

# Analisis Faktor Faktor yang Memengaruhi Angka Harapan Hidup di Jawa Timur Tahun 2022 dengan Metode Regresi Ridge<sup>1</sup>

Johannes Pande Manurung<sup>1†</sup>, Luthfiano Alzaky<sup>1</sup>, Farid Yafi Suwandi<sup>1</sup>, Yulianti Nurdzanah<sup>1</sup>, Hafly Akeyla Pari<sup>1</sup>, Sachnaz Desta Oktarina<sup>1</sup>, Akbar Rizki<sup>1</sup>, Akmal Riza Wibisono<sup>1</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia

<sup>†</sup>corresponding author: pandejohannes@apps.ipb.ac.id

## Abstrak

Provinsi Jawa Timur telah diidentifikasi sebagai provinsi dengan angka harapan hidup terendah kedua di Pulau Jawa. Guna mencapai *Sustainable Development Goals* (SDGs), maka diperlukan penelitian yang bertujuan mengetahui faktor-faktor yang secara signifikan berpengaruh terhadap angka harapan hidup di Jawa Timur dengan membandingkan model dari metode regresi linear berganda, regresi robust, regresi lasso, dan regresi ridge. Faktor-faktor yang diteliti yaitu balita dengan imunisasi lengkap (persen), jumlah rumah sakit umum (unit), prevalensi balita *stunting* (persen), sumber air minum bersih (persen), rumah tangga yang memiliki akses terhadap sanitasi layak (persen), tingkat pengangguran terbuka (persen), harapan lama sekolah (tahun), dan indeks kedalaman kemiskinan (poin). Data pada penelitian ini diperoleh dari publikasi Badan Pusat Statistik Indonesia dan Kementerian Kesehatan Republik Indonesia yang terdiri dari 38 kabupaten dan kota di Jawa Timur. Hasil penelitian menunjukkan bahwa balita dengan imunisasi lengkap (persen), jumlah rumah sakit umum (unit), rumah tangga yang memiliki akses terhadap sanitasi layak (persen), dan harapan lama sekolah (tahun) berpengaruh signifikan terhadap angka harapan hidup di Jawa Timur pada tahun 2022 dengan regresi ridge sebagai model terbaik.

**Kata Kunci:** Angka harapan hidup, Jawa Timur, regresi linear berganda, regresi lasso, regresi ridge, regresi robust.

## 1. Pendahuluan

Angka harapan hidup merupakan lama hidup rata-rata penduduk dari sejak dilahirkan yang dapat dicapai pada suatu wilayah dengan asumsi tidak ada perubahan pola mortalitas menurut umur (Ramadhani *et al.* 2020). Angka harapan hidup berperan penting sebagai alat evaluasi kinerja pemerintah dalam meningkatkan kesejahteraan masyarakat secara keseluruhan dengan fokus pada peningkatan derajat kesehatan (Sinaga *et al.* 2019), hal ini sesuai dengan tujuan pembangunan berkelanjutan mengenai kehidupan sehat dan berkelanjutan pada *Sustainable Development Goals* (SDGs). Selain dalam bidang kesehatan, studi oleh Berkman *et al.* (2014) menunjukkan bahwa angka harapan hidup juga seringkali terkait dengan tingkat kesejahteraan sosial dan ekonomi. Keterkaitan ini dapat dilihat dari variasi angka harapan hidup di tiap daerah, misalnya di pulau Jawa.

Di Pulau Jawa, Jawa Timur merupakan provinsi dengan angka harapan hidup kedua terendah. Badan Pusat Statistik (BPS) mencatat, pada tahun 2022 angka

harapan hidup penduduk Jawa Timur adalah 71,74 tahun. Meskipun berada pada peringkat kedua terbawah, selama periode 2012 hingga 2022 Jawa Timur telah berhasil meningkatkan usia harapan hidupnya sebesar 1,6 tahun. Dalam kurun waktu tersebut, rata-rata angka harapan hidup meningkat 0,23 persen per tahun. Fluktuasi angka harapan hidup tersebut dipengaruhi oleh berbagai faktor, sehingga penting untuk mengidentifikasi faktor-faktor yang berpengaruh secara rasional.

Studi awal penelitian menggunakan regresi linier berganda dengan estimasi parameter Metode Kuadrat Terkecil (MKT) mengindikasikan adanya beberapa pencilan serta peubah perancu (*confounding variable*) dalam data. Menurut Hawkins (1980) dalam Sihombing *et al.* (2022), pencilan merupakan amatan yang berada cukup jauh dari pola amatan lainnya, sedangkan *confounding variable* adalah suatu peubah yang tidak signifikan secara hipotesis namun muncul dalam penelitian dan berpengaruh terhadap peubah respons dan peubah penjelas (Yasril dan Kasjono 2008 dalam Dwijayanthi *et al.* 2013). Oleh karena itu, untuk menangani permasalahan tersebut diperlukan metode dengan hasil estimasi parameter yang stabil terhadap peubah pengganggu dan pencilan.

Berdasarkan penjelasan tersebut, penelitian ini mengambil judul “Analisis Faktor Faktor yang Mempengaruhi Angka Harapan Hidup di Jawa Timur Tahun 2022 Dengan Metode Regresi Ridge”. Penelitian ini menggunakan metode regresi linear berganda dengan tiga metode perbandingan regresi, yaitu regresi robust, lasso, dan ridge. Tujuan penelitian ini adalah mendapatkan model terbaik dalam menggambarkan angka harapan hidup di Jawa Timur tahun 2022 dan menentukan peubah-peubah penjelas yang berpengaruh signifikan, sehingga pemerintah dapat menentukan program dan langkah yang tepat untuk meningkatkan angka harapan hidup guna peningkatan kesejahteraan masyarakat.

