

Regresi Ridge Parsial Untuk Data Yang Mengandung Masalah Multikolinearitas

Ghaida Azzahra*, Nar Herrhyanto dan Fitriani Agustina

Departemen Pendidikan Matematika

Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia

*Surel : ghaida.ghazi@gmail.com

ABSTRAK. Multikolinearitas merupakan kondisi dimana terdapat hubungan linear antar beberapa atau semua variabel bebas dari model regresi linear berganda. Multikolinearitas dapat mengakibatkan taksiran parameter koefisien regresi dari variabel-variabel tersebut tidak dapat ditaksir dengan pasti. Regresi ridge dapat mengatasi permasalahan multikolinearitas, namun terdapat keterbatasan dalam regresi ridge yaitu konstanta bias c diterapkan untuk semua variabel bebas terlepas dari tinggi rendahnya tingkat multikolinearitas. Artikel ini membahas mengenai pengembangan metode regresi ridge yang menambahkan konstanta bias c untuk variabel yang mengandung kolinearitas tinggi. Pengembangan dari regresi ridge ini dinamakan regresi ridge parsial. Metode regresi ridge parsial ini diterapkan pada studi kasus angka kematian bayi di Provinsi Jawa Timur tahun 2018. Hasil penelitian menunjukkan adanya multikolinearitas pada data studi kasus kematian bayi tersebut. Selain itu diperoleh informasi bahwa angka kematian bayi di Provinsi Jawa Timur tahun 2018 hanya dipengaruhi oleh variabel berat badan lahir rendah, dimana setiap satu kenaikan bayi yang memiliki berat badan lahir rendah, akan meningkatkan angka kematian bayi sebesar 0,0953.

Kata kunci : Multikolinearitas, regresi ridge, regresi ridge parsial, angka kematian bayi

TITLE. *Partial ridge regression for data that contains multicollinearity problems*

ABSTRACT. *Multicollinearity is a condition where there is a linear relationship between some or all of the independent variables from multiple linear regression models. Multicollinearity can result in the parameter estimates of the regression coefficients of these variables that cannot be estimated with certainty. Ridge regression can solve multicollinearity problems, but there are limitations in ridge regression, namely the bias constant c is applied to all independent variables*

regardless of the level of multicollinearity. This article discusses the development of a ridge regression method that adds a bias constant c for variables containing high collinearity. The development of this ridge regression is called partial ridge regression. This partial ridge regression method was applied to a case study of infant mortality in East Java Province in 2018. The results showed that there was multicollinearity in the case study data of infant mortality. In addition, information was obtained that the infant mortality rate in East Java Province in 2018 was only influenced by the variable low birth weight, where every one increase in babies who have low birth weight, will increase the infant mortality rate by 0.0953.

Keywords : *Multicollinearity, ridge regression, partial ridge, regression, infant mortality rate.*

1. PENDAHULUAN

Analisis regresi merupakan sebuah studi mengenai hubungan antara satu variabel tak bebas (*regressand variable*) dengan satu atau lebih variabel bebas (*regressor variable*) [3]. Hubungan yang diperoleh pada analisis regresi umumnya dinyatakan dalam bentuk sebuah persamaan matematik [7]. Dengan demikian, analisis regresi adalah suatu teknik analisis data untuk mengetahui hubungan fungsional antara satu variabel tak bebas dengan satu atau lebih variabel bebas yang dinyatakan dalam sebuah persamaan matematik. Pada umumnya, variabel tak bebas dinotasikan dengan Y dan untuk satu atau lebih variabel bebas dinotasikan dengan X_1, X_2, \dots, X_k . Berdasarkan jumlah variabel bebasnya terdapat dua macam analisis regresi linear, yaitu regresi linear sederhana (*simple linear regression*) dan regresi linear berganda (*multiple linear regression*). Regresi linear sederhana merupakan teknik analisis data untuk mengetahui hubungan fungsional antara variabel tak bebas dengan satu variabel bebas, sedangkan regresi linear berganda merupakan teknik analisis data untuk mengetahui hubungan fungsional antara variabel tak bebas dengan dua atau lebih variabel bebas.

Penaksiran parameter pada model regresi linear sederhana dan regresi linear berganda menggunakan metode *Ordinary Least Square* (OLS). Penaksiran parameter dengan metode OLS dilakukan dengan cara meminimalkan jumlah kuadrat kesalahan (*error*). Penaksir OLS merupakan penaksir tak bias linear yang paling baik atau disebut dengan *Best Linear Unbiased Estimator* (BLUE). penaksir OLS memiliki varians yang minimum dibandingkan dengan penaksir tak bias linear lainnya, sehingga parameter dari model regresi dapat ditaksir secara lebih akurat. Secara garis besar dalam analisis regresi terdapat asumsi-asumsi yang harus dipenuhi agar penaksir analisis regresi dapat ditaksir dengan metode OLS, yaitu : (1) Normalitas, (2) Non-autokorelasi, (3) Homoskedastisitas, dan (4) Tidak terjadi multikolinearitas [3].

