



KODE ARTIKEL : TEK-24-1-3-2

REGRESI RIDGE DAN SIMULASINYA PADA DATA KEMISKINAN

Siti Nursiati Sugeng, Budi Pratikno

Departemen Matematika, Universitas Jenderal Soedirman

*email korespondensi : siti.sugeng@mhs.unsoed.ac.id

ABSTRAK

Riset ini membahas ridge regression model dan simulasinya pada data kemiskinan di Jawa Tengah. Jika multikolinearitas terjadi pada model regresi berganda, koefisiennya akan membesar dan tidak stabil. Peneliti memiliki beberapa metode regresi untuk mengatasi multikolinearitas yaitu, elastic-net, lasso, dan regresi ridge. Selanjutnya, penelitian ini difokuskan pada regresi ridge karena secara efektif mengurangi varians koefisien regresi. Pada metode regresi ridge, peneliti menambahkan tetapan bias c pada kuadrat terkecil sehingga nilai koefisien regresi menjadi stabil. Dua metode yang digunakan untuk menentukan nilai tetapan bias, yaitu dipilih secara random ($c = 0,2$) dan atau melalui pendekatan Bayes. Hasil penelitian menunjukkan bahwa faktor yang berpengaruh secara signifikan terhadap kemiskinan di Provinsi Jawa Tengah tahun 2023 adalah variabel presentase angka melek huruf (X_3), variabel angka harapan hidup (X_4), dan variabel IPM (X_6). Sementara itu, variabel presentase jumlah kasus diare balita (X_1), variabel presentase rata-rata lama sekolah (X_2), dan variabel presentase tingkat partisipasi angkatan kerja (X_5) tidak berpengaruh secara signifikan. Dalam konteks ini, metode regresi ridge memberikan solusi pada penentuan faktor yang berpengaruh dalam model tersebut.

Kata kunci : regresi linier berganda, regresi ridge, multikolinearitas, data kemiskinan dan simulasinya

PENDAHULUAN

Kemiskinan merupakan salah satu permasalahan kompleks yang terjadi di Indonesia. Pemerintah melakukan upaya pengentasan kemiskinan dengan cara yang sistematis, terencana, dan bersinergi dengan dunia usaha dan masyarakat untuk mengurangi jumlah penduduk miskin. Menurut Badan Pusat Statistika (2023), jumlah penduduk miskin di Indonesia mencapai 25,90 juta orang per Maret 2023 sedangkan pada September 2022 jumlah penduduk miskin sebesar 26,36 juta orang. Adapun pulau Jawa menjadi salah satu pulau yang memiliki kontribusi cukup besar bagi pertumbuhan tingkat perekonomian Indonesia salah satunya Provinsi Jawa Tengah dengan jumlah penduduk sebanyak 13,29 juta orang dan jumlah penduduk miskin sebesar 3,79 juta orang per Maret 2023.

Faktor-faktor yang mempengaruhi kemiskinan di Provinsi Jawa Tengah dapat dianalisis dengan menggunakan teknik statistika, salah satunya dengan analisis regresi. Analisis regresi harus memenuhi asumsi klasik, antara lain tidak terjadi multikolinearitas. Menurut Gujarati (2006) jika terjadi multikolinearitas maka koefisien dalam model regresi menjadi tidak stabil. Salah satu metode untuk mengatasi permasalahan multikolinearitas adalah regresi ridge (Daoud, 2007). Regresi ridge merupakan metode estimasi koefisien regresi yang diperoleh melalui penambahan konstanta bias c sehingga nilai koefisien menjadi stabil.

Metode regresi ridge sudah banyak digunakan oleh para peneliti. Penelitian Rozyadi (2018) yang menerapkan regresi ridge pada IPM di Jawa Tengah yang menghasilkan nilai tetapan bias dan MSE pada teori regresi ridge (Lawless dan Wang, 1976) lebih kecil daripada teori regresi ridge (Hoerl, Kernald, dan Baldwin, 1975). Wasilaine et al., (2014) memodelkan pengaruh pertumbuhan bayi di Kota Masohi dengan hasil bahwa persamaan regresi yang baru setelah melalui proses metode regresi ridge tidak mengandung multikolinearitas lagi. Arisandi et al., (2021) melakukan penelitian membandingkan regresi linier berganda dan regresi ridge pada kasus curah hujan di Jawa Barat dengan hasil nilai mean absolute percentage error (MAPE) dan root mean square error (RMSE) lebih kecil daripada regresi linier berganda. Kusuma dan Wulansari (2020) melakukan penelitian mengenai kemiskinan di Jawa Barat yang menghasilkan bahwa metode regresi ridge memiliki tingkat keefisienan lebih baik. Tambunan et al., (2022) menganalisis kemiskinan di Jawa Barat yang menunjukkan bahwa nilai tetapan bias c menggunakan teori Al-Hassan (2010) menyebabkan koefisien regresi yang dihasilkan semakin menyusut.



