PERBANDINGAN REGRESI RIDGE DAN PRINCIPAL COMPONENT REGRESSION DALAM MENGATASI MULTIKOLINIERITAS PADA FAKTOR-FAKTOR YANG MEMENGARUHI KEMISKINAN DI INDONESIA

COMPARISON OF RIDGE REGRESSION AND PRINCIPAL COMPONENT REGRESSION IN OVERCOMING MULTICOLINIERITY OF FACTORS AFFECTING POVERTY IN INDONESIA

RAEHANI RAHMAWATI
Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mataram.
Jl. Majapahit No. 62 Mataram, Nusa Tenggara Barat, Indonesia. Tel./Fax. (0370) 63604.
email: raehanirahmawati17@gmail.com.

Abstrak. Analisis regresi merupakan suatu metode statistika yang bertujuan untuk mengukur korelasi atau hubungan antar variabel independen terhadap variabel dependen. Analisis regresi memiliki beberapa asumsi yang harus terpenuhi agar memperoleh suatu model yang baik. Salah satu asumsi yang harus terpenuhi adalah tidak adanya multikolinieritas antara variabel independen. Multikolinieritas merupakan suatu keadaan dimana terdapat hubungan atau korelasi antar variabel independen yang menyebabkan keakuratan model berkurang. Penelitian ini bertujuan untuk membandingkan regresi ridge dan PCR dalam mengatasi multikolinieritas. Regresi ridge merupakan modifikasi dari metode kuadrat terkecil bekerja dengan menambahkan tetapan bias pada diagonal utama matriks varians-kovarians. Perhitungan tetapan bias pada penelitian ini menggunakan metode Kibria. Principal component regression merupakan gabungan dari principal component analysis (PCA) dan analisis regresi, metode ini bekerja dengan meregresikan komponen-komponen utama. Adapun kriteria pembanding yang digunakan adalah R^2 dan R^2_{adj}. Data penelitian yang digunakan yaitu data faktor-faktor yang memengaruhi kemiskinan di Indonesia pada tahun 2021. Berdasarkan hasil perhitungan diperoleh bahwa keakuratan model regresi ridge sebesar 83,6%, sedangkan keakuratan model principal component regression sebesar 77,5% nilai tersebut dilihat dari besar nilai R^2. Selanjutnya nilai R^2_{adj} dari model regresi ridge lebih besar dari nilai R^2_{adj} model principal component regression yaitu $0{,}7991 > 0{,}761$. Dengan demikian dapat disimpulkan bahwa regresi ridge dengan parameter Kibria lebih baik dibandingkan dengan principal component regression dalam mengatasi multikolinieritas pada data faktor-faktor yang memengaruhi kemiskinan di Indonesia.

Kata kunci: Analisis Regresi, Estimasi Parameter Kibria, Kemiskinan, Multikolinieritas, *Principal Component Analysis*, Regresi *Ridge*.

Abstract. *Regression analysis is a statistical method that aims to measure the correlation or relationship between the independent variables on the dependent variable. Regression analysis has several assumptions that must be met in order to obtain a good model. One of the assumptions that must be fulfilled is the absence of multicollinearity between the independent variables. Multicollinearity is a situation where there is a relationship or correlation between independent variables which causes the accuracy of the model to decrease. This study aims to compare ridge regression and* PCR *in overcoming multicollinearity. Ridge regression is a modification of the least squares method which works by adding a bias constant to the main diagonal of the variance-covariance matrix. The calculation of the bias constant in this study uses the Kibria method. Principal component regression is a combination of principal component analysis (*PCA*) and regression analysis, this method works by regressing the main components. The comparison criteria used are R^2 and R^2_{adj}. The research data used is data on factors that influence poverty in Indonesia in 2021. Based on the calculation results, it is found that the accuracy of the ridge regression model is 83.6%, while the accuracy of the principal component regression model is 77.5%, this value is seen from the large value R^2. Furthermore, the R^2_{adj} value of the ridge regression model is greater than the R^2_{adj} value of the principal component regression model, namely 0.7991> 0.761. Thus it can be concluded that ridge regression with Kibria parameters is better than principal component regression in overcoming multicollinearity in data on factors that influence poverty in Indonesia.*