Multikolinearitas merupakan kondisi dimana terdapat hubungan linear yang nyata antar beberapa atau semua variabel bebas dari model regresi linear berganda [3]. Terdapat dua jenis multikolinearitas, yaitu multikolinearitas sempurna dan multikolinearitas tidak sempurna. Multikolinearitas sempurna terjadi, apabila suatu variabel bebas bergantung sepenuhnya pada variabel yang lainnya [6]. Apabila terjadi pelanggaran pada asumsi tidak terjadi multikolinearitas, maka taksiran parameter koefisien regresi tidak dapat ditaksir dengan pasti. Dengan demikian, nilai taksiran parameter koefisien regresinya akan menjadi bias. Gejala kondisi multikolinearitas dapat dideteksi dengan menggunakan (1) Nilai *Variance*

Inflation Factor (VIF) dan *Tolerance* (TOL), dan (2) *Eigenvalues* dan *Condition Index* (CI). Terdapat beberapa cara untuk mengatasi multikolinearitas yaitu informasi apriori, menggabungkan data *cross-sectional* dan data *time series*, mengeluarkan satu dari variabel yang berkolinear, transformasi variabel, penambahan data baru, mengurangi kolinearitas pada regresi polinomial, analisis komponen utama dan regresi ridge [3].

Regresi ridge adalah satu dari berbagai teknik analisis data untuk mengatasi permasalahan multikolinearitas pada datanya yang menghasilkan penaksir bias dari koefisien regresi [5]. Regresi ridge pada dasarnya merupakan metode OLS yang dimodifikasi dengan menambahkan sebuah konstanta bias c pada matriks korelasi ($X'X$) dan variabel bebasnya ditransformasikan dengan prosedur *centering* dan *rescaling* (pemusatan dan penskalaan). Pemilihan konstanta bias c merupakan hal yang harus diperhatikan dalam regresi ridge. Konstanta bias yang diharapkan adalah konstanta bias yang menghasilkan bias relatif kecil dan menghasilkan koefisien regresi yang relatif stabil. Suatu cara yang dapat digunakan dalam memilih konstanta bias c adalah dengan menggunakan nilai VIF dan *ridge trace* (jejak gulud).

Pada tahun 2014 T.L Wasilaine dkk melakukan penelitian mengenai regresi ridge, penelitian tersebut menghasilkan sebuah model regresi yang diperoleh dengan menambahkan konstanta bias c pada diagonal matriks $X^T X$ [8]. Pada tahun yang sama Hasriani melakukan penelitian yang berkaitan dengan perbandingan regresi ridge dengan principal component analysis dalam mengatasi masalah multikolinearitas, hasil dari penelitian ini adalah metode regresi ridge lebih baik dari analisis komponen utama yang dilihat dari nilai *mean square error* dan nilai koefisien determinasinya [4].

Pendekatan regresi ridge belum memberikan hasil yang benar-benar memuaskan dalam mengatasi masalah data yang mengandung multikolinearitas. Terdapat sejumlah masalah atau keterbatasan dalam regresi ridge, antara lain yaitu konstanta bias diterapkan untuk semua variabel bebas terlepas dari tinggi rendahnya tingkat multikolinearitas dan pemilihan konstanta bias c yang bersifat subjektif dan sembarang. Untuk mengatasi keterbatasan itu C.K. Chandrasekhar dkk mengusulkan sebuah pendekatan penyesuaian secara selektif konstanta ridge terkait dengan tingginya variabel kolinear untuk mengontrol ketidakstabilan varians dari estimasi koefisien, yang dinamakan regresi ridge [1]. Untuk memilih konstanta bias dari variabel-variabel kolinear, regresi ridge parsial menggunakan pendekatan *Singular Value Decomposition* (SVD). Artikel ini membahas mengenai

prosedur penggunaan regresi ridge parsial disertai penerapannya pada kasus angka kematian bayi di Provinsi Jawa Timur tahun 2018.

2. METODOLOGI

Teknik analisis data yang digunakan pada penelitian ini yaitu regresi ridge parsial. Regresi ridge parsial ini digunakan untuk menentukan model regresi dari data yang mengandung multikolinearitas. Penelitian ini terlebih dahulu melakukan studi literatur yaitu dengan cara mencari referensi teori yang berkaitan dengan regresi ridge dan regresi ridge parsial yang tersedia di berbagai sumber, seperti buku, jurnal, dan penelitian-penelitian sebelumnya yang membahas mengenai regresi ridge dan regresi ridge parsial. Setelah memperoleh teori-teori dari studi literatur, langkah selanjutnya yaitu menerapkan teori tersebut ke dalam sebuah kasus.

2.1. Regresi Ridge

Regresi ridge merupakan metode regresi yang digunakan untuk data yang mengandung multikolinearitas. Estimasi parameter pada regresi ridge dilakukan dengan menambahkan sebuah konstanta bias c pada diagonal matriks $X'X$, sehingga diperoleh sebagai berikut:

$$(X'X + cI)\hat{\beta}_R = X'Y$$

$$\hat{\beta}_R = (X'X + cI)^{-1}X'Y$$

dimana $0 \leq c \leq 1$. Suatu hal yang harus diperhatikan dalam regresi ridge adalah pada proses penentuan konstanta bias c . Pemilihan konstanta bias c dapat dilakukan dengan dua pendekatan yaitu dengan menggunakan plot *ridge trace* atau dengan menggunakan pendekatan nilai VIF.