Berdasarkan penelitian Kusuma dan Wulansari (2020) dan Tambunan dkk (2022) riset ini difokuskan pada simulasi analisis kemiskinan di Provinsi Jawa Tengah tahun 2023 dengan metode regresi ridge. Perbedaan penelitian ini dengan riset sebelumnya adalah pada konsep riil data dan teori regresi ridge yang digunakan, yaitu data kemiskinan di Provinsi Jawa Tengah tahun 2023 sedangkan teori terkait ridge regression adalah teori Hoerl dan Kennard (1970) tentang nilai c dapat dipilih secara random dalam interval 0 dan 1 ($0 < c < 1$). Selanjutnya riset ini bertujuan untuk mengetahui faktor yang diduga berpengaruh pada data kemiskinan diantaranya presentase dan pemodelan regresi ridge melalui penetapan bias c secara random.

MATERI DAN METODE

A. Data Penelitian

Data yang digunakan dalam penelitian adalah data kemiskinan di Provinsi Jawa Tengah tahun 2023 yang ditampilkan pada Tabel 1.

Tabel 1. Data Penelitian

Variabel	Keterangan	Satuan
X_1	Jumlah kasus penyakit diare balita	Persen (%)
X_2	Rata-rata lama sekolah	Persen (%)
X_3	Angka melek huruf	Persen (%)
X_4	Angka harapan hidup	Tahun
X_5	Tingkat partisipasi angkatan kerja	Persen (%)
X_6	IPM	Skor

B. Tahapan Penelitian

Prosedur penelitian dilakukan untuk melakukan analisis faktor kemiskinan melalui pendekatan metode regresi ridge adalah sebagai berikut.

Membuat statistika deskriptif untuk variabel terikat dan variabel bebas;

Membuat plot antara variabel terikat dengan masing-masing variabel bebas;

Melakukan uji asumsi klasik untuk mendeteksi adanya multikolinearitas antara variabel bebas dengan menggunakan nilai VIF. Jika didapatkan nilai $VIF > 10$, maka data mengalami masalah multikolearitas;

Melakukan transformasi data terhadap matriks X menjadi Z dan vektor Y menjadi Y^* melalui centering and rescaling;

Hitung matriks $Z^T Z$ yaitu matriks korelasi dari variabel bebas serta hitung $Z^T Y^*$ yaitu korelasi dari variabel bebas terhadap variabel tak bebas Y ;

Menghitung nilai VIF dengan berbagai nilai c ($0 < c < 1$) dimana nilai (c) dipilih secara random;

Menghitung nilai penaksir parameter $\hat{\beta}(c)$ dengan berbagai kemungkinan tetapan bias (c);

Menetapkan nilai tetapan bias (c) dengan pertimbangan nilai VIF dan $\hat{\beta}_R$ berdasarkan teori Hoerl dan Kennard;

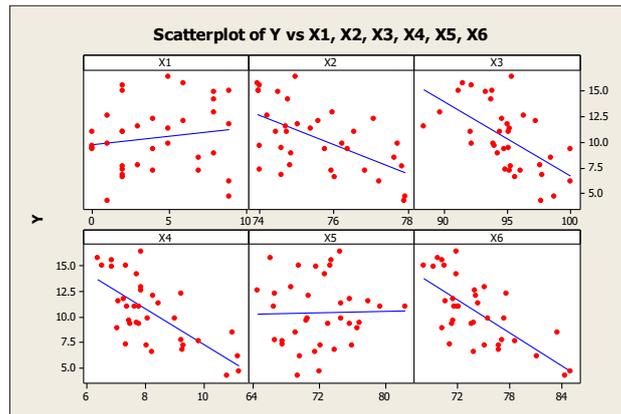
Membuat persamaan model regresi ridge dan ujinya;

Transformasikan model ke bentuk asal untuk interpretasi hasil model

HASIL DAN PEMBAHASAN

A. Scatter Plot

Langkah awal riset dilakukan dengan membuat scatterplot antara variabel terikat dengan masing-masing variabel bebas dibuat untuk mengetahui ada tidaknya hubungan antara variabel terikat dengan variabel bebas. Hasil scatterplot disajikan pada Gambar 1.



