Metode *Jackknife Ridge Regression* dalam Mengatasi Masalah Multikolinearitas

M. Addienul Haq[1], Joko Purwadi[2]
[1]*Program Studi Matematika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan*
Email: muhammadaddien16@gmail.com
[2]*Program Studi Matematika, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan*

Submitted :........................ Reviewed :.......................... Accepted:.....................

ABSTRAK

Pada analisis regresi berganda, terdapat beberapa asumsi-asumsi klasik yang harus dipenuhi, salah satunya adalah tidak adanya multikolinearitas. Metode yang digunakan untuk mengatasi multikolinearitas diantaranya yaitu *Jackknife Ridge Regression* (JRR). Metode JRR merupakan pengembangan dari metode *generalized ridge regression* dengan menekankan pengurangan bias pada estimator ridge dan bias yang dihasilkan tidak dijamin selalu bernilai kecil. Tujuan dari penelitian ini adalah mengetahui hasil estimasi parameter-parameter model regresi linear berganda dengan metode *Jackknife ridge regression* yang diterapkan pada data IPM di Provinsi Jawa Tengah. Hasil penelitian adalah diperolehnya hasil terbaik dalam mengatasi masalah multikolinearitas yaitu dengan metode JRR karena dapat menghasilkan nilai mean square erorr(MSE) dan variansi lebih kecil dari metode GRR.

Kata kunci : *Jackknife Ridge Regression, Generalized Ridge Regression,* Multikolinearitas

PENDAHULUAN

Analisis regresi merupakan analisis statistika yang memanfaatkan hubungan antara dua atau lebih variabel sehingga salah satu variabel dapat ditentukan dari variabel lainnya. Variabel tersebut terdiri dari variabel yang dijelaskan atau yang disebut variabel terikat (Y) dan variabel penjelas atau yang disebut variabel bebas (X). Dalam pembentukan model regresi dilakukan estimasi terhadap parameter regresi dalam model untuk menghasilkan estimasi terbaik menggunakan metode kuadrat terkecil. Metode kuadrat terkecil merupakan metode yang digunakan untuk mengestimasi $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ dengan cara meminimumkan jumlah kuadrat galat (Montgomery). Pada analisis regresi berganda, terdapat beberapa asumsi klasik yang harus terpenuhi, salah satunya adalah tidak adanya multikolinearitas.

Multikolinearitas merupakan suatu keadaan terjadinya hubungan antara variabel-variabel bebas dalam model regresi yang dapat mengakibatkan koefisien regresi yang dihasilkan menjadi sangat lemah atau tidak dapat memberi hasil analisis yang mewakili sifat atau pengaruh dari variabel bebas yang bersangkutan. Untuk mengetahui ada tidaknya multikolinearitas dapat menggunakan nilai *Variance Inflation Factor* (VIF). Jika nilai VIF >10 maka terdapat multikolinearitas pada variabel bebas dalam model regresi. Ada berbagai metode yang digunakan untuk mengatasi multikolinearitas yaitu *ridge regression* dan *generalized ridge regression*. Pada setiap metode memiliki kekurangan dan kelebihan namun setiap metode dapat digunakan untuk mengatasi masalah multikolinearitas.

Pada penelitian ini, penulis menggunakan metode *Jackknife Ridge Regression* dalam mengatasi masalah multikolinearitas. *Jackknife Ridge Regression* yaitu pengembangan dari metode *generalized ridge regression* dengan menekankan pegurangan bias pada estimator ridge dan bias yang dihasilkan tidak dijamin selalu bernilai kecil. Sedemikian sehingga, dengan menggunakan metode ini, bias dari estimator generalized ridge akan tereduksi. Tujuan dari penelitian ini adalah mengetahui hasil estimasi parameter-parameter model regresi linear berganda dengan metode *Jackknife Ridge Regression* yang diterapkan pada data Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Tengah dan untuk mengetahui kelayakan model yang dihasilkan *Jackknife Ridge Regression* dengan melihat nilai MSE dan variansi yang dihasilkan pada model.

METODE PENELITIAN

Alat dan Bahan

Dalam penelitian alat dan bahan yang dibutuhkan yaitu literatur yang dikumpulkan berupa buku, skripsi, artikel, jurnal atau modul yang dapat mendukung penelitian.

Jalannya Penelitian

Sumber data pada penelitian ini adalah berupa data sekunder. Data yang digunakan dalam penelitian ini yaitu data Indeks Pembangunan Manusi(IPM) yang diakses di website resmi BPS Provinsi Jawa Tengah.