## 2. Tinjauan Pustaka

### 2.1. Regresi Linear Berganda

Analisis regresi linear berganda adalah hubungan secara linear antara dua atau lebih peubah penjelas ( $X_1, X_2, \dots, X_p$ ) dengan peubah respons ( $Y$ ). Biasanya data yang digunakan berskala interval atau rasio (Maharadja *et al.* 2021). Misalkan  $k$  sebagai peubah penjelas, berikut merupakan model matematis regresi linier berganda (Huang 2015):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i,$$

atau

$$y = X\beta + \varepsilon, (1)$$

dimana  $\beta = (\beta_0, \beta_1, \dots, \beta_k)$  adalah vektor parameter yang akan diestimasi menggunakan MKT dan  $\varepsilon$  adalah vektor residual dari model. Proses estimasi parameter pada MKT dilakukan dengan meminimumkan kuadrat residual model sehingga diperoleh solusi untuk parameter  $\beta$  sebagai berikut (Huang 2015):

$$\hat{\beta} = (X^T X)^{-1} (X^T Y). \quad (2)$$

## 2.2 Pendeteksian Pencilan, Titik Leverage, dan Amatan Berpengaruh

Dalam analisis regresi, amatan yang berbeda dari pola umum dapat dikategorikan sebagai pencilan (*outlier*), titik leverage, atau amatan berpengaruh. Menurut Huang (2015), pencilan adalah keadaan titik-titik data yang tidak setipe dengan titik data lain yang terjadi karena titik data jauh dari pusat data, sedangkan titik leverage adalah amatan dengan nilai peubah penjelas  $X$  yang ekstrem atau di luar tren. Amatan berpengaruh adalah amatan yang memiliki pengaruh besar terhadap dugaan parameter regresi, nilai  $R^2$ , serta uji hipotesis pada saat amatan tersebut dikeluarkan dari model. *outlier* dan leverage memiliki kemungkinan yang sama untuk menjadi amatan berpengaruh.

## 2.3 Mencari Model Terbaik

Dalam mencari model terbaik untuk regresi linear berganda, salah metode yang umum digunakan adalah metode *stepwise*. Metode *stepwise* merupakan metode alternatif dalam pemilihan metode terbaik dengan cara memperhitungkan korelasi parsial model yang dihitung dari sisaan tiap peubah dalam proses analisisnya (Wohon *et al.* 2017). Selain itu, ada beberapa varian yang dapat dipertimbangkan, termasuk *robust regression* yang efektif dalam mengurangi dampak *outlier*, *lasso regression* yang melakukan seleksi fitur melalui penalti  $L_1$ , serta *ridge regression* yang mengurangi *overfitting* dengan menambahkan penalti  $L_2$ .

### 2.3.1 Regresi Robust

Regresi robust digunakan untuk menghasilkan estimasi yang cenderung tidak dipengaruhi keberadaan *outlier*. Pada umumnya, komponen pembobot ( $W$ ) akan ditambahkan pada metode estimasi parameter regresi robust (Chen 2002 dalam Husain dan Jamaluddin 2024).

$$\hat{\beta}_{(m)robust} = (X^t W X)^{-1} X^t W Y, \quad (3)$$

dimana  $\hat{\beta}_{(m)robust}$  merupakan koefisien regresi robust pada iterasi ke- $m$  yang berukuran  $(k + 1) \times 1$ .

Salah satu metode estimasi, yaitu estimator-M, dapat meminimalkan fungsi residu. Fungsi pembobot yang sering digunakan pada estimasi ini adalah Huber

yang nilainya didapatkan melalui fungsi objektif dalam regresi robust (Fox 2002 dalam Husain dan Jamaluddin 2024).

### 2.3.2 Regresi Ridge

Regresi ridge dikembangkan untuk menangani multikolinearitas yang mungkin berpengaruh terhadap nilai koefisien regresi. Regresi ridge adalah modifikasi dari metode *Ordinary Least Square* (OLS) yang menghasilkan penduga yang tidak terlalu rentan terhadap multikolinieritas dengan mengurangi varians penduga. Estimasi parameter pada model regresi Ridge adalah (Yanke *et al.* 2022):

$$\hat{\beta}_{ridge} = \arg \min_{\beta} \left\{ \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^k x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

Sedangkan estimator regresi ridge yang dipengaruhi oleh besarnya tetapan bias dengan besarnya nilai berada pada selang adalah (Lestari *et al.* 2022):

$$\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1} X'y, \lambda \geq 0$$

### 2.3.4 Regresi Lasso

Metode ini berfungsi sebagai teknik penyeleksi variabel dan pendugaan parameter, serta mengurangi koefisien regresi dari variabel-variabel yang sangat berkorelasi dengan galat sehingga koefisien tersebut mendekati nol atau bahkan mencapai nol. Oleh karena itu, lasso dapat melakukan seleksi variabel dan menangani masalah multikolinearitas dalam waktu yang bersamaan. Menurut Prabowo (2015), berikut merupakan penduga koefisien pada LASSO:

$$\hat{\beta}_{lasso} = \arg \min_{\beta} \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^k x_{ij} \beta_j \right)^2$$

dengan syarat  $\sum_{j=1}^k |\beta_j| \leq t$ , dimana  $t = \sum_{j=1}^k |\hat{\beta}_j| \leq t$ .

## 2.4 Uji Asumsi Klasik

Dalam penyusunan model regresi linear berganda, prasyarat yang harus dipenuhi dalam rangka memenuhi asumsi parameter dan koefisien regresi yang tidak bias adalah uji asumsi klasik (Abdullah 2020). Umumnya, ada empat jenis pengujian asumsi klasik yang perlu dipertimbangkan dalam pembentukan model regresi linear, yaitu uji normalitas, uji multikolinearitas, uji autokorelasi, dan uji heteroskedastisitas.

