

PENERAPAN METODE *GENERALIZED RIDGE REGRESSION* DALAM MENGATASI MASALAH MULTIKOLINEARITAS

NI KETUT TRI UTAMI¹, I KOMANG GDE SUKARSA²,
I PUTU EKA NILA KENCANA³

^{1,2,3} Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali,
e-mail: ¹ami.tara14@gmail.com, ²sukarsakomang@yahoo.com, ³i.putu.enk@gmail.com

Abstract

Ordinary least square is parameter estimation method for linier regression analysis by minimizing residual sum of square. In the presence of multicollinearity, estimators which are unbiased and have a minimum variance can not be generated. Multicollinearity refers to a situation where regressor variables are highly correlated. Generalized Ridge Regression is an alternative method to deal with multicollinearity problem. In Generalized Ridge Regression, different biasing parameters for each regressor variables ($k = k_1, k_2, \dots, k_p$) were added to the least square equation after transform the data to the space of orthogonal regressors. The analysis showed that Generalized Ridge Regression was satisfactory to overcome multicollinearity.

Keywords: *Linear regression, parameter estimation, multicollinearity, Generalized Ridge Regression*

1. Pendahuluan

Analisis regresi merupakan sebuah alat statistik yang memberi penjelasan tentang pola hubungan (model) antara dua peubah atau lebih [1]. Dalam pembentukan model regresi dilakukan pendugaan terhadap parameter atau koefisien regresi (β) dengan menggunakan metode kuadrat terkecil dengan jalan meminimumkan jumlah kuadrat galatnya. Penduga yang memiliki sifat tak bias dan mempunyai ragam minimum tidak dapat dihasilkan apabila terjadi multikolinearitas.

Multikolinearitas merupakan situasi dimana terjadi korelasi yang tinggi antar peubah-peubah bebas yang mengakibatkan determinan dari matriks $X'X$ akan mendekati 0 sehingga akan menyebabkan matriks tersebut hampir *singular* yang mengakibatkan nilai dari penduga parameternya tidak stabil. Suatu metode formal untuk mendeteksi multikolinearitas adalah *Variance Inflation Factors* (VIF) yang menyatakan seberapa besar ragam koefisien regresi dugaan membesar di atas nilai idealnya. Nilai $VIF > 5$ menunjukkan bahwa peubah-peubah bebas

¹ Alumni Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

^{2,3} Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

yang terlibat dalam masalah multikolinearitas.

Pada penelitian ini, penulis akan menerapkan metode *Generalized Ridge Regression* (GRR) dalam menganalisis data yang mengalami multikolinearitas dimana ditambahkan konstanta bias yang berbeda untuk setiap peubah bebas ($k = k_1, k_2, \dots, k_p$) pada metode kuadrat terkecil setelah sebelumnya dilakukan proses ortogonalisasi terhadap peubah bebas.

Generalized Ridge Regression merupakan *Generalized Ridge Regression* merupakan pengembangan dari prosedur *Ordinary Ridge Regression* yang memungkinkan terdapat parameter bias (k) berbeda untuk setiap peubah bebas [2,3]. Pembahasan mengenai *Generalized Ridge Regression* akan lebih sederhana apabila dilakukan transformasi terhadap data sehingga peubah bebas menjadi peubah bebas yang *orthogonal* [4]. Pertama-tama, asumsikan bahwa Λ merupakan matriks $p \times p$ dimana anggota dari diagonal utamanya merupakan nilai eigen ($\lambda_1, \lambda_2, \dots, \lambda_p$) dari matriks $X'X$ dan jika $T_{p \times p}$ merupakan matriks orthogonal dari vektor eigen yang bersesuaian dengan λ , maka $T'X'XT = \Lambda$.

Misalkan $Z = XT$ dan $\alpha = T'\beta$, maka model linear persamaan regresi menjadi

$$y = X\beta + \varepsilon = (ZT')(T\alpha) + \varepsilon = Z\alpha + \varepsilon \quad (1)$$

Penduga kuadrat terkecil dari α merupakan solusi dari $(Z'Z)\hat{\alpha} = Z'y$ yang setara dengan $\Lambda\hat{\alpha} = Z'y$. Maka penduga kuadrat terkecil menjadi $\hat{\alpha} = \Lambda^{-1}Z'y$. Vektor penduga parameter awal dapat dihitung dengan menggunakan persamaan $\hat{\beta} = T\hat{\alpha}$. Mengacu kepada persamaan (1) sebagai bentuk kanonik dari model, penduga *generalized ridge* merupakan solusi dari $(\Lambda + K)\hat{\alpha}_{GR} = Z'y$ dimana K merupakan matriks diagonal dengan anggota (k_1, k_2, \dots, k_p). Koefisien *generalized ridge* pada model awal yaitu $\hat{\beta}_{GR} = T\hat{\alpha}_{GR}$.