Plot *ridge trace* merupakan plot yang menggambarkan pola dari estimator koefisien regresi dengan berbagai kemungkinan nilai konstanta bias c yang berbeda-beda. Konstanta bias c ditambahkan kedalam diagonal matriks korelasi, sehingga dengan berbagai nilai c akan mengakibatkan nilai *Variance Inflation Factor* (VIF) mendekati 1. Demikian sehingga dengan menggunakan plot *ridge trace* dapat diperoleh estimator koefisien regresi yang relatif lebih stabil.

2.2. Regresi Ridge Parsial

Regresi ridge parsial merupakan pengembangan dari regresi ridge. Regresi ridge parsial pertama kali dikembangkan oleh Chandrasekhar dkk pada tahun 2016 dalam artikel jurnal yang berjudul “*Partial Ridge Regression Under Multicollinearity*” [1]. Regresi ridge parsial

dikembangkan untuk mengatasi keterbatasan dari regresi ridge. Pada model regresi ridge, konstanta bias c ditambahkan ke semua variabel bebas terlepas dari tinggi rendahnya tingkat kolinearitas diantara variabel-variabel, sedangkan pada regresi ridge parsial konstanta bias c hanya ditambahkan untuk variabel bebas yang mengandung tingkat kolinearitas tinggi.

Dasar dari pendekatan regresi ridge parsial adalah dekomposisi dari matriks $X'X$ dan pemilihan konstanta bias c . Maksudnya yaitu Matriks $X'X$ yang berukuran $k \times k$ didekomposisi menjadi matriks $U\Delta U'$, dimana Δ adalah matriks diagonal dengan elemen diagonalnya adalah nilai-nilai eigen ($\lambda_j, j = 1, 2, \dots, k$) dari matriks $X'X$. Apabila nilai eigen λ_j dari matriks $X'X$ mendekati nol, maka hal tersebut mengindikasikan bahwa adanya hubungan linear pada variabel regressor. Nilai-nilai eigen λ_1 sampai λ_k pada elemen diagonal matriks $X'X$ disusun berurut dari nilai terbesar hingga nilai terkecil ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$). Tingginya variabel yang mengandung kolinearitas berkaitan dengan rendahnya nilai eigen. Perbandingan nilai eigen $\delta = \frac{\lambda_j}{\lambda_1}$ dapat digunakan untuk mengukur tinggi rendahnya tingkat kolinearitas antar variabel. Terdapat kesepakatan umum bahwa jika nilai δ kurang dari 0,001 mengindikasikan bahwa variabel-variabel tersebut memiliki kolinearitas yang tinggi [1].

Dengan menggunakan pendekatan *singular value decomposition* (SVD) dimana $X'X = U\Delta U'$ maka estimator parameter koefisien regresi ditentukan menggunakan perumusan sebagai berikut :

$$\hat{\beta}_{RP} = (U\Delta U' + cI)^{-1} X'Y$$

$$\hat{\beta}_{RP} = \left(U \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_k \end{pmatrix} U' + \begin{pmatrix} c_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & c_k \end{pmatrix} \right)^{-1} X'Y$$

Pada model regresi ridge konstanta bias c ditambahkan untuk semua variabel, sehingga untuk $c_1 = c_2 = \dots = c_k = c$. Sedangkan pada model regresi ridge parsial, konstanta bias c ditambahkan hanya untuk variabel yang mengandung kolinearitas tinggi atau dapat dikatakan variabel yang memiliki nilai eigen yang rendah dan kurang dari δ . Untuk variabel yang kolinearitasnya rendah konstanta bias c menjadi 0 ($c_1 = c_2 = \dots = c_j = 0$), sedangkan untuk variabel yang mengandung kolinearitas tinggi adalah $c_{j+1} = c_{j+2} = \dots = c_k = c$.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini digunakan data angka kematian bayi di Provinsi Jawa Timur pada tahun 2018. Data tersebut merupakan data sekunder yang diperoleh dari publikasi profil kesehatan Provinsi Jawa Timur tahun 2018 dengan jumlah Kabupaten/Kota sebanyak 38 Kabupaten/Kota. Data dapat diakses melalui website *dinkes.jatimprov.go.id* [2].

Variabel yang terdapat pada penelitian ini meliputi satu variabel *regressand* (Y) yaitu angka kematian bayi, dan meliputi lima variabel *regressor* (X) yang terdiri dari jumlah penderita pneumonia yang ditemukan dan ditangani (X_1), jumlah bayi dengan berat badan lahir rendah (X_2), cakupan kasus gizi buruk (X_3), cakupan pemberian vitamin A pada bayi (X_4), dan cakupan pelayanan kesehatan bayi (X_5).