Gambar 1. Scatterplot

Berdasarkan Gambar 1, scatterplot dari variabel terikat dengan masing-masing variabel bebas memiliki hubungan linier satu sama lain (adanya multikolinearitas). Oleh karena itu, data memenuhi persyaratan untuk dilakukan pemodelan regresi linier berganda.

B. Pemodelan Regresi Linier Berganda

Pemodelan regresi linier berganda dilakukan dengan menggunakan bantuan software SPSS, dengan hasil diperoleh pada Tabel 2.

Tabel 22. Regresi Linier Berganda

Koefisien:					
	Estimasi	Std. Error	t-value	p-value	R-square
Intersep	99,4123	45,5292	2,183	0,0375	0,6056
X_1	0,2119	0,1301	1,628	0,1147	
X_2	-0,1928	0,5578	-0,346	0,7322	
X_3	-0,3466	0,2101	-1,650	0,1102	
X_4	0,1879	1,3779	0,136	0,8925	
X_5	-0,1413	0,1033	-1,368	0,1823	
X_6	-0,4552	0,4202	-1,083	0,2880	

Dari Tabel 2 ini, maka diperoleh persamaan regresi linier berganda untuk data kemiskinan Provinsi Jawa Tengah adalah

$$Y = 99,412 + 0,212X_1 - 0,193X_2 - 0,347X_3 + 0,188X_4 - 0,141X_5 - 0,455X_6$$

C. Uji Asumsi Klasik

1. Uji Normalitas

Uji normalitas digunakan untuk menguji apakah sampel data berasal dari populasi yang memiliki distribusi normal. Metode yang digunakan adalah uji Shapiro Wilk. Hasil pengujian uji normalitas terdapat pada Tabel 3'

Tabel 3. Uji Normalitas

Uji Normalitas Shapiro-Wilk	
W_{hitung}	W_{tabel}
0,9889	0,934



Pada Tabel 3 diperoleh nilai W_{hitung} sebesar 0,9889, karena $W_{hitung} > W_{tabel}$ maka H_0 gagal ditolak. Sehingga error berdistribusi normal.

2. Uji Autokorelasi

Uji autokorelasi menggunakan uji Durbin Watson untuk menilai apakah terdapat korelasi antara periode sekarang dengan periode waktu sebelumnya. Berikut diperoleh hasil uji autokorelasi tercantum pada Tabel 4.

Tabel 4. Uji Autokorelasi

Uji Autokorelasi Durbin Watson	
DW	p-value
1,7004	0,1103

Mengacu pada Tabel 4 dan standar keputusan autokorelasi Durbin Watson, disimpulkan bahwa tidak terjadi autokorelasi pada model.

3. Uji Heteroskedastisitas

Uji Breush Pagan digunakan untuk mendeteksi adanya gejala heteroskedastisitas dalam model regresi. Uji Heteroskedastisitas disajikan pada Tabel 5.

Tabel 5. Uji Heteroskedastisitas

Uji Heteroskedastisitas Breush Pagan		
BP	p-value	df
4,0995	0,6632	6

Tabel 5 meunjukkan bahwa nilai Breush Pagan sebesar 4,0995 dengan p-value 0,5997. Karena nilai p-value $> 0,05$, maka dapat disimpulkan H_0 diterima sehingga tidak terjadi heteroskedastisitas.

4. Uji Multikolinearitas

Dalam mendeteksi multikolinearitas pada model regresi menggunakan nilai VIF. Apabila nilai VIF dari suatu variabel bebas lebih besar dari 10 maka terindikasi adanya masalah multikolinearitas. Nilai VIF masing-masing variabel bebas tercantum pada Tabel 6.

Tabel 6. Uji Multikolinearitas

X_1	X_1	X_1	X_1	X_1	X_1
1,051588	3,623139	2,092533	21,699716	1,214209	21,56892

Dari Tabel 6, dapat dilihat hasil uji multikolinearitas menggunakan nilai VIF diperoleh nilai VIF variabel X_4 dan X_6 yaitu sebesar 21,699716 dan 21,568962 mengalami multikolinearitas karena nilai VIF > 10 .

