



SKEWNESS

Jurnal Statistika, Aktuaria dan Sains Data

Volume 1, No. 2, Oktober 2024

Model Regresi dengan Metode LASSO untuk Mengatasi Multikolinieritas Studi Kasus Banyaknya Gizi Buruk di Kabupaten Banyumas

Melda Juliza^{1*}, Felinda Arumningtyas², Novita Eka Chandra³

^{1,2,3} Program Studi Statistika, Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Jenderal Soedirman, Purwokerto, Indonesia

E-mail korespondensi: melda.juliza@unsoed.ac.id

Abstrak. Uji asumsi klasik pada analisis regresi salah satunya yaitu multikolinieritas. Terjadinya multikolinieritas berarti antar variabel prediktor saling berkorelasi. Jika data penelitian mengandung multikolinieritas akan mengakibatkan koefisien regresi yang dihasilkan tidak efisien. Salah satu solusi untuk mengatasi masalah tersebut dapat menggunakan metode LASSO (*least absolute shrinkage and selection operator*). Metode ini menyusutkan koefisien regresi dari variabel prediktor yang memiliki korelasi tinggi menjadi tepat pada nol atau mendekati nol. Data penelitian yang digunakan adalah banyaknya kasus gizi buruk pada balita di 40 puskesmas di Kabupaten Banyumas. Berdasarkan uji VIF terdapat dua variabel prediktor yang mengalami masalah multikolinieritas, yaitu variabel jumlah bayi yang mendapat Imunisasi Polio (X_4) dan jumlah bayi yang mendapat Imunisasi DPT-HB3 (X_6). Sehingga masalah tersebut akan diatasi menggunakan metode LASSO. Berdasarkan hasil pengujian metode LASSO, diperoleh bahwa koefisien yang signifikan terdapat pada variabel pemberian ASI eksklusif pada bayi (X_1) dan variabel pemberian munisasi DPT-HB3 (X_6).

Kata kunci: Gizi buruk; Multikolinieritas; Regresi; *Least absolute shrinkage*.

1 Pendahuluan

Status balita gizi kurang merupakan status gizi yang berdasarkan pada indeks berat badan menurut tinggi badan (BB/TB $>-2s.d-3SD$) sebesar 5,08% balita, dan balita gizi buruk (BB/TB $<-3SD$) sebesar 0,28% [1]. Menurut Depkes RI [2], gizi buruk dipengaruhi oleh banyak faktor yang saling berkesinambungan. Oleh karena itu, perlu adanya kajian mendalam terkait faktor-faktor yang mempengaruhi jumlah balita penderita gizi buruk di Kabupaten Banyumas berdasarkan Kecamatan Tahun 2023. Berdasarkan Data Profil Kesehatan Kabupaten Banyumas 2024, balita berat badan kurang (BB/U) sebesar 11,52% balita, sedangkan balita stunting dengan status gizi berdasarkan pada indeks tinggi badan menurut umur (TB/U) yaitu sebesar 11,35%.

Analisis regresi merupakan teknik analisis yang dapat digunakan untuk menjelaskan hubungan fungsional antar variabel [3]. Analisis regresi linier berganda adalah teknik statistika yang digunakan untuk menganalisis pengaruh antara dua atau lebih variabel prediktor terhadap satu variabel respons [4].

Menurut Gujarati dan Porter [5], terdapat beberapa asumsi yang harus dipenuhi dalam regresi linier berganda yaitu hubungan antara variabel respons dan prediktor adalah linier dalam parameter, tidak adanya hubungan linier antar-variabel prediktor (tidak adanya multikolinieritas), nilai rata-rata dari error adalah nol, tidak ada korelasi di antara error-errornya, dan variansi setiap error adalah sama atau homoskedastisitas.