3.1. Pendeteksian Multikolinearitas

Pada subbab ini akan memaparkan mengenai cara mendeteksi multikolinearitas. Pendeteksian multikolinearitas dapat dilakukan dengan menggunakan pendekatan nilai *Variance Inflation Factor* (VIF) dan *Tolerance* (TOL). Penentuan nilai VIF dan nilai TOL dilakukan dengan menggunakan bantuan program *Number Cruncher Statistical System* (NCSS). Nilai VIF dan nilai TOL yang diperoleh dari hasil pengolahan dengan bantuan program NCSS dapat dilihat pada Tabel 1.

Tabel 1. Nilai VIF dan Nilai TOL dari Variabel Bebas Penelitian

Variabel	VIF	TOL
X_1	2,4941	0,4010
X_2	3,7456	0,2670
X_3	1,8684	0,5352
X_4	170,4536	0,0059
X_5	156,5581	0,0064

Berdasarkan Tabel 1 diperoleh informasi bahwa variabel cakupan pemberian vitamin A pada bayi (X_2) dan variabel cakupan pelayanan kesehatan bayi (X_5) mempunyai nilai VIF yang melebihi 10 dan nilai TOL yang kurang dari 0,10; dengan demikian dapat dikatakan bahwa pada data penelitian terjadi masalah multikolinearitas.

3.2. Transformasi Data

Proses transformasi data dilakukan dengan tujuan untuk menghilangkan. Hal ini akan memudahkan dalam perhitungan untuk mencari model regresi ridge parsial. Transformasi data dilakukan dengan prosedur

centering (pemusatan) dan *scaling* (penskalaan). Hasil transformasi data terdapat pada Tabel 2.

Tabel 2, Hasil Transformasi Data

Y*	Z₁	Z₂	Z₃	Z₄	Z₅
-0,1364	-0,0948	-0,0905	0,0167	-0,1457	-0,1377
0,0777	0,0159	-0,0459	0,1393	-0,0698	-0,0693
-0,1556	-0,0774	-0,0241	-0,0863	-0,0983	-0,1027
0,0009	-0,0768	-0,0682	-0,0780	-0,0115	0,0003
0,0887	0,0575	-0,0264	-0,0369	0,0220	0,0046
0,1491	0,0854	0,0919	-0,0768	0,1446	0,1548
-0,0595	0,1694	0,3221	-0,0572	0,3969	0,3969
0,1079	-0,0891	0,0771	0,1387	0,0002	-0,0090
0,1491	0,2076	0,6126	0,8728	0,3292	0,3201
0,0887	0,0932	0,0641	0,2084	0,1326	0,1087
0,1820	0,0527	0,1086	-0,0661	-0,0752	-0,0820
0,0420	-0,0419	0,0645	-0,0191	-0,0939	-0,1053
0,3741	-0,1498	0,1922	0,0018	0,0501	0,0504
0,0805	0,0472	0,0525	-0,0727	0,1693	0,1421
0,1408	0,4906	-0,1262	-0,0887	0,2970	0,3268
0,1052	0,1681	0,0265	0,0369	0,0257	0,0411
0,2561	0,1622	0,1648	0,0613	0,0940	0,0865
-0,0046	-0,0757	0,0334	0,0137	-0,0094	-0,0016
-0,1281	-0,1104	-0,1016	-0,0238	-0,0927	-0,0943
-0,0623	-0,0596	-0,0965	0,1447	-0,1026	-0,1048
0,0146	-0,1123	-0,0589	-0,0125	-0,0802	-0,0681
0,1326	0,3094	0,1133	-0,0709	0,0485	0,0469
0,1436	-0,0406	0,0687	-0,0012	0,0256	0,0281
-0,0513	0,0850	-0,0042	-0,0411	0,0310	0,0252
-0,0595	0,3375	-0,0028	-0,0185	0,1243	0,0991
0,0722	0,0583	-0,0886	-0,0298	-0,0199	-0,0474
0,0256	-0,1949	0,1160	-0,0601	-0,0118	-0,0174
0,0420	-0,0621	-0,0278	-0,0822	-0,0704	-0,0377
-0,2050	-0,1818	0,0488	-0,0786	-0,0137	-0,0078
-0,2571	-0,1391	-0,2074	-0,0893	-0,1753	-0,1680
-0,2159	-0,1589	-0,2232	-0,0905	-0,2065	-0,2102
-0,0705	-0,0215	-0,1118	-0,0738	-0,0690	-0,0500
-0,1693	-0,1966	-0,1592	-0,0798	-0,1849	-0,1842

Y^*	Z_1	Z_2	Z_3	Z_4	Z_5
-0,2351	-0,1468	-0,2037	-0,0774	-0,1892	-0,1853
-0,2763	-0,1420	-0,2292	-0,0917	-0,2135	-0,2080
-0,2434	-0,1461	-0,1884	-0,0929	-0,2032	-0,2011
0,2973	0,1860	0,1337	0,0542	0,4389	0,4545
-0,2406	-0,2078	-0,2060	-0,0923	-0,1936	-0,1945