Tahapan penelitian yang akan dilakukan penulis dalam penelitian ini adalah sebagai barikut:

1. Mencari data IPM di website BPS Provinsi Jawa Tengah.
2. Melakukan Analisis regresi linear berganda pada data.
3. Membakukan data.
4. Melakukan proses orthogonalisasi pada variabel-variabel bebas.
5. Menentukan nilai *K*.
6. Menentukan estimasi awal *jackknife ridge regression.*
7. Proses iterasi menguji model *jackknife ridge regression.*
8. Memastikan data sudah tidak ada multikolinearitas.
9. Menguji kelayakan model yang dihasilkan *jackknife ridge regression.*

HASIL DAN PEMBAHASAN

1. Mendeteksi Adanya Multikolinearitas pada Model Regresi

   Langkah awal yang dilakukan yaitu membuat model regresi linear berganda menggunakan metode kuadrat terkecil terhadap data Indeks Pembangunan Manusia (IPM) yang ditunjukan pada tabel estimasi koefisien regresi (Tabel 1)

Tabel 1. Estimasi Koefisien Regresi

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0,001511466	2,178913307	2,744925	0,011044
X_1	0,4916398	0,027438061	16,23679	8,64E-15
X_2	0,9733792	0,116579173	7,917709	2,84E-08
X_3	1,252104	0,098688414	13,84605	3,15E-13

X_4	-0,000929127	0,008361671	-0,93023	0,361147
X_5	0,02499962	0,014757756	0,152519	0,880002
X_6	0,000958662	3,83363E-05	25,73234	1,65E-19
X_7	0,003285702	0,013317545	-0,51773	0,6092
X_8	5,82E-07	2,60968E-05	-1,0668	0,296256

Hasil analisis menggunakan metode kuadrat terkecil terhadap data IPM menghasilkan model regresi yaitu:

$$
\begin{aligned}
Y = 0.001511466 + 0.4916398X_1 + 0.9733792X_2 + 1.252104X_3 \\
- 0.000929127X_4 + 0.02499962X_5 + 0.000958662X_6 \\
+ 0.003285702X_7 + 0.0000005822412X_8
\end{aligned}
$$

Setelah mendapatkan model regresi, langkah selanjutnya yaitu uji kecocokan model regresi secara overall dengan melakukan uji F. Dalam melakukan uji F hipotesis yang digunakan sebagai berikut:

$H_0 :\ \beta_1 = \beta_2 = \cdots = \beta_8 = 0\ \ (model\ regresi\ tidak\ signifikan)$

$H_0 :\ \beta_1 \neq 0\ \ (model\ regresi\ signifikan)$

Tabel 2. Analisis Ragam

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	694,7199239	86,83999	2105,513	2,17617E-33
Residual	25	1,031102521	0,041244		
Total	33	695,7510265			

dengan keputusan tolak H_0 apabila $F_{hitung} > F_{tabel}$. Dari Tabel 2 didapatkan hasil $F_{hitung} = 2105{,}513$ dan nilai $F_{tabel} = 2{,}320$ maka dapat disimpulkan bahwa H_0 ditolak artinya variabel-variabel bebas secara overall signifikan terhadap variabel terikat.

Langkah selanjutnya adalah melakukan pengujian secara parsial (uji t) yang bertujuan untuk mengetahui signifikan atau tidaknya pengaruh masing-masing variabel bebas terhadap variabel terikat. Dalam melakukan uji t hipotesis yang digunakan adalah:

$H_0 : \beta_i = 0, i = 1,2, \ldots, n$ (tidak ada pengaruh masing-masing variabel bebas terhadap variabel terikat)

$H_0 : \beta_i \neq 0, i = 1,2, \ldots, n$ (ada pengaruh masing-masing variabel bebas terhadap variabel terikat)

dengan keputusan tolak H_0 apabila $\left|t_{hitung}\right| > t_{\left(\frac{\alpha}{2};n-p-1\right)}$. Dari Tabel 1 didapatkan nilai t_{hitung} didapatkan hasil untuk masing-masing variabel bebasnya. Dengan nilai $t_{\left(\frac{\alpha}{2};n-p-1\right)}$ adalah 2,378 dapat disimpulkan bahwa hanya X_1, X_2, X_3 dan X_6 yang signifikan secara statistik terhadap variabel Y sedangkan variabel bebas lain tidak.

Pada hasil analisis regresi linear berganda diperoleh nilai *adjusted* R^2 yang besar yaitu 0,998043761 yang artinya 99,8% variasi IPM dapat dijelaskan oleh variabel-variabel bebas yang telah ditentukan, tetapi tidak diikuti dengan hasil hipotesis yang berpengaruh signifikan dari masing-masing koefisien regresi. Hal ini mengindikasikan adanya penyimpangan yang terjadi pada model regresi.