Key words: *Kibria Estimator, Multicollinearity, Poverty, Principal Component Regression, Regression Analysis, Ridge Regression.*

PENDAHULUAN

Analisis regresi merupakan suatu metode statistika yang bertujuan untuk mengukur korelasi atau hubungan antar variabel independen terhadap variabel dependen. Analisis regresi memiliki beberapa asumsi yang harus terpenuhi agar memperoleh suatu model yang baik. Salah satu asumsi yang harus terpenuhi adalah tidak adanya multikolinieritas antara variabel independen (Kutner, et al. 2005) .

Multikolinieritas merupakan suatu keadaan dimana terdapat hubungan atau korelasi antar variabel independen yang menyebabkan keakuratan model berkurang (Montgomery, Peck and Vining 2012). Oleh karena itu, sebelum melakukan analisis lebih lanjut masalah multikolinieritas perlu diselesaikan. Terdapat beberapa metode untuk menyelesaikan masalah multikolinieritas diantaranya least square, principal component regression (PCR), regresi ridge, least absolute shrinkage and selection operator (LASSO), dan akar laten (Hastie, Tibshirani and Friedman 2008). Akan tetapi, dalam penerapan metode-metode tersebut diperoleh bahwa regresi ridge dan PCR merupakan metode terbaik untuk menyelesaikan multikolinieritas karena memiliki nilai average mean square error (AMSE) paling kecil dibandingkan metode lainnya (Herawati, et al. 2018).

Regresi ridge dan PCR bersama-sama berguna untuk menghilangkan multikolinieritas, akan tetapi kedua metode ini berbeda. Regresi ridge yang merupakan modifikasi dari metode kuadrat terkecil bekerja dengan menambahkan tetapan bias pada diagonal utama matriks varians-kovarians. Penambahan tetapan bias pada regresi ridge menghasilkan koefisien regresi yang stabil, sehingga mengakibatkan berkurangnya nilai varians (Younker 2012). Sementara itu metode PCR merupakan gabungan dari principal component analysis (PCA) dan analisis regresi. PCR bekerja dengan mereduksi variabel-variabel independen awal ke dalam variabel baru yang disebut dengan komponen utama yang saling independen, kemudian meregresikan komponen-komponen utama tersebut sehingga memperoleh suatu model regresi. Regresi ridge dan PCR sama-sama memperoleh hasil analisis yang memiliki varians lebih kecil (Montgomery, Peck and Vining 2012).

Berdasarkan beberapa penelitian yang membandingkan metode regresi ridge dengan PCR diperoleh hasil yang berbeda-beda tergantung studi kasus yang dipilih. Hal tersebut menandakan kebaikan model dari kedua metode tersebut tidak bersifat umum.

Oleh karena itu, dalam penelitian ini regresi ridge dan PCR akan diaplikasikan ke dalam faktor-faktor yang memengaruhi kemiskinan di Indonesia. Regresi ridge dan PCR akan digunakan untuk menyelesaikan masalah multikolinieritas dalam faktor-faktor yang memengaruhi kemiskinan di Indonesia.