3.3. Matriks Korelasi

Matriks korelasi $Z'Z$ dan $Z'Y^*$ yang diperoleh dari data transformasi adalah sebagai berikut :

$$Z'Z = \begin{pmatrix} 1 & 0,4209 & 0,2529 & 0,7457 & 0,7414 \\ 0,4209 & 1 & 0,6602 & 0,7294 & 0,7122 \\ 0,2529 & 0,6602 & 1 & 0,3909 & 0,3729 \\ 0,7457 & 0,7294 & 0,3909 & 1 & 0,9965 \\ 0,7414 & 0,7122 & 0,3729 & 0,9965 & 1 \end{pmatrix}$$

$$Z'Y^* = \begin{pmatrix} 0,5281 \\ 0,6472 \\ 0,3253 \\ 0,6460 \\ 0,6411 \end{pmatrix}$$

3.4. Pemilihan Konstanta Bias c

Pemilihan konstanta bias c diperoleh melalui pendekatan nilai VIF dan plot *ridge trace*. Hasil perhitungan nilai VIF $\hat{\beta}(c)$ dengan berbagai nilai c tersaji pada Tabel 3.

Tabel 3. Nilai VIF $\hat{\beta}(c)$ dengan Berbagai Nilai c

c	VIF $\hat{\beta}_1(c)$	VIF $\hat{\beta}_2(c)$	VIF $\hat{\beta}_3(c)$	VIF $\hat{\beta}_4(c)$	VIF $\hat{\beta}_5(c)$
0,000	2,4941	3,7456	1,8684	170,4536	156,5581
0,001	2,4480	3,6134	1,8542	97,4756	89,6587
0,002	2,4169	3,5322	1,8413	63,1802	58,2182
0,003	2,3919	3,4718	1,8288	44,3707	40,9731
0,004	2,3700	3,4216	1,8168	32,9525	30,5033
0,005	2,3498	3,3772	1,8050	25,5039	23,6723
0,006	2,3307	3,3364	1,7934	20,3762	18,9688
0,007	2,3125	3,2980	1,7820	16,6958	15,5921

0,008	2,2948	3,2615	1,7708	13,9644	13,0854
0,009	2,2776	3,2263	1,7598	11,8812	11,1728
0,010	2,2608	3,1922	1,7489	10,2558	9,6800
0,020	2,1076	2,8914	1,6477	3,8158	3,7501
0,030	1,9741	2,6388	1,5582	2,2071	2,2528
0,040	1,8557	2,4216	1,4785	1,5509	1,6317
0,050	1,7498	2,2328	1,4068	1,2081	1,3006
0,060	1,6546	2,0674	1,3421	0,9999	1,0949
0,070	1,5684	1,9216	1,2832	0,8600	0,9536
0,080	1,4901	1,7923	1,2294	0,7590	0,8494
0,090	1,4185	1,6769	1,1800	0,6821	0,7687
c	VIF $\hat{\beta}_1(c)$	VIF $\hat{\beta}_2(c)$	VIF $\hat{\beta}_3(c)$	VIF $\hat{\beta}_4(c)$	VIF $\hat{\beta}_5(c)$
0,100	1,3529	1,5735	1,1345	0,6213	0,7038
0,200	0,9088	0,9378	0,8173	0,3432	0,3931
0,300	0,6680	0,6424	0,6342	0,2420	0,2749
0,400	0,5186	0,4773	0,5136	0,1882	0,2117
0,500	0,4179	0,3738	0,4280	0,1547	0,1723
0,600	0,3461	0,3039	0,3640	0,1317	0,1455
0,700	0,2928	0,2539	0,3146	0,1150	0,1260
0,800	0,2519	0,2166	0,2753	0,1021	0,1112
0,900	0,2196	0,1879	0,2434	0,0920	0,0995
1,000	0,1937	0,1652	0,2171	0,0837	0,0901

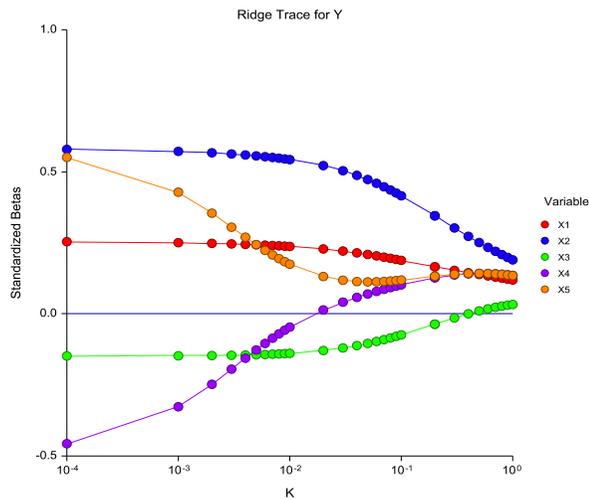
Berdasarkan Tabel 3 dapat dilihat bahwa untuk konstanta bias c dari 0,000 sampai 1,000; diketahui nilai VIF $\hat{\beta}(c)$ semakin besar untuk nilai c semakin kecil. Nilai VIF yang diambil yaitu nilai VIF yang relatif dekat dengan satu. Nilai parameter $\hat{\beta}(c)$ untuk berbagai nilai c tersaji pada Tabel 4.