Multikolinieritas antar variabel prediktor dalam regresi linier berganda dapat menyebabkan tingginya variansi dalam model. Dengan demikian, model yang terbentuk tidak dapat mempresentasikan populasi dengan baik. Suatu model dikatakan model yang baik jika model yang terbentuk memiliki error kecil dan variansi kecil. Alternatif penyelesaian regresi linier berganda jika terdapat permasalahan antar variabel prediktor yang mengandung multikolinieritas adalah menerapkan metode LASSO (*Least Absolute Shrinkage and Selection Operator*).

Multikolinieritas dapat menghasilkan koefisien regresi dari analisis regresi berganda menjadi sangat lemah atau tidak memberikan hasil analisis yang mampu menjelaskan pengaruh variabel prediktor yang terlibat [6]. Adanya multikolinieritas juga akan mengakibatkan:

1. Nilai *standar error* cenderung akan semakin besar seiring dengan tingginya tingkat korelasi antarvariabel prediktor.
2. Jika terdapat multikolinieritas, nilai selang kepercayaan cenderung menjadi lebih besar akibat besarnya standar error, sehingga sangat sulit untuk menyangkal hipotesis nol dalam penelitian.
3. Jika terdapat multikolinieritas yang tinggi, mengakibatkan kemungkinan atau risiko gagal menolak hipotesis yang salah meningkat.
4. Jika multikolinieritas tinggi, memungkinkan diperoleh nilai koefisien determinasi (R^2) yang tinggi juga, akan tetapi tidak dapat menjelaskan sifat maupun pengaruh dari variabel prediktor yang bersangkutan.

Terdapat beberapa riset terdahulu yang sudah menelaah terkait permasalahan gizi buruk, di antaranya Janah dan Kartini [7] yang menganalisa variabel apa saja yang dianggap signifikan mempengaruhi terjadinya kasus gizi buruk di Kabupaten Bojonegoro menggunakan metode regresi linier berganda. Adapun variabel yang digunakan adalah persentase bayi baru lahir mendapat inisiasi menyusui dini, persentase bayi mendapat ASI eksklusif, persentase balita mendapat kapsul vitamin A, persentase balita memiliki KMS atau buku KIA, persentase balita ditimbang empat kali atau lebih dalam enam bulan

terakhir, dan persentase balita kurus mendapatkan makanan tambahan. Hasil riset menyatakan bahwa variabel yang berpengaruh signifikan terhadap kejadian balita yang mengalami gizi buruk adalah persentase balita ditimbang empat kali atau lebih dalam enam bulan terakhir, dan persentase balita kurus mendapatkan makanan tambahan.

Hananti, dkk. [8] memodelkan kasus gizi buruk balita di Indonesia menggunakan *panel quantile regression model*. Penelitian ini menggunakan pemodelan regresi data panel melalui pendekatan fixed effects model (FEM). Hasil penelitian menunjukkan bahwa dampak kasus gizi buruk balita dengan faktor kemiskinan di level kuantil rendah (0,25) akan menyebabkan kasus gizi buruk juga rendah, sedangkan pada level kuantil yang sedang (0,5) akan menyebabkan kasus gizi buruk yang sedang, begitu juga untuk level kuantil yang tinggi (0,75) akan menyebabkan kasus gizi buruk yang tinggi.

Artikel ini akan mengkaji mengenai faktor-faktor yang mempengaruhi gizi buruk dengan menggunakan variabel yang kemungkinan berpengaruh. UNICEF [9] menyatakan bahwa penyebab tidak langsung terjadinya gizi buruk yakni persediaan pangan yang meliputi oleh pemberian ASI Eksklusif bayi, pola asuh anak meliputi oleh rumah tangga berperilaku hidup bersih dan sehat, pelayanan kesehatan dasar meliputi oleh lima imunisasi dasar, serta sanitasi dan air bersih. Umumnya faktor-faktor yang mempengaruhi gizi buruk saling berkaitan, hal ini menyebabkan terjadinya multikolinieritas antar variabel prediktor dalam analisis regresi berganda. Sebagai alternatif penentuan faktor-faktor yang mempengaruhi gizi buruk di Kabupaten Banyumas 2023 ditelaah dengan regresi LASSO.