Pengujian normalitas ditujukan untuk mendeteksi sebaran model dalam rangka memenuhi asumsi sisaan yang menyebar normal (Usmadi 2020), sedangkan pengujian multikolinearitas bertujuan mendeteksi adanya hubungan linear antar

peubah penjelas (Ryan 1997 dalam Sriningsih 2018). Selanjutnya pengujian autokorelasi, pengujian ini bertujuan menemukan adanya hubungan antara data pada periode sekarang dengan periode sebelumnya sehingga memastikan bahwa sisaan dari model tersebut adalah independen satu sama lain, sedangkan pengujian heteroskedastisitas digunakan untuk mengidentifikasi adanya variasi yang berbeda dalam sisaan antara berbagai pengamatan (Ghozali dalam Abdullah 2020).

## 2.5 Pengujian Signifikansi Parameter dengan Uji Simultan (F) dan Uji Parsial (t)

Menurut Priyatno (2013) dalam Languju *et al.* (2016), uji F dilakukan ketika ada suatu hubungan linier antara peubah respons ( $y$ ) dan keseluruhan peubah penjelas ( $x_1, x_2, x_3, \dots, x_k$ ), untuk mengetahui signifikansinya, sedangkan Uji t digunakan untuk mengetahui pengaruh signifikansi peubah penjelas secara parsial terhadap peubah respons.

## 3. Metodologi Penelitian

### 3.1 Data

Penelitian ini menggunakan data sekunder yang diperoleh dari hasil survei sosial ekonomi nasional (Susenas) tahun 2022 yang dilakukan oleh Badan Pusat Statistik (BPS 2022) dan Survei Status Gizi Indonesia (SSGI) tahun 2022 yang dilakukan oleh Kementerian Kesehatan (Kemenkes 2022). Data yang digunakan terdiri atas sebelas peubah penjelas dan satu peubah respons dari masing-masing kota/kabupaten di Jawa Timur pada tahun 2022. Dasar pemilihan peubah yang digunakan dalam penelitian berasal dari penelitian terdahulu yang telah dilakukan. Keterangan peubah-peubah yang digunakan tercantum pada Tabel 1.

Tabel 1 Peubah respons dan peubah penjelas

Peubah	Keterangan	Satuan	Sumber Data	Korelasi
$Y$	Angka Harapan Hidup	Persen	BPS 2022	-
$X_1$	Persentase balita dengan imunisasi lengkap	Persen	BPS 2022	Positif (+)
$X_2$	Jumlah rumah sakit umum	Unit	BPS 2022	Positif (+)
$X_3$	Prevalensi balita <i>stunting</i>	Persen	Kemenkes 2022	Negatif (-)
$X_4$	Persentase sumber air minum bersih	Persen	BPS 2022	Positif (+)

$X_5$	Persentase rumah tangga yang memiliki akses terhadap sanitasi layak	Persen	BPS 2022	Positif (+)
$X_6$	Tingkat Pengangguran Terbuka	Persen	BPS 2022	Positif (+)
$X_7$	Harapan Lama Sekolah	Tahun	BPS 2022	Positif (+)
$X_8$	Indeks kedalaman kemiskinan	Poin	BPS 2022	Negatif (-)

### 3.2 Analisis Data

Analisis data dilakukan menggunakan *software* R. Proses analisis data dilakukan dengan ketentuan berikut:

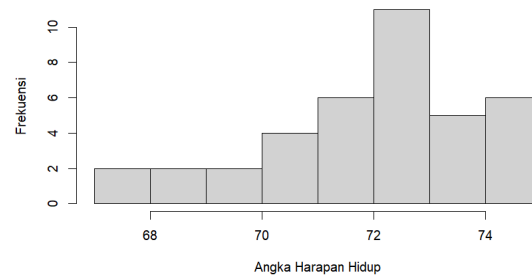
- Melakukan eksplorasi data guna memperoleh informasi umum mengenai angka harapan hidup di Jawa Timur 2022.
- Melakukan pengujian multikolinearitas antar peubah penjelas. Suatu model regresi dapat dikatakan bebas dari indikasi multikolinearitas apabila seluruh peubah penjelas memiliki nilai VIF kurang dari 10. Apabila ada paling sedikit satu peubah penjelas dengan nilai VIF lebih dari 10, perlu dilakukan eliminasi peubah penjelas tertentu hingga diperoleh nilai VIF untuk seluruh peubah penjelas kurang dari 10.
- Melakukan pendugaan parameter model regresi menggunakan uji simultan (F) dan uji parsial (t).
- Melakukan seleksi peubah penjelas dengan metode *stepwise* serta menduga parameter model regresi terbaik berdasarkan peubah yang telah terseleksi.
- Melakukan uji asumsi klasik terhadap model regresi terbaik.
- Mendeteksi titik pencilan, titik leverage, dan amatan berpengaruh dengan metode *Different of Hits* (DFFITS).
- Mengatasi anomali dengan regresi kekar (*robust*) estimasi-M, regresi *lasso*, dan regresi *ridge*, lalu membandingkan hasil pendugaan semua metode untuk mencari model terbaik serta faktor-faktor yang berpengaruh terhadap angka harapan hidup di Jawa Timur 2022.