Penyimpangan-penyimpangan diatas diakibatkan oleh koefisien regresi yang tidak dapat memberikan hasil analisis yang mewakili pengaruh dari variabel bebas yang bersangkutan. Hal tersebut mengindikasikan adanya hubungan korelasi yang kuat antar variabel bebasnya (multikolinearitas). Untuk mengetahui adanya multikolinearitas pada variabel bebas akan dilakukan analisis terhadap nilai koefisien korelasi antar variabel bebas dan nilai VIF dari masing-masing variabel bebas.

Tabel 3. Koefisien korelasi antar variabel bebas

	X1	X2	X3	X4	X5	X6	X7	X8
X1	1							
X2	0,4893	1						
X3	0,3298	0,9191	1					
X4	-0,0841	0,1933	0,1367	1				
X5	0,4554	0,2054	0,0508	-0,0659	1			
X6	0,1176	0,4963	0,3512	0,2599	0,4636	1		
X7	0,0907	-0,141	-0,193	-0,0824	-0,0787	-0,0458	1	
X8	0,3202	0,3043	0,028	0,1569	0,2171	0,1846	-0,2885	1

Tabel 4. Nilai VIF Variabel Bebas

Prediktor	X1	X2	X3	X4	X5	X6	X7	X8
VIF	2,4036	8,836	12,2473	1,2623	1,6002	3,516	1,2282	3,3306

Nilai VIF pada Tabel 4 menunjukan terdapat masalah multikolinearitas pada Variabel bebas X_3 karena nilai VIF lebih dari 10. Dari hasil uraian diatas maka dapat disimpulkan bahwa model mengandung multikolinearitas. Untuk mengatasi masalah tersebut diperlukan metode *jackknife ridge regression*.

2. Proses Transformasi dengan Metode *Centering and Rescaling*

   Metode *centering and rescaling* merupakan salah satu bentuk transformasi data yang bertujuan mempermudah analisis data yang memiliki satuan yang berbeda-beda. Dalam hal ini yang akan dibakukan adalah model regresi linear berganda. Proses penskalaan adalah rata-rata dibagi dengan standar deviasi.

$$Y_i^* = \frac{(Y_i - \bar{Y})}{\sqrt{S_y}}$$

$$X_{ji}^* = \frac{\left(X_{ji} - \bar{X}_J\right)}{\sqrt{S_{X_j}}}$$

Pada proses *centering*(pemusatan) dapat mengakibatkan hilangnya β_0 yang membuat perhitungan untuk mencari model regresi lebih sederhana.

3. Penyelesaian Masalah Multikolinearitas dengan *Jackknife Ridge Regression*

Pada analisis regresi menggunakan *jackknife ridge regression* data yang digunakan adalah data yang sudah mengalami proses transformasi. Setelah data ditransformasi menggunakan *centering and rescaling,* dilakukan analisis regresi linear berganda terhadap data kemudian dicari nilai eigen baru dari hasil transformasi dan dibentuk $Z = XT$ dengan T merupakan vektor eigen dari matriks $X'X$. Kemudian dilakukan estimasi dengan menggunakan metode kuadrat terkecil sebagai penduga koefisien dari *Z* yang merupakan estimasi γ_{LS} dari koefisien *X*.

Untuk mengatasi masalah multikolinearitas pada metode *jackknife ridge regression*, kemudian dicari nilai tetapan bias *k* yang kemudian digunakan dalam menghitung nilai $K = kI$ dengan K adalah matriks diagonal yang akan ditambahkan pada diagonal matriks Δ. Nilai konstanta bias *k* yang diperoleh yaitu 0,05057904.

Langkah selanjutnya mensubstitusikan nilai K kedalam persamaan $\hat{\gamma}_{JR} = (I - (A^{-1}K)^2)\gamma$ dan diperoleh estimasi parameter untuk metode *jackknife ridge regression.*

Tabel 5. Estimasi Parameter, MSE dan Variansi Metode JRR

Parameter	Jackknife Ridge Regression
γ_1	0,485265785

γ_2	0,062155136
γ_3	0,072265908
γ_4	0,068328042
γ_5	0,086731125
γ_6	-0,230501016
γ_7	-0,009157629
γ_8	-0,182828397
MSE	0,001930592
Variansi	0,001901238