Adapun tujuan penelitian ini adalah membentuk model regresi dengan metode LASSO terhadap faktor-faktor yang mempengaruhi gizi buruk Kabupaten Banyumas tahun 2023 yang terdeteksi masalah multikolinieritas. Harapannya, hasil penelitian ini dapat menjadi bahan referensi dinas terkait.

2 Metodologi

Draper dan Smith [10] menyatakan bahwa hubungan antara satu variabel respons dengan satu atau lebih variabel prediktor dinyatakan dalam regresi linier. Hubungan tersebut secara umum dapat diformulakan seperti berikut ini:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (1)$$

Dengan:

Y_i : Variabel respons
 i : 1, 2, ..., n

- $\beta_0, \beta_1, \dots, \beta_k$: Parameter
 $X_{i1}, X_{i2}, \dots, X_{ik}$: Variabel prediktor
 ε_i : Error

Di dalam analisis regresi terdapat beberapa asumsi klasik, salah satunya adalah tidak terjadinya multikolinieritas. Pemeriksaan ada atau tidaknya multikolinieritas dalam model regresi dapat dilihat berdasarkan nilai VIF (*variance inflation factor*). Jika nilai VIF lebih dari 10, maka dikatakan terjadi multikolinieritas dalam model. Nilai VIF dapat dihitung menggunakan rumus berikut [11]:

$$VIF_j = \frac{1}{(1-R_j^2)}, j = 1, 2, 3, \dots, k \quad (2)$$

Dengan R_j^2 adalah koefisien determinasi dari variabel prediktor X_j yang diregresikan dengan variabel lainnya.

Menurut James, dkk [12] untuk mengatasi masalah multikolinieritas, salah satu caranya yaitu dengan *shrinkage* (menyusutkan) parameter yang ditaksir. Teknik *shrinkage* sering dikenal sebagai teknik regularisasi. Regularisasi memungkinkan pengurangan parameter mendekati nol relatif terhadap perkiraan kuadrat terkecil.

Tibshirani [13] menyatakan bahwa LASSO (Least Absolute Shrinkage and Selection Operator) dapat menyusutkan koefisien regresi dari variabel prediktor yang berkorelasi tinggi dengan error mendekati 0 atau tepat sama dengan 0, sehingga LASSO dapat mengestimasi parameter secara simultan serta berperan sebagai metode seleksi variabel sekaligus juga dapat mengatasi multikolinieritas. Penduga koefisien LASSO didapat dengan meminimumkan jumlah kuadrat error, dan dapat diformulasikan sebagai berikut [13]:

$$\hat{\beta}^{(lasso)} = \arg \min_{\beta} \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (3)$$

dengan fungsi kendala $\sum_{j=1}^k |\beta_j| \leq t$ dan $t = \sum_{j=1}^k |\hat{\beta}_j|$

Nilai t adalah pengontrol besarnya penyusutan pada penduga koefisien LASSO dengan $t \geq 0$. Nilai t yang kecil berimbas pada beberapa koefisien regresi menjadi nol. Oleh karena itu, variabel-variabel yang berpengaruh tinggi dalam model akan terpilih, sedangkan variabel-variabel yang berpengaruh rendah akan tereliminasi. Dengan demikian hal tersebut menyebabkan solusi LASSO memberikan hasil yang efisien. Bila nilai $t < t_0$ dengan $t_0 = \sum_{j=1}^k |\hat{\beta}_j|$ akan menjadikan koefisien OLS menyusut mendekati nol atau tepat

sama dengan nol. Sehingga dapat dikatakan bahwa LASSO berperan sebagai seleksi variabel. Namun, bila $t > t_0$ maka penduga koefisien LASSO memberikan hasil sama dengan penduga OLS (*Ordinary Least Square*). Penduga koefisien LASSO diperoleh secara bertahap dengan menetapkan nilai koefisien awal semua variabel prediktor bernilai nol. Metode LASSO dapat menyeleksi variabel mana saja yang dapat digunakan untuk membentuk model. Model yang paling optimal dapat digunakan untuk melakukan klasifikasi data baru dengan menggunakan variabel-variabel yang signifikan saja.

