X7    1    1.5809 32.290 14.831
+ X5    1    0.8619 33.009 15.426
```

```
Step:  AIC=12.94
Y ~ X4 + X2 + X3 + X1

        Df Sum of Sq    RSS    AIC
<none>               30.104 12.938
+ X6     1   2.10321 28.000 12.982
+ X7     1   1.55061 28.553 13.510
+ X5     1   0.33385 29.770 14.637
```

Tidak ada model yang lebih baik

Model Dugaan Terbaik Metode Forward :

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$$

Lampiran 4 : Tahapan Metode *Stepwise*

Metode *Stepwise* → $Menambah/mengeliminasi\ Peubah\ dengan\ AIC\ terkecil$

```
Start:  AIC=84.8
Y ~ 1

       Df Sum of Sq    RSS    AIC
+ X4    1    518.70  61.10 26.049
+ X1    1    481.62  98.17 38.854
+ X2    1    403.89 175.90 54.601
+ X6    1    314.58 265.21 65.686
+ X5    1    277.87 301.92 69.187
+ X7    1    231.07 348.72 73.078
+ X3    1    120.03 459.76 80.541
<none>              579.79 84.804
```

```
Step:  AIC=26.05
Y ~ X4

       Df Sum of Sq    RSS    AIC
+ X2    1     19.68  41.42 17.555
+ X7    1     16.34  44.75 19.644
+ X3    1     10.09  51.01 23.175
+ X1    1      9.93  51.17 23.261
+ X5    1      5.38  55.72 25.562
+ X6    1      4.52  56.57 25.971
<none>               61.10 26.049
- X4    1    518.70 579.79 84.804
```

```
Step:  AIC=17.55
Y ~ X4 + X2

       Df Sum of Sq     RSS    AIC
+ X3    1     7.550  33.871 14.122
+ X7    1     5.716  35.705 15.545
+ X1    1     5.377  36.044 15.801
+ X6    1     3.725  37.696 17.010
+ X5    1     3.166  38.255 17.408
<none>               41.421 17.555
- X2    1    19.675  61.096 26.049
- X4    1   134.481 175.903 54.601
```

```
Step:  AIC=14.12
Y ~ X4 + X2 + X3

       Df Sum of Sq     RSS    AIC
+ X1    1     3.768  30.103 12.938
<none>               33.871 14.122
+ X6    1     2.219  31.653 14.293
+ X7    1     1.581  32.290 14.831
+ X5    1     0.862  33.009 15.426
- X3    1     7.550  41.421 17.555
- X2    1    17.134  51.005 23.175
- X4    1   120.283 154.155 53.037
```

```
Step:  AIC=12.94
Y ~ X4 + X2 + X3 + X1

       Df Sum of Sq    RSS    AIC
<none>              30.103 12.938
+ X6    1    2.1032 28.000 12.982
+ X7    1    1.5506 28.553 13.510
- X1    1    3.7679 33.871 14.122
+ X5    1    0.3339 29.770 14.637
- X3    1    5.9407 36.044 15.801
- X2    1   13.7515 43.855 21.096
- X4    1   28.9189 59.022 29.116
```

Tidak ada model yang lebih baik

Model Dugaan Terbaik Metode Stepwise :

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \beta_4 X4 + e$$