Selanjutnya, pertimbangan untuk pemilihan parameter bias pada K berdasarkan pada nilai MSE. MSE diminimumkan dengan memilih

$$k_j = \frac{\sigma^2}{\alpha_j^2}, \quad j = 1, 2, \dots, p \quad (2)$$

Untuk menentukan nilai k_j digunakan pendekatan iteratif [3]. Ketika kuadrat panjang vektor parameter ($\hat{\alpha}'_{GR}\hat{\alpha}_{GR}$) yang diduga tidak mengalami perubahan yang signifikan dari iterasi $i - 1$ ke iterasi i , maka proses iterasi dihentikan.

Setelah mendapatkan penduga koefisien regresi dari metode *Generalized Ridge Regression*, dipastikan masalah multikolinearitas telah diatasi dengan melihat nilai *Variance Inflation Factors* (VIF). $VIF_j(K)$ sebagai fungsi dari K merupakan unsur diagonal ke j dalam matriks

$$(X'X + K)^{-1}X'X(X'X + K)^{-1} \quad (3)$$

2. Metode Penelitian

Data yang digunakan yaitu data mengenai kebutuhan akan tenaga kerja pada 17 Rumah Sakit Angkatan Laut U.S dari tabel 13.3 dalam buku Bowerman dan O'Connell (1997). Tahapan penelitian diawali dengan melakukan analisis

regresi linear berganda pada data. Langkah selanjutnya adalah mendeteksi adanya multikolinieritas dengan melihat nilai korelasi antar peubah bebas dan nilai VIF. Apabila ditemukan adanya multikolinieritas pada data, maka dilakukan pendugaan parameter dengan *Generalized Ridge Regression*. Langkah pertama yang dilakukan adalah mentransformasi data melalui proses *centering* dan *scaling* data. Tahapan selanjutnya adalah proses ortogonalisasi pada peubah-peubah bebas. Setelah itu, ditentukan nilai K yang merupakan matriks diagonal dengan anggota (k_1, k_2, \dots, k_5) dan penduga parameter regresi *generalized ridge* dari peubah bebas ortogonal dengan menggunakan metode iteratif. Penduga awal untuk k_j yaitu $k_j^0 = \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}, j = 1, 2, \dots, 5$ dengan $\hat{\sigma}^2$ merupakan MSE dimana $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n-p-1}, \hat{y} = X\hat{\alpha}$ dan $\hat{\alpha}_j$ adalah penduga parameter regresi dari peubah bebas ortogonal yang merupakan solusi dari $y = Z\alpha + \varepsilon$. Pendugaan awal dari k_j digunakan untuk menghitung pendugaan awal *generalized ridge* untuk peubah bebas ortogonal $\hat{\alpha}_{GR,j}^0 = (\Lambda + K^0)^{-1}Z'y$, dimana $K^0 = \text{diag}(k_1^0, k_2^0, \dots, k_5^0)$. Pendugaan awal $\hat{\alpha}_{GR,j}^0$ kemudian digunakan untuk menghitung pendugaan k_j^1 . Nilai k_j^1 ini dapat digunakan untuk menghitung pendugaan dari $\hat{\alpha}_{GR,j}^1$ dan seterusnya. Proses iterasi dihentikan ketika $|(\hat{\alpha}'_{GR}\hat{\alpha}_{GR})^i - (\hat{\alpha}'_{GR}\hat{\alpha}_{GR})^{i-1}| \leq 0,01$. Setelah iterasi terhenti akan diperoleh nilai penduga koefisien *generalized ridge* dari peubah bebas ortogonal ($\hat{\alpha}_{GR}$). Langkah selanjutnya yaitu menentukan penduga koefisien *Generalized Ridge Regression*, dimana $\hat{\beta}_{GR} = T\hat{\alpha}_{GR}$. Langkah terakhir adalah memastikan bahwa sudah tidak terjadi multikolinieritas dengan melihat nilai VIF yang kemudian dilanjutkan dengan melakukan uji terhadap model secara simultan dengan uji F dan uji individu untuk koefisien regresi dengan uji t.