D. Standarisasi Data

Data yang digunakan pada penelitian ini memiliki satuan yang berbeda antara variabel terikat dengan variabel bebas, sehingga perlu dilakukan standarisasi data menggunakan centering dan rescaling. Berikut adalah data yang sudah dilakukan proses standarisasi pada Tabel 7 dengan menggunakan bantuan software NCSS 2024.

Tabel 7. Data Standarisasi



Y^*	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
0,03115	-0,12195	-0,10625	0,01343	-0,11324	-0,22861	-0,09069
0,11204	-0,17813	-0,17500	0,09314	-0,05040	-0,31162	-0,01381
⋮	⋮	⋮	⋮	⋮	⋮	⋮
-0,18841	-0,12195	-0,12311	0,18699	0,13550	0,07645	0,09630
-0,14271	-0,06578	-0,09198	0,17671	0,12895	-0,22695	0,11031

E. Matriks Korelasi

Matriks korelasi diperoleh dari data yang telah distandarisasi tujuan untuk mengetahui hubungan antara variabel terikat dengan variabel bebas atau antar variabel bebas Hasil matriks korelasinya dengan menggunakan bantuan software NCSS 2024 dapat dilihat pada Tabel 8.

Tabel 8. Matriks Korelasi

	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Y^*
Z_1	1	0,140980	0,023236	0,111270	-0,092688	0,091138	0,151124
Z_2	0,140980	1	0,418787	0,802680	-0,073049	0,823570	-0,586947
Z_3	0,023236	0,418787	1	0,6881854	-0,220291	0,626787	-0,594783
Z_4	0,111270	0,802680	0,6881854	1	-0,243243	0,971058	-0,694681
Z_5	-0,092688	-0,073049	-0,220291	-0,243243	1	-0,276856	0,017834
Z_6	0,091138	0,823570	0,626787	0,971058	-0,276856	1	-0,698155
Y^*	0,151124	-0,586947	-0,594783	-0,694681	0,017834	-0,698155	1

F. Menentukan Nilai c

Pencarian nilai c yang optimal bertujuan untuk menghasilkan nilai VIF < 10. Dengan melihat VIF plot dan Ridge Trace peneliti dapat menentukan nilai c saat VIF mendekati 1 dan koefisien regresi ridge stabil pada nilai c tersebut. Tabel 9 merupakan output nilai c optimal dengan melihat nilai VIF menggunakan NCSS.

Tabel 9. Nilai VIF

c	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
0,200000	0,7021	0,9198	0,8611	0,5807	0,7222	0,5786

Selanjutnya, koefisien dari masing-masing variabel seperti pada Tabel 10.

Tabel 10. Koefisien Regresi Ridge

c	Z_1	Z_2	Z_3	Z_4	Z_5	Z_6
0,000000	0,1982	-0,0781	-0,2832	0,0754	-0,1789	-0,5971

Berdasarkan Tabel 9 dan 10 terlihat bahwa nilai VIF masing-masing variabel kurang dari 10 dan masing-masing koefisien regresi ridge stabil. Dengan demikian, diperoleh persamaan regresi ridge yaitu

$$Y = 0,1739X_1 - 0,1318X_2 - 0,2323X_3 - 0,1932X_4 - 0,1203X_5 - 0,2546X_6$$

G. Uji Hipotesis



1. Uji Simultan

Uji simultan dengan menggunakan uji F . taraf signifikansi yang digunakan $\alpha = 0,05$ dengan daerah kritis jika $F_{hitung} > F_{tabel}$, maka H_0 ditolak. Hasil uji simultan menggunakan software NCSS terdapat pada Tabel 11.

Tabel 11. Uji Simultan

Analisis Varian untuk $c = 0,2$						
	DF	SS	MS	F_{hitung}	p-value	R-square
Intersep	1	1,97E-09	1,97E-09	5,7419	0,000539	0,5517

Dari Tabel 11 diperoleh hasil F_{hitung} sebesar 5,7419 dengan F_{tabel} sebesar 2,38. Karena $F_{hitung} > F_{tabel}$ maka H_0 ditolak sehingga terdapat pengaruh setidaknya satu variabel bebas terhadap variabel terikat.

2. Uji Parsial

Uji parsial dilakukan dengan menggunakan uji t . Taraf signifikan yang digunakan adalah $\alpha = 0,05$ dengan daerah kritis jika $t_{hitung} > t_{tabel}$ atau $-t_{hitung} < -t_{tabel}$, maka H_0 ditolak. Hasil pengujian secara parsial dengan software NCSS tercantum pada Tabel 12.