MATERI DAN METODE

Prosedur Kerja

Prosedur analisis data yang dilakukan dalam penelitian ini sebagai berikut:

1. Menentukan deskripsi data yaitu rata-rata dan simpangan baku dari data kemiskinan dan faktor-faktornya.
2. Melakukan standarisasi data penelitian dengan metode pemusatan dan penskalaan.
3. Melakukan estimasi parameter β dengan menggunakan metode *ordinary least square* atau metode kuadrat terkecil.
4. Mendeteksi multikolinieritas. Jika terdeteksi adanya multikolinieritas maka diselesaikan dengan menggunakan metode regresi *ridge* dan PCR.
5. Menyelesaikan masalah multikolinieritas dengan regresi *ridge*
   a. Menghitung konstanta bias (c) dengan metode Kibria.
   b. Melakukan estimasi parameter regresi *ridge* $(\hat{\beta}_{ridge})$ dengan estimator parameter Kibria.
   c. Menguji signifikansi parameter model regresi *ridge* dengan menggunakan uji F dan uji t dengan.
   d. Mengembalikan persamaan regresi *ridge* ke variabel awal X.
6. Menyelesaikan masalah multikolinieritas dengan PCR
   a. Menentukan nilai eigen (λ) dari matriks korelasi (ρ).
   b. Menentukan skor komponen utama (K) yang terbentuk.
   c. Melakukan regresi linier berganda variabel dependen (Y) terhadap komponen utama (K).
   d. Menguji signifikansi parameter model PCR dengan menggunakan uji F dan uji t.
   e. Merubah persamaan regresi linier berganda ke dalam bentuk variabel standar (Z).

f. Merubah kembali persamaan regresi linier berganda ke dalam bentuk variabel semula (X).

7. Melakukan uji asumsi klasik regresi linier berganda diantaranya uji normalitas, uji heteroskedastisitas, dan uji autokorelasi.
8. Membandingkan model regresi yang terbentuk dari metode regresi *ridge* dan PCR dengan menggunakan nilai R^2 dan nilai R^2_{adj}.

Analisis Data

Penelitian ini menggunakan data sekunder tentang jumlah penduduk miskin dan faktor yang memengaruhi kemiskinan berdasarkan provinsi di Indonesia tahun 2021 yang diperoleh dari Badan Pusat Statistik (BPS). Data berjumlah $n = 34$ yang merupakan jumlah provinsi di Indonesia dengan variabel-variabel sebagai berikut indeks penduduk miskin (y), kepadatan penduduk (X_1), tingkat pengangguran terbuka (X_2), PDRB perkapita (X_3), Indeks keparahan kemiskinan (X_4), rata-rata lama sekolah (X_5), indeks pembangunan manusia (X_6), dan pengeluaran perkapita (X_7). Data yang sudah diperoleh distandarisasi terlebih dahulu kemudian dilakukan proses berikut.

1. Multikolinieritas

Pengujian multikolinieritas dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF) yang terbentuk dari masing-masing variabel independen terhadap variabel dependen. Saat nilai VIF besar maka multikolinieritas pada model regresi tersebut semakin parah (Montgomery, Peck and Vining 2012). Berikut merupakan rumusan VIF untuk mendeteksi multikolinieritas (Draper and Smith 1998).

$$VIF = \frac{1}{1 - R^2} \tag{1}$$

Nilai VIF dapat juga dihitung menggunakan persamaan berikut

$$VIF = \left(\left(\frac{1}{n-1}\right)\boldsymbol{C}\right)^{-1} \tag{2}$$

Apabila diperoleh nilai VIF ≥ 10 maka terjadi multikolinieritas antara variabel independen.

2. Regresi Ridge

Regresi *ridge* merupakan pengembangan metode estimasi jumlah kuadrat terkecil yang berguna untuk menyelesaikan multikolinieritas dengan pembatas pada penjumlahan kuadrat koefisien. Multikolinieritas pada sebagian atau seluruh variabel independen mengakibatkan ketidak stabilan nilai estimasi yang dihasilkan oleh matriks $\boldsymbol{X^tX}$. Modifikasi metode estimasi jumlah kuadrat terkecil pada regresi *ridge* adalah adanya parameter *ridge c* pada diagonal utama matriks $\boldsymbol{X^tX}$. Modifikasi tersebut mengakibatkan koefisien estimator *Ridge* meningkat dengan penambahan parameter *Ridge c* (Draper and Smith 1998). Estimator regresi *ridge* dituliskan dalam formula berikut (Younker 2012).