## 4. Hasil dan Analisis Data

### 4.1 Eksplorasi Data

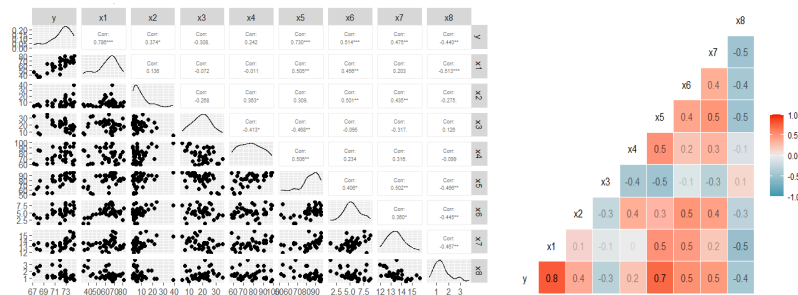
Peubah respons yang digunakan pada penelitian ini adalah angka harapan hidup yang diperoleh dari 38 kota/kabupaten di Jawa Timur pada tahun 2022. Eksplorasi data menggunakan histogram dilakukan untuk mengetahui persebaran data angka harapan hidup yang disajikan pada Grafik 1 sebagai berikut.

Gambar 1 Histogram angka harapan hidup di Jawa Timur tahun 2022



Dapat terlihat pada histogram bahwa sebaran angka harapan hidup di Jawa Timur pada tahun 2022 menjulur ke kiri. Sebaran ini mengindikasikan bahwa banyak kabupaten atau kota yang memiliki angka harapan hidup yang lebih rendah dibandingkan angka harapan hidup kabupaten atau kota lainnya.

Gambar 2 Scatter plot, histogram, dan matriks korelasi antar peubah



Melalui *scatter plot* dan histogram dapat terlihat sebaran data pada setiap peubah. Kenormalan peubah dapat diamati melalui garis yang ada di dalam histogram yang disediakan. Dalam matriks korelasi, korelasi antar peubah terlihat dengan jelas melalui warna korelasinya yang bervariasi. Warna merah menunjukkan korelasi positif, sedangkan warna biru menunjukkan korelasi negatif. Matriks korelasi ini memberikan informasi tentang kekuatan korelasi antara peubah melalui intensitas warnanya. Semakin pekat warna tersebut, semakin kuat korelasinya.

#### 4.2. Pendugaan Parameter Regresi Linear Berganda dan Pendeteksian Multikolinearitas

Pendugaan parameter regresi linear berganda dilakukan untuk mengidentifikasi peubah penjelas yang berpengaruh signifikan terhadap peubah respons. Pemodelan awal dilakukan menggunakan seluruh peubah penjelas. Dengan kriteria *p-value* kurang dari taraf nyata 5%, model regresi linear berganda penuh diuji dengan uji-t untuk menguji signifikansi setiap peubah penjelas dan uji-F untuk menguji signifikansi model.

Pendeteksian multikolinearitas dilakukan dengan mencari nilai VIF setiap peubah penjelas dengan bantuan *software* R. Suatu peubah penjelas dikatakan mengalami multikolinearitas jika nilai  $VIF \geq 10$ . Hasil pendugaan parameter regresi linear berganda dan pendeteksian multikolinearitas terdapat pada Tabel 2 sebagai berikut.

Tabel 2 Pendugaan parameter regresi linear berganda model penuh

Peubah	Koefisien	p-value	VIF
Intersep	55,127321	$5,59 \times 10^{-16}$	
$X_1$	0,101609	$2,60 \times 10^{-6} *$	1,964
$X_2$	0,035904	$1,74 \times 10^{-1}$	1,620
$X_3$	-0,010835	$6,715 \times 10^{-1}$	1,439
$X_4$	-0,007141	$6,648 \times 10^{-1}$	1,718
$X_5$	0.056189	$9,21 \times 10^{-3} *$	2,712
$X_6$	0.044335	$7 \times 10^{-1}$	1,793
$X_7$	0.407506	$7,328 \times 10^{-2}$	1,703
$X_8$	0.505827	$7,321 \times 10^{-2}$	1,755
p-value		$3,401 \times 10^{-9} *$	

Hasil analisis pendugaan parameter regresi linear berganda model penuh pada Tabel 2 menunjukkan hanya peubah persentase balita dengan imunisasi lengkap ( $X_1$ ) dan persentase rumah tangga yang memiliki akses terhadap sanitasi Layak ( $X_5$ ) yang berpengaruh signifikan terhadap peubah respons pada taraf nyata 5%. Nilai *Adjusted R-Squared* dalam model adalah 78,16% yang berarti peubah-peubah penjelas berkontribusi terhadap angka harapan hidup sebesar 78,16%. Adapun pengujian multikolinearitas menunjukkan bahwa seluruh peubah penjelas memiliki nilai  $VIF < 10$  yang bermakna tidak terdeteksi adanya multikolinearitas sehingga tidak diperlukan proses reduksi peubah. Langkah selanjutnya, perlu dilakukan pendugaan parameter regresi linear berganda dengan seleksi peubah menggunakan metode *stepwise* untuk menentukan model regresi linear berganda terbaik.

#### 4.3 Pendugaan Parameter Regresi dengan Metode *Stepwise*

Metode seleksi peubah dengan metode *stepwise* mengeliminasi peubah penjelas secara bertahap. Tahapan mengeliminasi peubah penjelas menggunakan *software* R dengan hasil terdapat pada Tabel 3.

Tabel 3 Pendugaan parameter regresi menggunakan metode *stepwise*

Model	AIC	Interpretasi
Model dengan $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ , dan $X_8$	1,29	Semua peubah ditambahkan



Model dengan $X_1, X_2, X_3, X_4, X_5, X_7$ , dan $X_8$	-0,51	Peubah $X_6$ dihilangkan
Model dengan $X_1, X_2, X_4, X_5, X_7$ , dan $X_8$	-2,33	Peubah $X_3$ dihilangkan
Model dengan $X_1, X_2, X_5, X_7$ , dan $X_8$	-4,16	Peubah $X_4$ dihilangkan

Seleksi peubah menggunakan metode *stepwise* pada Tabel 6 menghasilkan pemodelan terbaik berdasarkan AIC terendah dengan menghilangkan peubah prevalensi balita *stunting* ( $X_3$ ), persentase sumber air minum bersih ( $X_4$ ), dan tingkat pengangguran terbuka ( $X_6$ ). Langkah selanjutnya menguji signifikansi pengaruh peubah terhadap model dengan uji-F dan uji-t. Hasil pengujian menggunakan *software* R terdapat pada Tabel 4.