Data yang digunakan dalam penelitian ini yaitu data sekunder yang diambil dari publikasi Dinas Kesehatan (Dinkes) Kabupaten Banyumas yaitu Profil Kesehatan Kabupaten Banyumas 2024. Publikasi tersebut diperoleh melalui website resmi Dinas Kesehatan Banyumas, yakni www.dinkes.banyumaskab.go.id. Data untuk seluruh variabel dalam penelitian ini adalah data Tahun 2023 berdasarkan kecamatan di Kabupaten Banyumas, yakni terdapat 40 pengamatan ($n = 40$).

Terdapat dua variabel yang digunakan dalam penelitian ini, yakni variabel respons dan variabel prediktor. Variabel respons yang digunakan adalah data jumlah bayi gizi buruk Kabupaten Banyumas Tahun 2023 (Y). Adapun variabel prediktor yang diduga berpengaruh terhadap gizi buruk ada enam variabel, yakni:

Tabel 1 Variabel prediktor.

Variabel	Keterangan
X_1	Jumlah bayi usia 0-6 Bulan yang diberi ASI Eksklusif
X_2	Jumlah bayi yang mendapat Imunisasi Hepatitis B
X_3	Jumlah bayi yang mendapat Imunisasi BCG
X_4	Jumlah bayi yang mendapat Imunisasi Polio
X_5	Jumlah bayi yang mendapat Imunisasi Campak
X_6	Jumlah bayi yang mendapat Imunisasi DPT-HB3

Adapun tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan data jumlah bayi gizi buruk dan faktor-faktor yang mempengaruhinya di Kabupaten Banyumas Tahun 2023
2. Membentuk model regresi linier dengan metode OLS
3. Melakukan uji multikolinieritas dengan melihat nilai variance inflation factor (VIF)
4. Pemodelan regresi linier dengan metode LASSO
5. Melakukan interpretasi dari model yang didapatkan
6. Membuat simpulan berdasarkan masalah yang telah ditelaah.

3 Hasil dan Pembahasan

3.1 Regresi Linier Berganda

Berdasarkan hasil analisis dengan bantuan *software* R, diperoleh penduga parameter model regresi linier berganda dengan OLS adalah sebagai berikut:

Tabel 2 Penduga parameter regresi linier berganda.

Variabel	Penduga Parameter
konstanta	0,537686
X_1	-0,006214
X_2	0,013945
X_3	-0,008631
X_4	0,010313
X_5	0,013162
X_6	-0,026937

Berdasarkan nilai penduga parameter pada **Tabel 2**, diperoleh model regresi linier berganda pada kejadian bayi gizi buruk Kabupaten Banyumas Tahun 2023 seperti berikut ini:

$$\hat{Y} = 0,537686 - 0,006214X_1 + 0,013945X_2 - 0,008631X_3 + 0,010313 X_4 + 0,013162X_5 - 0,026937X_6 \quad (4)$$

Selanjutnya dilakukan uji signifikansi parameter model, yaitu secara serentak dan secara individu. Uji serentak menggunakan uji *F* dengan hipotesis sebagai berikut:

H_0 : Tidak ada pengaruh variabel prediktor terhadap variabel respons secara bersama-sama (model tidak sesuai atau tidak signifikan)

H_1 : Ada pengaruh variabel prediktor terhadap variabel respons secara bersama-sama (model sesuai atau signifikan)

Hasil uji *F* dengan bantuan *software* R diperoleh *p_value* sebesar $0,01601 < 0,05$. Hal ini berarti bahwa pada tingkat signifikansi 5% tolak H_0 atau dikatakan bahwa model signifikan. Dengan demikian, dapat disimpulkan bahwa secara bersama-sama ada pengaruh variabel prediktor terhadap variabel respons. Selanjutnya, setelah uji secara serentak, periksa uji secara individu, menggunakan uji *t* dengan hipotesis sebagai berikut:

H_0 : Variabel prediktor tidak berpengaruh signifikan terhadap variabel respons

H_1 : Variabel prediktor berpengaruh signifikan terhadap variabel respons

dengan kriteria penolakan akan menolak H_0 jika $p_value < \alpha$. Pada **Tabel 3** berikut disajikan hasil uji signifikansi parameter secara individu.