$$\hat{\beta}_{ridge} = arg\min_{\beta}\left(\sum_{i=1}^{n}(y_i - \sum_{j=1}^{k} x_{ij}\beta_j)^{\ 2} + c\sum_{j=1}^{k}\beta_j^2\right) \tag{3}$$

Persamaan (3) *equivalent* dengan

$$\boldsymbol{\hat{\beta}_{ridge}} = (\boldsymbol{X^tX} + c\boldsymbol{I})^{-1}\boldsymbol{X^tY} \tag{4}$$

Estimasi parameter regresi *ridge* dengan metode Kibria merupakan pengembangan metode estimasi sebelumnya yaitu metode Hoerl dan Kennard. Metode Kibria menggunakan nilai median untuk menentukan tetapan bias, median digunakan dikarenakan tidak sensitif terhadap data yang mengandung pencilan. Nilai tetapan bias dengan metode Kibria dirumuskan sebagai berikut.

$$c^0 = median\left\{\frac{\hat{\sigma}^2}{\boldsymbol{\hat{\alpha}}_{ols,j}^2}, \qquad j = 1, 2, \dots, k\right\} \tag{5}$$

Dengan demikian didapatkan $\boldsymbol{\hat{\alpha}}_{ridge}$ sebagai berikut.

$$\boldsymbol{\hat{\alpha}}_{ridge}^1 = (\boldsymbol{\Lambda} + c^0\boldsymbol{I})^{-1}\boldsymbol{W^tY}^* \tag{6}$$

Kemudian, nilai parameter $\boldsymbol{\hat{\beta}}_{ridge}$ dihitung menggunakan tetapan bias c^0 dengan formula sebagai berikut.

$$\boldsymbol{\hat{\beta}}_{ridge}^1 = (\boldsymbol{Z^tZ} + c^0\boldsymbol{I})^{-1}\boldsymbol{Z^tY} \tag{7}$$

Dalam Persamaan (7) $\boldsymbol{\hat{\beta}}_{ridge}^1$ menyatakan estimator awal dari metode regresi *ridge* dan digunakan untuk menentukan nilai c^1 sebagai berikut.

$$c^1 = median\left\{\frac{\hat{\sigma}^2}{\left(\boldsymbol{\hat{\alpha}}_{ridge,j}^1\right)^2}, \qquad j = 1, 2, \dots, k\right\} \tag{8}$$

Proses iterasi akan terus berlanjut hingga diperoleh suatu nilai estimator yang stabil atau konsisten, hal tersebut dapat dilihat dari nilai eigen yang konvergen. Selain itu, iterasi dapat dihentikan hingga diperoleh nilai $error \leq 0{,}01$ (Younker 2012).

3. Principal Component Regression

Principal component regression merupakan suatu metode yang mengkombinasikan analisis komponen utama dengan analisis regresi. Pengkombinasian ini bermanfaat agar variabel independen dalam analisis regresi terbebas dari multikolinieritas, sehingga diperoleh model yang lebih akurat. Prosedur analisis dengan menggunakan regresi komponen utama sebagai berikut. Prosedur analisis dengan menggunakan regresi komponen utama sebagai berikut (Mattjik and Sumertajaya 2011).

a. Menstandarisasi data variabel penelitian.

b. Menghitung nilai eigen (λ), vektor eigen ($\boldsymbol{a}$), dan komponen utama (K). Nilai eigen (λ) ditentukan melalui rumusan berikut.

$$|\boldsymbol{\rho} - \lambda \boldsymbol{I}| = 0 \tag{9}$$

Sementara itu, untuk menentukan vektor eigen $\boldsymbol{a_j}$ dari matriks korelasi $\boldsymbol{\rho}$ dihitung melalui rumusan berikut.