Tabel 4 Pendugaan parameter regresi linear berganda model terbaik

Peubah	Koefisien	p-value
Intersep	55,127321	$< 2 \times 10^{-16} *$
$X_1$	0,10465	$8,23 \times 10^{-6} *$
$X_2$	0,03854	$8,61 \times 10^{-2}$
$X_5$	0,05533	$1,2 \times 10^{-3} *$
$X_7$	0,41750	$5,47 \times 10^{-2}$
$X_8$	0,48875	$6,53 \times 10^{-2}$
$R^2$ -adjusted		0,7992
p-value		$2,829 \times 10^{-11} *$

Pendugaan parameter regresi linear berganda membentuk model terbaik yang diperoleh yaitu

$$\hat{Y} = 55,1273 + 0,10465X_1 + 0,03854X_2 + 0,0553X_5 + 0,4175X_7 + 0,4887X_8.$$

Namun, terdapat kesalahan interpretasi peubah indeks kedalam kemiskinan ( $X_8$ ) yang seharusnya pertambahan nilai indeks kedalam kemiskinan ( $X_8$ ) akan menurunkan nilai rata-rata angka harapan hidup. Kesalahan interpretasi dapat dikarenakan adanya amatan pencilan, titik leverage, maupun adanya korelasi yang signifikan antara peubah indeks kedalam kemiskinan ( $X_8$ ) dengan peubah penjelas lainnya. Gambar 2 menampilkan bahwa peubah indeks kedalam kemiskinan ( $X_8$ ) berkorelasi signifikan dengan peubah persentase balita dengan imunisasi lengkap ( $X_1$ ), persentase rumah tangga yang memiliki akses terhadap sanitasi layak ( $X_5$ ), dan harapan lama sekolah ( $X_7$ ) sehingga dapat dicurigai bahwa  $X_8$  merupakan peubah perancu (*confounding variable*).

#### 4.4 Pengujian Asumsi

Pengujian asumsi terhadap model regresi linear berganda baru dilakukan agar memastikan sifat *best linear unbiased estimator* terpenuhi. Asumsi dinilai terpenuhi jika diperoleh nilai  $p\text{-value} < \alpha$  (0,05). Asumsi yang diuji adalah asumsi nilai harapan sisaan sama dengan nol, homoskedastisitas ragam sisaan, sisaan saling bebas, serta normalitas sisaan. Hasil perhitungan dan pengujian menggunakan *software* R terdapat pada Tabel 5 sebagai berikut.

Tabel 5 Hasil uji asumsi sisaan model regresi linear berganda terbaik				
Asumsi	Jenis Uji	p-value	Keputusan	Keterangan
$E(\varepsilon) = 0$	Uji-t	1	Tidak tolak $H_0$	Asumsi terpenuhi
Homoskedastisitas	Uji Breusch-Pagan	0,1431	Tidak tolak $H_0$	Asumsi terpenuhi
Sisaan saling bebas	Uji Durbin-Watson	0,146	Tidak tolak $H_0$	Asumsi terpenuhi
Normalitas sisaan	Uji Shapiro-Wilk	0,8054	Tidak tolak $H_0$	Asumsi terpenuhi

#### 4.5 Pendeteksian Pencilan, Titik Leverage, dan Amatan Berpengaruh

##### 4.5.1 Pendeteksian Pencilan

Suatu amatan dapat dikatakan sebagai pencilan apabila nilai mutlak sisaan terbakukannya lebih dari 2 ( $|r_i| > 2$ ). Pada data ini, didapatkan tiga pencilan yang terletak pada amatan ke-3 yaitu kabupaten Trenggalek, amatan ke-9 yaitu kabupaten Jember, dan amatan ke-29 yaitu kabupaten Sumenep.

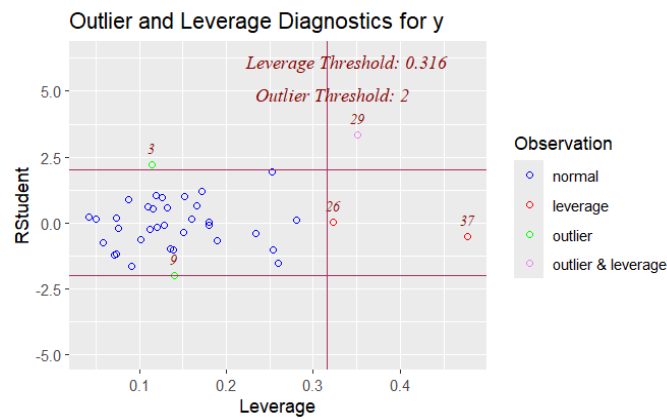
##### 4.5.2 Pendeteksian Titik Leverage

Suatu amatan dapat dikatakan sebagai titik leverage apabila nilai dari fungsi jarak antara amatan dengan nilai tengahnya lebih dari rata-rata jarak antar amatannya ( $h_{ii} > 2p/n$ ). Pada data ini, didapatkan tiga titik leverage yang terletak pada amatan ke-26 yaitu kabupaten Bangkalan, amatan ke-29 yaitu kabupaten Sumenep, dan amatan ke-37 yaitu kota Surabaya.