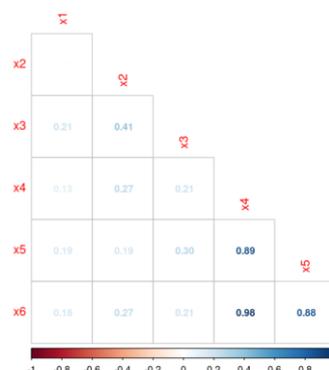
Tabel 3 Hasil uji signifikansi parameter secara individu

Variabel	Penduga	Standard error	t-hitung	p-value
Konstanta	0,537686	1,091385	0,493	0,6255
X_1	-0,006214	0,002327	-2,671	0,0117 *
X_2	0,013945	0,012103	1,152	0,2575
X_3	-0,008631	0,006915	-1,248	0,2207
X_4	0,010313	0,018945	0,544	0,5898
X_5	0,013162	0,008641	1,523	0,1373
X_6	-0,026937	0,018957	-1,421	0,1647

Berdasarkan **Tabel 3**, diperoleh bahwa hanya variabel X_1 yang signifikan, yaitu Jumlah bayi usia 0-6 Bulan yang diberi ASI Eksklusif mempunyai p_value sebesar $0,0117 < 0,05$. Hal ini berarti bahwa pada tingkat signifikansi 5% H_0 ditolak. Sehingga dapat disimpulkan bahwa terdapat pengaruh Jumlah bayi usia 0-6 Bulan yang diberi ASI Eksklusif terhadap Jumlah bayi gizi buruk. Adapun pengaruhnya adalah pengaruh arah negatif.

3.2 Multikolinieritas

Pengujian multikolinieritas dilakukan dengan memeriksa korelasi antar variabel prediktor. Korelasi yang kuat antar variabel prediktor ini dapat mengindikasikan adanya multikolinieritas.



Gambar 1 Korelasi antar variabel prediktor.

Plot korelasi pada **Gambar 1** memperlihatkan hubungan antar variabel prediktor pada data. Dapat dilihat bahwa terdapat korelasi yang kuat antara variabel X_4 dengan variabel prediktor X_5 dan X_6 . Korelasi bernilai positif dan mendekati satu, hal ini mengindikasikan adanya multikolinieritas antar variabel prediktor.

Selanjutnya memeriksa multikolinieritas antar variabel prediktor dalam regresi linier berganda. Uji multikolinieritas bertujuan untuk menguji apakah pada model regresi dijumpai adanya korelasi antar variabel prediktor. Uji multikolinieritas pada penelitian ini menggunakan statistik uji VIF (*Variance Inflation Factor*). Jika nilai VIF yang diperoleh > 10 , maka terjadi multikolinieritas antar variabel prediktor tersebut [5]. Berikut disajikan dalam **Tabel 4** nilai VIF untuk masing-masing variabel prediktor yang diperoleh dengan bantuan *software R*.

Tabel 4 Uji multikolinieritas

Variabel	Nilai VIF
X_1	1,170440
X_2	1,344614
X_3	1,396705
X_4	31,766700
X_5	5,512488
X_6	28,242163

Berdasarkan **Tabel 4**, dapat dilihat bahwa terdapat 4 variabel yang mempunyai nilai $VIF < 10$ yakni X_1, X_2, X_3 , dan X_5 , sedangkan 2 variabel lainnya yakni X_4 dan X_6 , mempunyai nilai $VIF > 10$. Artinya bahwa terdapat multikolinieritas antar variabel prediktor pada model regresi tersebut, yang akan menyebabkan tingginya variansi dalam model. Multikolinieritas dapat menghasilkan koefisien regresi dari analisis regresi berganda menjadi sangat lemah atau tidak memberikan hasil analisis yang mampu menjelaskan pengaruh variabel prediktor yang terlibat. Sehingga model yang terbentuk tidak dapat mempresentasikan populasi dengan baik. Sebagai alternatif, maka diterapkan metode LASSO sebagai penyelesaian regresi linier berganda jika antar variabel prediktor mengandung multikolinieritas.