Tabel 4. Nilai $\hat{\beta}(c)$ Untuk Berbagai Nilai c

c	$\hat{\beta}_1(c)$	$\hat{\beta}_2(c)$	$\hat{\beta}_3(c)$	$\hat{\beta}_4(c)$	$\hat{\beta}_5(c)$
0,000	0,2541	0,5788	-0,1480	-0,4568	0,5509
0,001	0,2506	0,5719	-0,1475	-0,3268	0,4283
0,002	0,2482	0,5669	-0,1467	-0,2478	0,3543
0,003	0,2462	0,5628	-0,1458	-0,1945	0,3050
0,004	0,2446	0,5593	-0,1449	-0,1560	0,2699

c	$\hat{\beta}_1(c)$	$\hat{\beta}_2(c)$	$\hat{\beta}_3(c)$	$\hat{\beta}_4(c)$	$\hat{\beta}_5(c)$
0,005	0,2431	0,5561	-0,1439	-0,1269	0,2438
0,006	0,2418	0,5532	-0,1429	-0,1041	0,2235
0,007	0,2406	0,5504	-0,1419	-0,0856	0,2075
0,008	0,2395	0,5478	-0,1409	-0,0703	0,1945
0,009	0,2384	0,5453	-0,1398	-0,0575	0,1838
0,010	0,2373	0,5429	-0,1388	-0,0465	0,1748
0,020	0,2285	0,5218	-0,1290	0,0141	0,1315
0,030	0,2212	0,5037	-0,1199	0,0414	0,1184
0,040	0,2148	0,4875	-0,1116	0,0581	0,1138
0,050	0,2092	0,4729	-0,1040	0,0700	0,1127
0,060	0,2041	0,4595	-0,0970	0,0791	0,1131
0,070	0,1996	0,4472	-0,0905	0,0864	0,1142
0,080	0,1955	0,4359	-0,0845	0,0926	0,1157
0,090	0,1918	0,4254	-0,0789	0,0978	0,1174
0,100	0,1884	0,4156	-0,0738	0,1023	0,1192
0,200	0,1656	0,3451	-0,0364	0,1273	0,1327
0,300	0,1531	0,3023	-0,0142	0,1370	0,1393
0,400	0,1448	0,2729	0,0003	0,1409	0,1420
0,500	0,1387	0,2511	0,0105	0,1422	0,1426
0,600	0,1337	0,2341	0,0178	0,1419	0,1420
0,700	0,1295	0,2202	0,0233	0,1409	0,1407
0,800	0,1257	0,2086	0,0275	0,1393	0,1390
0,900	0,1224	0,1987	0,0307	0,1375	0,1371
1,000	0,1193	0,1901	0,0332	0,1355	0,1350

Selanjutnya berdasarkan data pada Tabel 4 dibentuk plot *ridge trace* seperti tersaji pada Gambar 1.



Gambar 1. Ridge Trace

Berdasarkan pendekatan nilai VIF dan plot *ridge trace*, untuk berbagai kemungkinan nilai c , nilai VIF yang relatif dekat dengan satu diperoleh ketika konstanta bias c bernilai 0,20. Hal ini juga menunjukkan bahwa ketika $c = 0,20$ koefisien $\hat{\beta}$ relatif lebih stabil.

3.5. Persamaan Regresi Ridge

Persamaan regresi ridge dengan nilai c yang diambil sebesar 0,20 adalah sebagai berikut

$$Y^* = 0,1656Z_1 + 0,3451Z_2 - 0,0364Z_3 + 0,1273Z_4 + 0,1327Z_5$$

Selanjutnya, setelah memperoleh persamaan regresi ridge di atas, persamaan tersebut akan diubah kedalam bentuk semula. Sehingga model yang diperoleh adalah :

$$\hat{Y} = 39,9451 + 0,0049X_1 + 0,0584X_2 - 0,0079X_3 + 0,0002X_4 + 0,0008X_5$$

3.5.1. Uji Keberartian Regresi Secara Simultan

Hasil perhitungan statistika yang digunakan untuk uji keberartian koefisien regresi secara simultan dapat dilihat pada Tabel 5.

Tabel 5. Tabel Anava Untuk Regresi Ridge

Sumber Variasi	dk	JK	KT	F hitung
Regresi	5	61912,48	12382,5	5,5916
Residual	32	70863,73	2214,491	
Total	37	132776,2		

Dengan mengambil taraf nyata $\alpha = 0,05$ maka diperoleh $F_{\text{tabel}(5;32;0,05)} = 2,5123$. Karena $F_{\text{hitung}} = 5,5916 > F_{\text{tabel}} = 2,5123$, maka H_0 ditolak. Hal ini mempunyai makna bahwa secara simultan koefisien regresi tersebut berarti atau signifikan.

3.5.2. Uji Keberartian Regresi Secara Parsial

Hasil perhitungan statistika untuk uji keberartian koefisien regresi secara parsial dapat dilihat pada Tabel 6.