##### 4.5.3 Pendeteksian Amatan Berpengaruh

Berdasarkan metode DFFITS, amatan dapat dikatakan sebagai amatan berpengaruh apabila nilai  $DFFITS_i > 2\sqrt{p/n}$ . Pada data ini, didapatkan satu amatan berpengaruh yang terletak pada amatan ke-9 yaitu kabupaten Jember, amatan ke-11 yaitu kabupaten Bondowoso, amatan ke-12 yaitu kabupaten Situbondo, amatan ke-29 yaitu kabupaten Sumenep.

Gambar 3 Amatan berpengaruh, pencilan, dan titik leverage



Pencilan dan titik leverage yang teridentifikasi merupakan kabupaten atau kota yang terdapat di Jawa Timur sehingga menghapus amatan tersebut dapat menghilangkan sebagian informasi. Oleh sebab itu, pencilan serta titik leverage akan disimpan dan dilakukan pendugaan model regresi dengan metode regresi kekar (*robust regression*), regresi *lasso*, dan regresi *ridge* dalam menangani pencilan, titik leverage, dan peubah perancu (*confounding variable*).

#### 4.6 Model *Robust Regression* (Regresi Kekar)

Pembentukan model terbaik yang dapat menangani titik pencilan dan titik leverage adalah dengan metode regresi kekar estimasi-M dengan pembobot Huber. Nilai pembobot ( $w_i$ ) diperoleh dari nilai  $u_i$  dengan nilai  $u_i = \frac{e_i}{\frac{0.44707}{0.7947}} = \frac{e_i}{0.5626}$ . Hasil dugaan koefisien regresi kekar estimasi-M dengan pembobot Huber diuji dengan uji-F dan uji-t menggunakan *software R* pada iterasi ke-18 pada Tabel 6.

Tabel 6 Pendugaan parameter regresi kekar pada iterasi ke-18

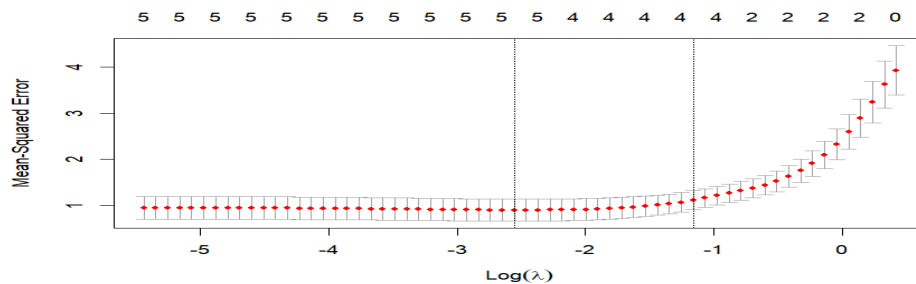
Peubah	Koefisien	p-value
Intersep	55,9668	$< 2 \times 10^{-16} *$
$X_1$	0.1086	$4,06 \times 10^{-10} *$
$X_2$	0,0432	$1,67 \times 10^{-2} *$
$X_5$	0,0552	$2,16 \times 10^{-4} *$
$X_7$	0,3026	$8,7 \times 10^{-2}$
$X_8$	0,2408	$2,94 \times 10^{-1}$
$R^2$ -adjusted		0,8644
p-value		$5,798 \times 10^{-14} *$

Dugaan model regresi kekar yang dihasilkan adalah  $\hat{Y} = 55,9668 + 0.1086X_1 + 0,0432X_2 + 0,0552X_5 + 0,3026X_7 + 0,2408X_8$  dengan dihasilkan dua peubah penjelas yang tidak signifikan yaitu harapan lama sekolah ( $X_7$ ) dan indeks kedalam kemiskinan ( $X_8$ ).

#### 4.7 Model Regresi *Lasso*

Pembentukan model yang dapat menangani anomali akibat adanya peubah perancu (*confounding variable*) adalah regresi *lasso*. Metode validasi silang dapat digunakan untuk menduga parameter *lasso* ( $\lambda$ ). Pembuatan grafik validasi silang menggunakan *software* R pada Gambar 4 sebagai berikut.

Gambar 4 Grafik validasi silang parameter *lasso* ( $\lambda$ )



Nilai parameter *lasso* ( $\lambda$ ) optimum yang diperoleh menggunakan *software* R adalah 0,707. Pendugaan koefisien regresi *lasso* menggunakan *software* R terdapat pada Tabel 8.

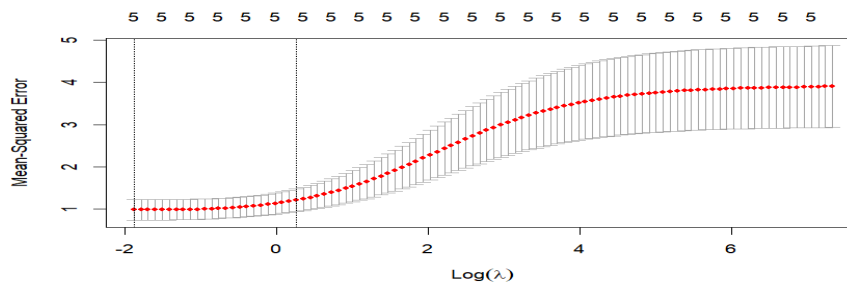
Tabel 8 Pendugaan koefisien regresi *lasso*

Peubah	Koefisien
Intersep	57,4145
$X_1$	0,0931
$X_2$	0,0308
$X_5$	0.0525
$X_7$	0.2997
$X_8$	0.1824

Dugaan model regresi *lasso* yang dihasilkan adalah  $\hat{Y} = 57,4145 + 0.0931X_1 + 0,0308X_2 + 0,0525X_5 + 0,2997X_7 + 0,1824X_8$  dengan nilai *Adjusted R-Squared* sebesar 0,81. Terdapat kesalahan interpretasi peubah indeks kedalam kemiskinan ( $X_8$ ) yang seharusnya memiliki koefisien negatif.