$$(\boldsymbol{\rho} - \lambda \boldsymbol{I})\boldsymbol{a_j} = 0 \tag{10}$$

c. Melakukan analisis regresi komponen utama yang terpilih terhadap variabel dependen. Demikian sehingga diperoleh model regresi komponen utama sebagai berikut.

$$Y = w_0 + w_1K_1 + w_2K_2 + \cdots + w_mK_m \tag{11}$$

d. Mentransformasi model regresi komponen utama ke dalam model regresi dengan variabel yang sudah distandarisasi, diperoleh model sebagai berikut.

$$Y = b_0 + b_1Z_1 + b_2Z_2 + \cdots + b_kZ_k + \varepsilon \tag{12}$$

e. Mentransformasi model regresi dengan variabel independen yang distandarisasi ke dalam model regresi dengan variabel independen asal, diperoleh model sebagai berikut.

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_kX_k + \varepsilon \tag{13}$$

4. Kebaikan Model

Kebaikan model (Goodness of Fit) merupakan suatu tahapan dalam analisis yang bertujuan untuk mengetahui seberapa baik suatu model regresi mengestimasi nilai variabel dependen. Kebaikan suatu model dapat diketahui dengan melihat nilai koefisien determinasi atau R^2. Suatu model regresi dikatakan sesuai apabila nilai R^2 mendekati 1. Perhitungan nilai R^2 menggunakan rumusan berikut (James, et al. 2013).

$$R^2 = \frac{SST - SSE}{SST} \tag{14}$$

Penerapan koefisien determinasi seringkali menimbulkan bias pada model yang terbentuk karena penambahan variabel independen. Nilai koefisien determinasi akan berubah saat menambahkan variabel independen ke dalam model, tanpa memerhatikan signifikansi dari variabel tersebut. Oleh karena itu terdapat metode pengembangan dari R^2 yaitu $adjusted - R^2$. Berikut adalah rumusan untuk menghitung $adjusted - R^2$ (Montgomery, Peck and Vining 2012).

$$R_{adj}^2 = 1 - \frac{SSE/_{(n-k)}}{SST/_{(n-1)}} \tag{15}$$

HASIL DAN DISKUSI

1. Deteksi multikolinieritas

Table 1. VIF pada Deteksi Multikolinieritas

Variabel	VIF	Keputusan
Z_1	3,043	Tidak ada multikolinieritas
Z_2	1,521	Tidak ada multikolinieritas
Z_3	2,264	Tidak ada multikolinieritas
Z_4	1,920	Tidak ada multikolinieritas
Z_5	4,101	Tidak ada multikolinieritas
Z_6	11,032	Ada multikolinieritas
Z_7	8,800	Tidak ada multikolinieritas

Berdasarkan Tabel 1 di atas dapat dilihat bahwa variabel Z_6 yaitu variabel Indeks Pembangunan Manusia memiliki nilai VIF > 10 yaitu sebesar $11{,}032$. Hal tersebut berarti model regresi mengalami masalah multikolinieritas, sehingga untuk mengatasi masalah tersebut digunakan metode regresi *ridge* dan *principal component regression*.

2. Regresi ridge

Hasil perhitungan tetapan bias regresi ridge dengan metode kibria beserta nilai estimasi parameter regresi ridge disajikan dalam tabel berikut