3. Hasil dan Pembahasan

Mendeteksi Adanya Multikolinieritas pada Model Regresi

Analisis regresi linear berganda dengan menggunakan metode kuadrat terkecil terhadap data kebutuhan akan tenaga kerja pada 17 Rumah Sakit Angkatan Laut U.S menghasilkan model regresi linear berganda yaitu $Y = 1.963 - 15,85X_1 + 0,0559X_2 + 1,59X_3 - 4,219X_4 - 394,3X_5$ dengan nilai koefisien determinasi (R^2) sebesar 0,987.

Untuk mendeteksi adanya multikolinieritas pada peubah bebas akan dilihat nilai korelasi antar peubah bebas dan nilai VIF dari masing-masing peubah bebas.

Tabel 1. Nilai Koefisien Korelasi

	X ₁	X ₂	X ₃	X ₄	X ₅
X ₁	1	0,907	1	0,936	0,671
X ₂		1	0,907	0,910	0,447
X ₃			1	0,933	0,671
X ₄				1	0,463
X ₅					1

Tabel 2. Nilai VIF Peubah Bebas

<i>Predictor</i>	VIF
X1	9.597,6
X2	7,9
X3	8.933,1
X4	23,3
X5	4,3

Dari tabel 1 terlihat bahwa korelasi antar peubah bebas cukup besar yaitu mendekati satu yang menunjukkan bahwa terjadi kolinearitas sangat kuat antar peubah bebas. Nilai VIF dari peubah bebas pada tabel 2 menunjukkan bahwa peubah bebas X₁, X₂, X₃, dan X₄ terlibat dalam masalah multikolinieritas karena memiliki nilai VIF yang lebih besar dari 5. Berdasarkan uraian-uraian tersebut, maka dapat dipastikan terjadi pelanggaran terhadap asumsi multikolinieritas. Oleh karena itu, diperlukan metode alternatif dalam mengatasi masalah multikolinieritas yaitu dengan *Generalized Ridge Regression*.

Penyelesaian Masalah Multikolinieritas dengan *Generalized Ridge Regression*

Pada *Generalized Ridge Regression*, nilai konstanta bias k_1, k_2, \dots, k_5 diperoleh melalui proses iterasi sampai ditemukan penduga koefisien regresi yang stabil. Iterasi berhenti pada iterasi kedua, karena pada iterasi ketiga nilai $(\Lambda + K)$ menjadi singular karena nilai k_1, k_2, \dots, k_5 yang ditambahkan cukup besar. Nilai konstanta bias k_1, k_2, \dots, k_5 yang diperoleh dari proses iterasi kedua yaitu 17,7499, $4,2348 \times 10^8$, $4,5454 \times 10^{18}$, $8,3003 \times 10^7$, $5,6253 \times 10^{19}$. Setelah itu dilakukan pendugaan koefisien regresi untuk *Generalized Ridge Regression*. Penduga koefisien regresi pada data terstandarisasi dan pada peubah awal beserta nilai VIF dari masing-masing peubah bebas dapat dilihat pada tabel 3.

Tabel 3. Penduga Koefisien Regresi untuk *Generalized Ridge Regression*

Independent Variable	Regression Coefficient	Standard Error	Standardized Regression Coefficient	VIF	t_0	$t_{0,025,11}$
Intercept	-1.420,48					
X_1	6,4929	2,9190	0,188	$3,4812 \times 10^{-3}$	2,3329	2,201
X_2	0,0459	0,0221	0,1755	$2,3258 \times 10^{-9}$	2,1777	2,201
X_3	0,213	0,0958	0,1879	$5,5987 \times 10^{-17}$	2,3315	2,201
X_4	9,1916	4,3544	0,1784	$2,2988 \times 10^{-7}$	2,2139	2,201
X_5	453,3054	296,7532	0,1291	$3,5448 \times 10^{-14}$	1,6021	2,201

Tabel 4. Analisis Ragam (ANOVA) *Generalized Ridge Regression*

Source	DF	SS	MS	F_0	$F_{(0,05,5,11)}$
Regression	5	69.657.543	13.931.509	4,3344	3,20
Residual Error	11	35.355.826	3.214.166		
Total	16	494.712.541			