#### 4.8 Model Regresi *Ridge*

Regresi *ridge* dapat menstabilkan nilai dugaan koefisien regresi saat terjadi multikolinieritas dan menangani anomali akibat adanya peubah perancu (*confounding variable*). Metode ini memodifikasi MKT sehingga dihasilkan penduga koefisien regresi yang bias (Kutner et al. 2005 dalam Kusuma dan Wulansari 2017). Metode validasi silang dapat digunakan untuk menduga parameter *ridge* ( $\lambda$ ). Pembuatan grafik validasi silang menggunakan *software* R pada Gambar 5.

Gambar 5 Grafik validasi silang parameter *ridge* ( $\lambda$ )

Nilai parameter *ridge* ( $\lambda$ ) optimum yang diperoleh menggunakan *software* R adalah 0,1525. Dugaan koefisien pada regresi *ridge* untuk peubah Pendugaan koefisien regresi *ridge* menggunakan *software* R terdapat pada Tabel 9 sebagai berikut.

Tabel 9 Pendugaan koefisien regresi *ridge* ( $\lambda = 0,1525$ )

Peubah	Koefisien	p-value
Intersep	57,1165	$< 2 \times 10^{-16} *$
$X_1$	0,0861	$1,9 \times 10^{-13} *$
$X_2$	0,0353	$5,27 \times 10^{-2}$
$X_5$	0,0524	$1,07 \times 10^{-5} *$
$X_7$	0,3466	$3,12 \times 10^{-2} *$
$X_8$	0,2256	$2,55 \times 10^{-1}$
p-value	$1,44 \times 10^{-10} *$	

Dugaan model regresi *ridge* dengan parameter  $\lambda = 0,1525$  adalah  $\hat{Y} = 57,1165 + 0,0861X_1 + 0,0353X_2 + 0,0524X_5 + 0,3466X_7 + 0,2256X_8$  dengan nilai *Adjusted R-Squared* sebesar 0,808. Terdapat kesalahan interpretasi peubah indeks kedalam kemiskinan ( $X_8$ ) sehingga digunakan parameter *ridge* yang lain hingga diperoleh agar koefisien  $X_8$  menjadi negatif yaitu  $\lambda = 0,5$ . Pendugaan koefisien regresi *ridge* dengan parameter  $\lambda = 0,5$  menggunakan *software* R terdapat pada Tabel 10 sebagai berikut.

Tabel 10 Pendugaan koefisien regresi *ridge* ( $\lambda = 0,5$ )

Peubah	Koefisien	p-value
Intersep	60,2458	$< 2 \times 10^{-16} *$
$X_1$	0,0646	$4,71 \times 10^{-13} *$
$X_2$	0,0305	$3,89 \times 10^{-2} *$
$X_5$	0,0449	$3,36 \times 10^{-7} *$
$X_7$	0,2900	$1,63 \times 10^{-2} *$
$X_8$	-0,0184	$9,01 \times 10^{-1}$
p-value	$2,14 \times 10^{-8} *$	

Dugaan model regresi *ridge* dengan parameter  $\lambda = 0,5$  adalah  $\hat{Y} = 60,2458 + 0,0646X_1 + 0,0305X_2 + 0,0449X_5 + 0,2900X_7 - 0,0184X_8$  dengan nilai *Adjusted R-Squared* sebesar 0,756. Dugaan model ini memberikan interpretasi yang tepat terhadap peubah indeks kedalam kemiskinan ( $X_8$ ). Meskipun tidak berpengaruh signifikan secara hipotesis, peubah indeks kedalam kemiskinan ( $X_8$ ) berpengaruh terhadap signifikansi peubah penjelas lainnya.

#### 4.9 Pemilihan Model Terbaik dengan Membandingkan Metode Kuadrat Terkecil, Metode Regresi Robust, Regresi Lasso, dan Regresi Ridge

Model regresi terbaik dipilih dengan membandingkan setiap metode pendugaan yang telah digunakan. Nilai koefisien, *Adjusted R-Squared*, serta signifikansinya terhadap peubah respon dapat dilihat pada Tabel 11 sebagai berikut.

Tabel 11 Dugaan koefisien regresi setiap metode

Peubah	MKT	Robust	Lasso	Ridge ( $\lambda = 0,1525$ )	Ridge ( $\lambda = 0,5$ )
	Koefisien	Koefisien	Koefisien	Koefisien	Koefisien
Intersep	55,1273 *	55,9668 *	57,4145	57,1165 *	60,2458 *
$X_1$	0,10465 *	0,1086 *	0,0931	0,0861 *	0,0646 *
$X_2$	0,03854	0,0432 *	0,0308	0,0353	0,0305 *
$X_5$	0,05533 *	0,0552 *	0,0525	0,0524 *	0,0449 *
$X_7$	0,41750	0,3026	0,2997	0,3466 *	0,2900 *
$X_8$	0,48875	0,2408	0,1824	0,2256	-0,0184
$Adj-R^2$	0,7992	0,8644	0,81	0,808	0,756

Model regresi terbaik adalah model regresi yang dapat mengatasi kesalahan interpretasi peubah indeks kedalam kemiskinan ( $X_8$ ). Selain itu, model tersebut harus memiliki nilai *Adjusted R-Squared* > 0,60 sehingga model terbaik yang diperoleh adalah model regresi *ridge* dengan parameter  $\lambda = 0,5$ , yaitu:

$$\hat{Y} = 60,2458 + 0,0646X_1 + 0,0305X_2 + 0,0449X_5 + 0,2900X_7 - 0,0184X_8.$$