Table 2. Nilai Estimasi Parameter Kibria Beserta Nilai VIF

Iterasi	Nilai c	$\hat{\alpha}_{ridge}$	$Error$	$\hat{\beta}^*_{ridge}$	VIF
0	0,8787	-0,2128	0,1409	0,1056	0,4947
		-0,208		-0,1351	0,3886
		0,5433		-0,1374	0,4895
		0,0964		0,8503	0,3721
		-0,1228		-0,037	0,5432
		-0,5408		0,1238	0,7183
		-0,3273		-0,1402	0,6733
1	2,102	-0,1426	0,111	0,0954	0,1313
		-0,1815		-0,1298	0,1203
		0,493		-0,1236	0,1321
		0,0896		0,7991	0,1162
		-0,1169		-0,0241	0,138
		-0,5228		0,0541	0,1531
		-0,3244		-0,1159	0,1497
2	2,8894	-0,1176	0,0519	0,0916	0,0776
		-0,1677		-0,1263	0,0732
		0,4653		-0,1159	0,078
		0,0856		0,7705	0,0712
		-0,1134		-0,0218	0,0804
		-0,5118		0,0259	0,0861
		-0,3226		-0,1102	0,0848
3	3,38320	-0,106	0,0279	0,0897	0,0593
		-0,1601		-0,1242	0,0566
		0,4495		-0,1115	0,0596
		0,0834		0,7541	0,0553
		-0,1113		-0,0214	0,0611
		-0,5052		0,0118	0,0646
		-0,3215		-0,1083	0,0638
4	3,7128	-0,0994	0,0171	0,0885	0,0505
		-0,1554		-0,1229	0,0485
		0,4395		-0,1087	0,0508
		0,0819		0,7436	0,0475
		-0,11		-0,0215	0,0519
		-0,5009		0,0036	0,0545
		-0,3207		-0,1074	0,0539
5	3,9414	-0,0953	0,0112	0,0877	0,0455
		-0,1523		-0,122	0,0439
		0,4328		-0,1069	0,0458
		0,0809		0,7366	0,043
		-0,1091		-0,0216	0,0467
		-0,4979		-0,0017	0,0488
		-0,3202		-0,107	0,0484
6	4,1038	-0,0926	0,0077	0,0872	0,0424
		-0,1502		-0,1214	0,041
		0,4282		-0,1056	0,0426
		0,0802		0,7317	0,0402
		-0,1084		-0,0218	0,0435

Iterasi	Nilai c	$\hat{\alpha}_{ridge}$	*Error*	$\hat{\beta}^{*}{}_{ridge}$	VIF
		-0,4958		-0,0052	0,0453
		-0,3198		-0,1068	0,0449

Berdasarkan Tabel 2 di atas diperoleh nilai *error* $\leq 0{,}01$ yaitu sebesar 0,007 pada iterasi ke enam dengan nilai tetapan bias sebesar 4,1038. Berdasarkan nilai tetapan bias c^6 pada Tabel 2 di atas, diperoleh estimasi parameter dengan metode *ridge* yang tidak mengandung multikolinieritas. Hal tersebut dapat dilihat pada nilai VIF yang lebih kecil dari 10, sehingga metode estimasi parameter *ridge* dengan metode Kibria dapat menyelesaikan masalah multikolinieritas.oleh karena itu diperoleh model regresi dari metode regresi ridge sebagai berikut.

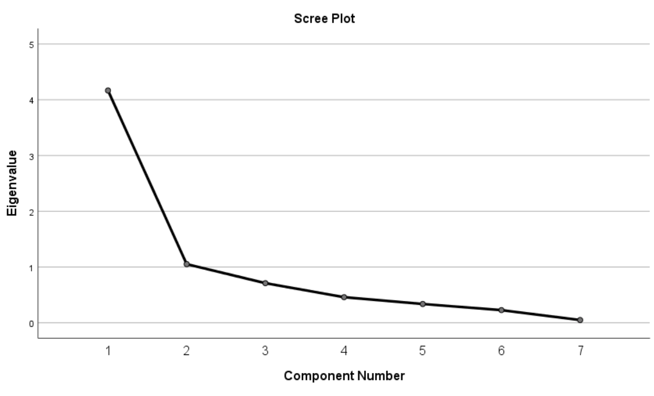
$$Y = -8{,}1e^{-7} + 0{,}0872Z_1 - 0{,}1214Z_2 - 0{,}1056Z_3 + 0{,}7317Z_4 - 0{,}0218Z_5 - 0{,}0052Z_6 - 0{,}1068Z_7 \qquad (16)$$