Tabel 12. Uji Parsial

X	$\hat{\beta}(c)$	$se \hat{\beta}(c)$	t_{hitung}	t_{tabel}	Keputusan
X_1	0,1739	0,1060098	1,6404	1,68957	Tidak Signifikan
X_2	-0,1318	0,1213138	-1,0864		Tidak Signifikan
X_3	-0,2323	0,1174252	-1,9782		Signifikan
X_4	-0,1932	0,09642004	-2,0037		Signifikan
X_5	-0,1203	0,1075275	-1,1187		Tidak Signifikan
X_6	-0,2546	0,09625722	-2,6449		Signifikan

Dari Tabel 12, diperoleh variabel yang presentase angka melek huruf (X_3), variabel angka harapan hidup (X_4), dan variabel IPM (X_6) berpengaruh secara signifikan terhadap variabel terikat Y karena nilai $t_{hitung} > t_{tabel}$, sedangkan variabel presentase jumlah kasus diare balita (X_1), variabel presentase rata-rata lama sekolah (X_2), dan variabel presentase tingkat partisipasi angkatan kerja (X_5) tidak berpengaruh secara signifikan terhadap variabel terikat Y karena nilai $t_{hitung} < t_{tabel}$.

SIMPULAN

Berdasarkan penelitian yang telah dilakukan, dapat diperoleh kesimpulan (1) persamaan model regresi ridge untuk data kemiskinan di Provinsi Jawa Tengah adalah $Y = 0,1739X_1 - 0,1318X_2 - 0,2323X_3 - 0,1932X_4 - 0,1203X_5 - 0,2546X_6$, dimana persamaan ini menunjukkan koefisien yang lebih stabil dengan nilai VIF nya mendekati 1 dengan tetapan bias c sebesar 0,2, dan (2) variabel yang berpengaruh secara signifikan pada kemiskinan di Provinsi Jawa Tengah pada tahun 2023 pada penelitian menggunakan metode regresi ridge yaitu variabel presentase angka melek huruf (X_3), variabel angka harapan hidup (X_4), dan variabel IPM (X_6).

DAFTAR PUSTAKA

Arisandi, R., Ruhiat. D., dan Marlina, E. (2021). Implementasi Ridge Regression untuk Mengatasi Gejala Multikolinearitas pada Pemodelan Curah Hujan Berbasis Data Time Series Klimatologi. *Jurnal Riset Matematika dan Sains Terapan*. 1(1). 1–11.

Badan Pusat Statistika Provinsi Jawa Tengah. (2023). Jawa Tengah dalam Angka Tahun 2023. Jawa Tengah: Badan Pusat Statistik Provinsi Jawa Tengah.

Daoud, J. I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*. 949(1). 1-6.



- Gujarati, D. N. (2006). *Dasar-dasar Ekonometrika*. Jakarta: Erlangga.
- Hoerl, A. E. & Kennard. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*. Vol. 12, 55-67.
- Kusuma. G. W., dan Wulansari. I. Y. (2020). Analisis Kemiskinan dan Kerentanan Kemiskinan dengan Regresi Ridge. Lasso. dan Elastic-Net di Provinsi Jawa Tengah Tahun 2017. *Jurnal Seminar Nasional Official Statistics*. 2019(1). 503–513.
- Rozyadi. M. Z. (2018). Penerapan Metode Regresi Ridge untuk Mengatasi Masalah Multikolinearitas pada Kasus Indeks Pembangunan Manusia di Provinsi Jawa Tengah. Skripsi. Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Tambunan. R. F., Statistika. P., Matematika. F., & Alam. P. (2022). Pemodelan New Ridge Regression Estimator pada Tingkat Kemiskinan di Kabupaten / Kota Provinsi Jawa Barat Tahun 2020. *Jurnal Bandung Conference Series: Statistics*. 2(2). 317–323.
- Wasilaine. T. L., Talakua. M. W., dan Lesnussa. Y. A. (2014). Model Regresi Ridge untuk Mengatasi Model Regresi Linier Berganda yang Mengandung Multikolinearitas (Studi Kasus : Data Pertumbuhan Bayi di Kelurahan Namaelo RT 001 . Kota Masohi). *Jurnal Barenkeng*. 8(1). 31–38.