## 5. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan dengan menggunakan metode regresi *ridge*, didapatkan sebuah model regresi linear terbaik yang memenuhi seluruh asumsi sebagai berikut:

$$Y = 60,2458 + 0,0646X_1 + 0,0305X_2 + 0,0449X_5 + 0,29X_7 - 0,0184X_8.$$

Mengacu pada model regresi linear berganda di atas, angka harapan hidup di Provinsi Jawa Timur pada tahun 2022 dipengaruhi oleh beberapa faktor, yaitu persentase balita dengan imunisasi lengkap ( $X_1$ ), jumlah rumah sakit umum ( $X_2$ ),

persentase rumah tangga dengan akses terhadap sanitasi layak ( $X_5$ ), harapan lama sekolah ( $X_7$ ), dan indeks kedalaman kemiskinan ( $X_8$ ). Dalam model terbaik, faktor persentase balita dengan imunisasi lengkap ( $X_1$ ), jumlah rumah sakit umum ( $X_2$ ), persentase rumah tangga dengan akses terhadap sanitasi layak ( $X_5$ ), harapan lama sekolah ( $X_7$ ) berpengaruh secara signifikan sedangkan indeks kedalaman kemiskinan ( $X_8$ ) tidak berpengaruh signifikan. Oleh sebab itu, indeks kedalaman kemiskinan ( $X_8$ ) merupakan peubah perancu yang mana tidak berpengaruh signifikan terhadap peubah respons, tetapi muncul sebagai peubah yang berpengaruh terhadap signifikansi peubah bebas lainnya.

### Daftar Pustaka

- Abdullah RR. 2020. Pengaruh Current Ratio (CR), Debt Equity Ratio (DER), dan ukuran perusahaan (size) terhadap return on equity. Studi pada perusahaan manufaktur sub sektokonsumsi yang Terdaftar di BEI Periode 2015-2018 [skripsi]. Jakarta: Universitas Pendidikan Indonesia.
- Berkman LF, Kawachi I, Glymour MM. 2014. *Social Epidemiology*. New York : Oxford University Press.
- Dwijayanthi KN, Igusti AYU, Suciptawati NLP. 2013. Hubungan pengaruh penolong kelahiran terhadap status kelahiran bayi dengan kontrol variabel confounding di kabupaten buleleng (studi kasus: puskesmas sukasada II). *E-Jurnal Matematika*. 2(4): 29-32.
- Huang D, Cabral R, Torre FDL. 2015. Robust regression. *IEEE transactions on pattern analysis and machine intelligence*. 38(2): 363-375.
- Husain A, Jamaluddin SRW. 2024. Pemodelan data angka kematian bayi menggunakan regresi robust. *SAINTEK: Jurnal Sains, Teknologi & Komputer*. 1(1): 1-7.
- Kusuma GW, Wulansari IY. 2019. Analisis kemiskinan dan kerentanan kemiskinan dengan regresi ridge, lasso, dan elastic-net di provinsi jawa tengah tahun 2017. *seminar nasional official statistics*. 2019(1): 503-513
- Languju O, Mangantar M, Tasik HHD. 2016. Pengaruh return on equity, ukuran perusahaan, price earning ratio dan struktur modal terhadap nilai perusahaan property and real estate terdaftar di bursa efek indonesia. *Jurnal Berkala Ilmiah Efisiensi*. 16(2): 387–98.
- Lestari PS, Martha S, Debataraja NN. 2022. Penerapan metode regresi ridge pada kasus angka kematian bayi di provinsi Jawa Timur. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*. 11(4): 603–610.
- Maharadja AN, Maulana I, Dermawan BA. 2021. Penerapan metode regresi linear berganda untuk prediksi kerugian negara berdasarkan kasus tindak pidana

- korupsi. *Journal of Applied Informatics and Computing*. 5(1): 95-102.
- Prabowo FKH, Wilandari Y, Rusgiyono A. 2015. Pemodelan pertumbuhan ekonomi jawa tengah menggunakan pendekatan Least Absolute Shrinkage and Selection Operator (Lasso). *Jurnal Gaussian*. 4(4): 855-864.
- Ramadhani E, Salwa N, Mazaya MS. 2020. Identifikasi faktor-faktor yang memengaruhi angka harapan hidup di Sumatera Tahun 2018 Menggunakan Analisis Regresi Spasial Pendekatan Area. *Journal of Data Analysis*. 3(2): 62-75.
- Sihombing PR, Suryadiningrat, Sunarjo DA, Yuda YPAC. 2022. Identifikasi data outlier (pencilan) dan kenormalan data pada data univariat serta alternatif penyelesaiannya. *Jurnal Ekonomi dan Statistik Indonesia*. 2(3). 307-316
- Sinaga SP, Wanto A, Solikhun S. 2019. Implementasi jaringan syaraf tiruan resilient backpropagation dalam memprediksi angka harapan hidup masyarakat Sumatera Utara. *Jurnal Infomedia: Teknik Informatika, Multimedia, dan Jaringan*. 4(2): 81-88.
- Sriningsih M, Hatidja D, Prang JD. 2018. Penanganan multikolinearitas dengan menggunakan analisis regresi komponen utama pada kasus impor beras di provinsi sulut. *Jurnal Ilmiah Sains*. 18(1):18-24.
- Usmadi. 2020. Pengujian persyaratan analisis (uji homogenitas dan uji normalitas. *Inovasi Pendidikan*. 7(1): 50-62.
- Wohon SC, Hatidja D, Nainggolan N. 2017. Penentuan model regresi terbaik dengan menggunakan metode stepwise. *Jurnal Ilmiah Sains*. 17(2): 80-88.
- Yanke A, Zendrato NE, Soleh AM. 2022. handling multicollinearity problems in Indonesia' s economic growth regression modeling based on endogenous economic. *Indonesian Journal of Statistics and Its Applications*. 6(2): 228–244.