Selanjutnya Persamaan (16) ditransformasi ke variabel awal, sehingga diperoleh model regresi dalam Persamaan berikut

$$Y = 0{,}53923 + 0{,}00003X_1 - 0{,}06670X_2 - 0{,}000003X_3 + 1{,}46340X_4 - 0{,}02344X_5 - 0{,}00132X_6 - 0{,}00005X_7$$

3. *Principal component regression*

Principal component regression merupakan kombinasi dari *principal component analaysis* dan analisis regresi. Proses analisis dimulai dengan melakukan standarisasi data dan kemudian menentukan banyak komponen utama yang terbentuk berdasarkan nilai eigen yang diperoleh. Berikut adalah jumlah komponen utama yang terbentuk



Gambar 1. scree plot nilai eigen

Berdasarkan gambar di atas, sebanyak 2 komponen memiliki nilai eigen lebih besar dari satu yaitu K_1 dan K_2. Komponen K_1 memiliki nilai eigen sebesar 4,166, nilai tersebut mampu menjelaskan keragaman sebesar 59,517% dari total cumulative. Sementara itu, komponen K_2 dengan nilai eigen sebesar 1,052 mampu menjelaskan keragaman sebesar 15,025% dari total cumulative. diperoleh dua komponen utama yang terbentuk dengan persamaan komponen utama sebagai berikut

$$K_1 = 0{,}350\mathrm{Z}_1 + 0{,}317\mathrm{Z}_2 + 0{,}366\mathrm{Z}_3 - 0{,}294\mathrm{Z}_4 + 0{,}403\mathrm{Z}_5 + 0{,}448\mathrm{Z}_6 + 0{,}440\mathrm{Z}_7$$

$$K_2 = -0{,}531\mathrm{Z}_1 - 0{,}004\mathrm{Z}_2 - 0{,}451\mathrm{Z}_3 - 0{,}668\mathrm{Z}_4 + 0{,}042\mathrm{Z}_5 + 0{,}251\mathrm{Z}_6 + 0{,}059\mathrm{Z}_7$$

Selanjutnya, dilakukan analisis regresi menggunakan komponen utama yang diperoleh, sehingga diperoleh hasil sebagai berikut

Table 3. Estimasi Parameter PCR

Predictor	Koefisien	Standar error	t	P-value
Konstanta	10,427	0,454	22,973	0,000
K_1	-1,783	0,226	-7,902	0,000
K_2	-3,001	0,449	-6,679	0,000

Berdasarkan nilai estimasi parameter pada Tabel 3 di atas, maka dapat dibentuk suatu model regresi dari komponen utama sebagai berikut.

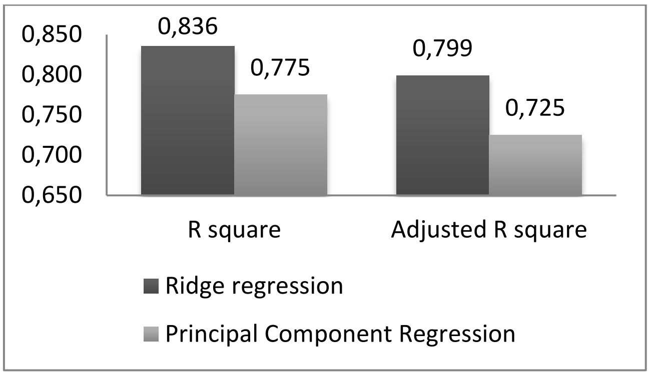
$$Y = 10{,}427 - 1{,}783K_1 - 3{,}001K_2 \tag{17}$$

Selanjutnya Persamaan (17) di transformasi ke dalam variabel awal, sehingga diperoleh model regresi sebagai berikut

$$\begin{aligned} Y = {} & 48{,}9713 + 0{,}0004X_1 - 0{,}3042X_2 + 0{,}00002X_3 + 5{,}0938X_4 - 0{,}9097X_5 \\ & -0{,}3942X_6 - 0{,}0004X_7 \end{aligned} \tag{18}$$