Model regresi untuk metode *Generalized Ridge Regression* adalah $Y = -1.420,48 + 6,4929X_1 + 0,0459X_2 + 0,213X_3 + 9,1916X_4 + 453,3054X_5$. Nilai koefisien determinasi (R^2) dari model tersebut adalah 0,9913 dengan MSE sebesar 3.214.166. Metode *Generalized Ridge Regression* dapat mengatasi masalah multikolinieritas dengan baik, dapat dilihat pada tabel 3 dimana nilai VIF dari masing-masing peubah bebas lebih kecil dari 5 yang berarti peubah-peubah bebas yang terlibat dalam model sudah tidak terlibat masalah multikolinieritas.

Pengujian model secara simultan dengan menggunakan uji F menghasilkan nilai F_0 yang lebih besar daripada $F_{(0,05,5,11)}$ ($4,3344 > 3,20$) sehingga mengakibatkan H_0 ditolak, maka dapat disimpulkan bahwa peubah bebas $X_1 \dots X_5$ secara simultan mempunyai kontribusi yang signifikan pada model. Untuk mengetahui signifikan tidaknya pengaruh masing-masing peubah bebas secara individu terhadap peubah tak bebas maka dilakukan uji t. Nilai t_0 dari masing-masing peubah bebas dapat dilihat pada tabel 3. Sesuai dengan kaidah keputusan, jika $|t_0| > t_{0,025,11}$ maka tolak H_0 , maka dari lima peubah bebas, terdapat dua peubah bebas yang tidak signifikan yaitu X_2 dan X_5 .

4. Kesimpulan dan Saran

Pada data yang mengalami masalah multikolinieritas, metode kuadrat terkecil (*Ordinary Least Square*) tidak dapat melakukan pendugaan koefisien regresi dengan tepat. Metode *Generalized Ridge Regression* merupakan salah satu metode alternatif yang dapat mengatasi masalah multikolinieritas dengan sangat baik, dibuktikan dari nilai VIF dari masing-masing peubah bebas yang lebih kecil dari 5.

Metode *Generalized Ridge Regression* pada penelitian ini menggunakan pendekatan iteratif yang diusulkan oleh Hoerl dan Kennard (1970a) untuk menentukan nilai K dan penduga koefisien regresinya. Pada penelitian selanjutnya pendugaan koefisien regresi dapat menggunakan pendekatan non-iteratif, yaitu

solusi eksplisit dari *Generalized Ridge Regression* yang diusulkan oleh Hemmerle (1975). Selain metode *Generalized Ridge Regression* yang mengatasi masalah multikolinearitas dengan lebih menekankan pada pengurangan ragam sampel, dapat pula dilakukan penelitian dengan menggunakan metode *Jacknife Ridge Regression* yang pertama kali diusulkan oleh Hinkley (1977) dimana metode ini mengatasi masalah multikolinearitas dengan lebih menekankan pada pengurangan bias pada penduga ridge.

Daftar Pustaka

- [1] Draper, N.R and H. Smith. 1998. *Applied Regression Analysis, Third Edition*. New York: John Wiley and Sons, Inc.
- [2] Hoerl, A.E. and R.W. Kennard. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". *Technometrics*, 12 (1970a), 55-67. (Republished in *Technometrics*. 42 (2000), 80-86). http://scholar.google.co.id/scholar_url?hl=id&q=http://citeseerx.ist.psu.edu/viewdoc/download%3Fdoi%3D10.1.1.157.3796%26rep%3Drep1%26type%3Dpdf&sa=X&scisig=AAGBfm2FYmMwataOUaFGiDv2ODGgB6z69g&oi=scholarOJS20110100005_36910681. Diakses tanggal 9 Desember 2011.
- [3] Hoerl, A.E. and R.W. Kennard. 1970. "Ridge Regression: Applications to Nonorthogonal Problems". *Technometrics*, Vol. 12, No. 1. (Feb., 1970b), pp. 69-82. <http://statgen.ucr.edu/download/course/STAT288/hoerl70b.pdf>. Diakses tanggal 14 Juni 2012.
- [4] Montgomery, D.C. and E.A. Peck. 1991. *Introduction to Linear Regression Analysis, Second Edition*. New York: John Wiley and Sons, Inc.