3.3 Metode LASSO

Setelah dilakukan seleksi variabel dengan bantuan *software R*, didapat penduga koefisien LASSO seperti yang disajikan pada **Tabel 5** berikut ini:

Tabel 5 Penduga koefisien LASSO

Variabel	Penduga koefisien LASSO
Konstanta	0,889210657
X_1	-0,005444756
X_2	.
X_3	.
X_4	.
X_5	.
X_6	-0,001938672

Berdasarkan **Tabel 5**, diperoleh bahwa variabel X_1 dan X_6 masuk dalam model. Dengan demikian model regresi LASSO dapat ditulis sebagai berikut:

$$\hat{Y} = 0,889210657 - 0,005444756X_1 - 0,001938672X_6 \quad (5)$$

Multikolinieritas yang terdapat pada variabel X_4 dan X_6 telah teratasi dengan metode LASSO. Hal ini terlihat dengan adanya perubahan nilai koefisien parameter pada variabel X_6 , yaitu -0,026937 menjadi -0,001938672 sehingga menghasilkan persamaan regresi yang lebih sederhana.

4 Kesimpulan

Berdasarkan hasil analisis regresi yang dilakukan dengan bantuan software R, maka dapat disimpulkan bahwa model regresi dengan metode LASSO (*Least Absolute Shrinkage and Selection Operator*) menghasilkan 2 variabel yang berpengaruh secara signifikan terhadap jumlah bayi gizi buruk di Kabupaten Banyumas Tahun 2023, yaitu jumlah bayi (0-6 Bulan) yang diberi ASI Eksklusif dan jumlah bayi yang mendapat Imunisasi DPT-HB3.

Referensi

- [1] Kesehatan, D. (2024). *Profil Kesehatan Kabupaten Banyumas Tahun 2023*. Purwokerto: Dinas Kesehatan Kabupaten Banyumas.
- [2] RI, D. K. (2005). *Rencana Aksi Nasional Pencegahan dan Penanggulangan Gizi Buruk 2005-2006*. Jakarta: Departemen Kesehatan RI.
- [3] Sudjana. (2013). *Metode Statistika Edisi 7*. Bandung: Tarsito.
- [4] Dewanti, D. (2023). *Metode Statistika Populer untuk Penelitian*. Bogor: Exsight.
- [5] Gujarati, D. N., & Porter, D. C. (2009). *Basics Econometrics Fifth Edition*. New York: McGraw-Hill/Irwin.

- [6] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. USA: John Wiley & Sons.
- [7] Janah, M., & Kartini, A. Y. (2022). Penerapan Metode Regresi Linier Berganda Pada Kasus Balita Gizi Buruk Di Kabupaten Bojonegoro. *Jurnal Statistika Dan Komputasi (STATKOM)*, 1(2), 2963-0398. doi:<https://doi.org/10.32665/statkom.v1i2.1170>
- [8] Hananti, H., Jaya, I. G., & Ginanjar, I. (2023). Pemodelan Kasus Gizi Buruk Balita di Indonesia Menggunakan Panel Quantile Regression Model. *Statistika*, 23(2), 116-122. doi:<https://doi.org/10.29313/statistika.v23i2.2025>
- [9] UNICEF. (1998). *The States on teh World's Children*. New York: Oxford University Press.
- [10] Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. New York: John Wiley & Sons.
- [11] Montgomery, D. C., & Runger, G. C. (2011). *Applied Statistics and Probability for Engineers Fifth Edition*. New York: John Wiley & Sons.
- [12] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Application in R*. New York : Springer.
- [13] Tibshirani, R. (2011). Regression Shrinkage and Selection via the Lasso: A Retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1), 273-282.