Tabel 6. Uji Keberartian Koefisien Regresi Ridge

Bi	Sbi	$ t_{\text{hitung}} $	t_{tabel}	Kesimpulan
0,00485	0,00361	1,34376	2,35184	Tidak Berarti
0,05837	0,02115	2,75965		Berarti
-0,00790	0,02533	0,31192		Tidak Berarti
0,00016	0,000096	1,68972		Tidak Berarti
0,00082	0,0005	1,64296		Tidak Berarti

Berdasarkan Tabel 6 diketahui bahwa koefisien regresi ridge yang berarti atau signifikan adalah koefisien regresi ridge variabel X_2 .

$$\hat{Y} = 39,9451 + 0,0584X_2$$

Untuk persamaan regresi ridge dapat disimpulkan bahwa setiap satu kenaikan bayi yang memiliki berat badan lahir rendah maka akan meningkatkan angka kematian bayi sebesar 0,0584.

3.6. Persamaan Regresi Ridge Parsial

3.6.1. Dekomposisi Matriks

Dasar dari pendekatan regresi ridge parsial adalah mendekomposisi matriks korelasi, yaitu dengan *Singular Value Decomposition* (SVD). Dekomposisi nilai singular tersebut tidak lepas dari perihal nilai eigen dari matriks $Z'Z$. Sebelum dilakukan dekomposisi matriks, terlebih dahulu dihitung nilai eigen dari matriks $Z'Z$ dengan hasil tersaji pada Tabel 7.

Tabel 7. Nilai Eigen

Variabel	Nilai Eigen
λ_1	3,4801
λ_2	0,9600
λ_3	0,3931
λ_4	0,1638
λ_5	0,0031

Singular value decomposition merupakan dekomposisi suatu matriks kedalam tiga komponen matriks. Matriks korelasi $Z'Z$ merupakan matriks simetris, sehingga dekomposisi nilai singular yang terbentuk adalah sebagai berikut :

$$Z'Z = U\Delta U'$$

Dimana

$$U = \begin{pmatrix} -0,4153 & 0,4430 & -0,6863 & 0,4001 & -0,0141 \\ -0,4522 & -0,3826 & 0,4257 & 0,6835 & -0,0261 \\ -0,3199 & -0,7466 & -0,4808 & -0,3302 & -0,0037 \\ -0,5124 & 0,2139 & 0,2359 & -0,3387 & 0,7220 \\ -0,5080 & 0,2328 & 0,2469 & -0,3859 & -0,6913 \end{pmatrix}$$

$$\Delta = \begin{pmatrix} 3,4801 & 0 & 0 & 0 & 0 \\ 0 & 0,9600 & 0 & 0 & 0 \\ 0 & 0 & 0,3931 & 0 & 0 \\ 0 & 0 & 0 & 0,1638 & 0 \\ 0 & 0 & 0 & 0 & 0,0031 \end{pmatrix}$$

$$U' = \begin{pmatrix} -0,4153 & -0,4522 & -0,3199 & -0,5124 & -0,5080 \\ 0,4430 & -0,3826 & -0,7466 & 0,2139 & 0,2328 \\ -0,6863 & 0,4257 & -0,4808 & 0,2359 & 0,2469 \\ 0,4001 & 0,6835 & -0,3302 & -0,3387 & -0,3859 \\ -0,0141 & -0,0261 & -0,0037 & 0,7220 & -0,6913 \end{pmatrix}$$

3.6.2. Perbandingan Nilai Eigen

Sebelum diperoleh persamaan regresi ridge parsial, terlebih dahulu akan ditentukan variabel yang memiliki kolinearitas tinggi dengan pengukuran δ . Hasil δ yang diperoleh berdasarkan nilai eigen pada matriks $Z'Z$ tersaji pada tabel 8.

Tabel 8. Nilai-nilai δ

Variabel	Nilai δ
δ_1	1
δ_2	0,2758591
δ_3	0,1129452
δ_4	0,0470616
δ_5	0,0008839

Kesepakatan umum menyatakan bahwa nilai $\delta = 0,001$ dapat digunakan sebagai indikasi dalam menentukan kolinearitas

tinggi. Konstanta bias c ditambahkan ketika nilai eigen yang rendah dan kurang dari δ . Pada Tabel 8 dapat dilihat bahwa variabel yang kurang dari 0,001 yaitu hanya δ_5 , sehingga konstanta bias c yang ditambahkan pada parameter regresi ridge parsial hanya satu. Untuk $c_1 = c_2 = c_3 = c_4 = 0$ dan $c_5 = 0,2$, sehingga persamaan regresi ridge parsial adalah sebagai berikut:

$$Y^* = 0,2474Z_1 + 0,5635Z_2 - 0,1520Z_3 + 0,0929Z_4 + 0,0171Z_5$$

Selanjutnya, setelah memperoleh persamaan regresi ridge parsial di atas, persamaan tersebut akan diubah kedalam bentuk semula. Sehingga model yang diperoleh adalah :

$$\hat{Y} = 28,0636 + 0,0073X_1 + 0,0953X_2 - 0,0330X_3 + 0,0001X_4 + 0,0001X_5$$

3.6.3. Uji Keberartian Regresi Secara Simultan

Hasil perhitungan statistika yang digunakan untuk uji keberartian koefisien regresi secara simultan dapat dilihat pada Tabel 9.