Selanjutnya diperoleh persamaan regresi ridge menggunakan dengan jackknife ridge regression

$$\begin{aligned} Y = \; & 0{,}4852657785Z_1 + 0{,}062155136Z_2 + 0{,}072265908Z_3 \\ & + 0{,}068328042Z_4 + 0{,}086731125Z_5 - 0{,}230501016Z_6 \\ & - 0{,}009157629Z_7 - 0{,}182828397Z_8 \end{aligned}$$

Atau dapat dikonversikan kedalam bentuk regresi yang asli kedalam bentuk regresi yang asli dengan mengalikan estimasi parameter jackknife ridge regression dengan matriks dan diperoleh persamaan:

$$\begin{aligned} Y = \; & 0{,}1935204X_1 + 0{,}1807551662X_2 + 0{,}371639469X_3 - 0.007976249X_4 \\ & + 0.00224565X_5 + 0.37100329X_6 - 0{,}005267081X_7 \\ & - 0{,}015460064X_8 \end{aligned}$$

4. Transformasi ke Bentuk Awal

Selanjutnya dikembalikan kedalam bentyuk awal dengan menggunakan transformasi persamaan regresi menggunakan rumus:

$$\beta_i = \frac{S_y}{S_i}\beta_i^*$$

$$\beta_0 = Y - \beta_1 X_1 - \beta_2 X_2 - \cdots - \beta_p X_p$$

Jika ditulis dalam persamaan maka model awal yang terbentuk dengan *jackknife ridge regression* adalah

$$Y = 6.058846 + 0.4442277X_1 + 0.9230317X_2 + 1.365234X_3 - 0.007729201X_4 + 0.003365829X_5 + 0.0009872812X_6 - 0.008110791X_7 - 0.00002879868X_8$$

5. Uji Signifikansi Model Regresi

Tabel 6. Uji Signifikansi Uji *t*

Prediktor	β_i	S_{β_i}	t_{hitung}	t_{tabel}
X1	0.4442277	0.003849908	115,3865757	2.378
X2	0.9230317	0.00748166	123,4285594	2.378
X3	1,365234	0.002747998	496,6631773	2.378
X4	-0.007729201	0.001937414	-3,991156132	2.378
X5	0.003365829	0.000766214	4,403726603	2.378
X6	0.0009872812	7,43E+00	13,28265639	2.378
X7	-0.008110791	0.000901736	-8,9920048	2.378
X8	-0.00002879868	1,66E-01	-17,30003724	2.378

$H_0 : \beta_i = 0, i = 1,2, \dots, n$ (tidak ada pengaruh masing-masing variabel bebas terhadap variabel terikat)

$H_0 : \beta_i \neq 0, i = 1,2, \dots, n$ (ada pengaruh masing-masing variabel bebas terhadap variabel terikat)

dengan keputusan tolak H_0 apabila $|t_{hitung}| > t_{\left(\frac{\alpha}{2};n-p-1\right)}$ yang artinya jika H_0 ditolak maka parameter layak digunakan dalam model regresi.

Berdasarkan kriteria uji diatas maka dapat disimpulkan bahwa parameter layak digunakan dalam model regresi.

KESIMPULAN

Berdasarkan pembahasan pada Bab 4 terdapat kesimpulan yaitu:

1. Bentuk estimasi model regresi linear berganda menggunakan metode *jackknife ridge regression* adalah

$$\hat{\gamma}_{JR} = (I - (A^{-1}K)^2)\gamma$$

dan bentuk bias parameter $\hat{\gamma}_{JR}$ adalah

$$Bias\ \hat{\gamma}_{JRR} = -K^2A^{-2}\gamma$$

2. Hasil estimasi parameter-parameter model regresi linear berganda dengan metode *jackknife ridge regression* yang diterapkan pada data IPM di Provinsi Jawa Tengah yaitu:

$$\begin{aligned} Y = 6.058846 &+ 0.4442277X_1 + 0.9230317X_2 + 1.365234X_3 \\ &- 0.007729201X_4 + 0.003365829X_5 + 0.0009872812X_6 \\ &- 0.008110791X_7 - 0.00002879868X_8 \end{aligned}$$

DAFTAR PUSTAKA

Badan Pusat Statistik Provinsi Jawa Tengah (BPS), diakses dari https://jateng.bps.go.id/

Gujarati. 2004. Basic Econometrics Fourth Edition. Singapura: McGraw Hill.

Khurana, M., Chaubey, Y.P., dan Chandra, S..2012. Jackknifing The Ridge Regression Estimator: A Revisit. Marcel Dekker, Inc: America.

Maddala, G.S. 1992. Introduction Econometrics Second Edition. New York: MacmillanPublishingCompany.