Tabel 9. Tabel Anava Unuk Regresi Ridge Parsial

Sumber Variasi	dk	JK	KT	F hitung
Regresi	5	68626,14	13725,2	6,8474
Residual	32	64142,35	2004,45	
Total	37	132768,49		

Dengan mengambil taraf nyata $\alpha = 0,05$ maka diperoleh $F_{tabel(5;32;0,05)} = 2,5123$. Karena $F_{hitung} > F_{tabel}$ yaitu $6,8474 > 2,5123$ maka H_0 ditolak. Hal ini mempunyai makna bahwa secara simultan koefisien regresi tersebut berarti atau signifikan.

3.6.4. Uji Keberartian Regresi Secara Parsial

Hasil perhitungan statistika untuk uji keberartian koefisien regresi secara parsial dapat dilihat pada Tabel 10.

Tabel 10. Uji Keberartian Koefisien Regresi Ridge Parsial

b_i	S_{b_i}	$ t_{hitung} $	t_{tabel}	Kesimpulan
0,00725	0,00360	2,01323	2,35184	Tidak Berarti
0,09530	0,02062	4,62068		Berarti
-0,03297	0,02633	1,25236		Tidak Berarti
0,00012	0,00016	0,75602		Tidak Berarti
0,00011	0,00076	0,13876		Tidak Berarti

Berdasarkan Tabel 10 diketahui bahwa koefisien regresi ridge parsial yang berarti atau signifikan adalah koefisien regresi variabel X_2

$$\hat{Y} = 28,0636 + 0,0953X_2$$

Untuk persamaan regresi ridge dapat disimpulkan bahwa setiap satu kenaikan bayi yang memiliki berat badan lahir rendah maka akan meningkatkan angka kematian bayi sebesar 0,0953.

3.7. Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan memperhatikan nilai MSE. Hasil perhitungan nilai MSE untuk model regresi ridge dan model regresi ridge parsial dapat dilihat pada Tabel 11.

Tabel 11. Nilai MSE

TEKNIK ANALISIS DATA	MSE
Regresi Ridge	2214,491
Regresi Ridge Parsial	2004,45

Berdasarkan Tabel 11 diketahui bahwa nilai MSE yang paling kecil adalah 2004,45; dengan demikian model terbaik yang terpilih adalah model regresi ridge parsial.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan pada penelitian ini maka dapat disimpulkan :

1. Prosedur penggunaan pada metode regresi ridge parsial yaitu mentransformasikan data, membentuk matriks korelasi $Z'Z$ dan $Z'Y^*$, membentuk matriks dekomposisi $U\Delta U'$ berdasarkan nilai eigen, menentukan variabel yang berkolinearitas tinggi δ , menentukan konstanta bias c , menginput konstanta bias c , mengestimasi parameter dan membentuk model regresi ridge parsial.
2. Penerapan regresi ridge parsial pada kasus angka kematian bayi di Provinsi Jawa Timur tahun 2018 dengan menambahkan konstanta bias $c = 0,20$ pada variabel yang berkolinearitas tinggi, maka diperoleh persamaan regresi ridge parsial sebagai berikut :

$$\hat{Y} = 28,0636 + 0,0953X_2$$

3. Hasil perbandingan persamaan regresi ridge dengan regresi ridge parsial masing-masing memiliki nilai MSE sebesar 2214,491 dan 2004,45. Nilai

MSE yang terkecil adalah 2004,45, sehingga persamaan regresi yang terbaik adalah persamaan regresi ridge parsial.

5. DAFTAR PUSTAKA

- [1] Chandrasekhar, C., Bagyalakshmi, H., Srinivasan, M., & Gallo, M. (2016). Partial Ridge Regression Under Multicollinearity. *Journal of Applied Statistics*.
- [2] Dinas Kesehatan. 2018. *Profil Kesehatan Provinsi Jawa Timur Tahun 2018*. Surabaya: Dinas Kesehatan Provinsi Jawa Timur.
- [3] Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics Fifth Edition*. New York: McGraw-Hill Irwin.
- [4] Hasriani. (2014). *Perbandingan Regresi Ridge (Regresi Gulud) dan Principal Component Analysis (Analisis Komponen Utama) dalam Mengatasi Masalah Multikolinearitas*. Makassar: Universitas Islam Negeri Alauddin.
- [5] Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression : Biased Estimation for Nonorthogonal Problems. *Technometrics*, 55-67.
- [6] Sembiring, R. (2003). *Analisis Regresi Edisi Kedua*. Bandung: ITB.
- [7] Sudjana. (2005). *Metoda Statistika*. Bandung: PT Tarsito.
- [8] Wasilaine, T., Talakua, M., & Lesnussa, Y. (2014). Model Regresi Ridge Untuk Mengatasi Model Regresi Linier Berganda Yang Mengandung Multikolinearitas. *Barekeng Jurnal Ilmu Matematika dan Terapan*, 31-37.