4. Kebaikan model

Setelah diperoleh model dari metode regresi ridge dan metode principal component regression, langkah yang terakhir yaitu membandingkan kedua model regresi dari metode-metode tersebut. Kebaikan dari model regresi akan dilihat berdasarkan nilai R^2 dan R^2_{adj}, berikut disajikan nilainya ke dalam histogram di bawah ini



Gambar 2. Nilai R^2 dan R^2_{adj} dari regresi ridge dan principal component regression

Berdasarkan histogram pada Gambar 2 di atas, dapat dilihat bahwa nilai perbandingan nilai R^2 dan R^2_{adj} regresi ridge lebih besar jika dibandingkan dengan nilai R^2 dan R^2_{adj} dari principal component regression. Hal tersebut menandakan model regresi yang dihasilkan oleh metode regresi ridge dengan estimasi parameter Kibria lebih baik dari model regresi yang dihasilkan oleh metode principal component regression.

Nilai R^2_{adj} dari metode regresi ridge adalah sebesar 79,91%, nilai tersebut menggambarkan besarnya hubungan antara variabel dependen dengan variabel independen yang menjelaskannya. Sebesar 20,09% sisanya adalah variabel lain diluar variabel independen memiliki pengaruh terhadap variabel dependen. Selanjutnya yaitu nilai R^2_{adj} yang dihasilkan dari metode principal component regression, sebesar 76,1% variabel dependen dipengaruhi oleh variabel independen dan sisanya sebesar 23,9% dipengaruhi oleh variabel lain.

KESIMPULAN

Kebaikan model regresi dalam penelitian ini dilihat dari nilai R^2 dan R^2_{adj}. Berdasarkan hasil perhitungan diperoleh bahwa keakuratan model regresi ridge sebesar 83,6%, sedangkan keakuratan model principal component regression sebesar 77,5% nilai tersebut dilihat dari besar nilai R^2. Selanjutnya nilai R^2_{adj} dari model regresi ridge lebih besar dari nilai R^2_{adj} model principal component regression yaitu $0{,}7991 > 0{,}761$. Dengan demikian dapat disimpulkan bahwa regresi ridge dengan parameter Kibria lebih baik dibandingkan dengan principal component regression dalam mengatasi multikolinieritas pada data faktor-faktor yang memengaruhi kemiskinan di Indonesia.

UCAPAN TERIMAKASIH

Pertama penulis mengucapkan terima kasih kepada Ibu yang telah mendukung dan mendo’akan dalam kelancaran penyusunan artikel ini. Selanjutnya, terima kasih penulis ucapkan kepada Universitas Mataram yang telah memberikan wadah bagi penulis untuk melakukan penelitian sehingga penelitian ini bisa diselesaikan dengan baik. Terimakasih juga penulis ucapkan kepada dosen pembimbing Ibu Lisa Harsyiah, S.Pd., M.Si. dan Bapak Zulhan Widya Baskara, S.Si., M.Si. yang telah membantu dan membimbing penulis dengan sabar dan ikhlas dalam melakukan penelitian ini.

DAFTAR PUSTAKA

Draper, N. R., & Smith, H., 1998. Applied Regression Analysis (Thrid Edition ed.), Wiley, New York.

Hastie, T., Tibshirani, R., & Friedman, J., 2008. The Elements of Statistical Learning (Second Edition ed.), Springer, California.

Herawati, N., Nisa, K., Setiawan, E., Nusyirwan, & Tiryono., 2018. Regularized Multiple Regression Methods to Deal with Severe Multicollinearity. International Journal of Statistics and Applications, 8(4).

James, G., Witten, D., Hastie, T., & Tibshirani, R., 2013. An Introduction to Statistical Learning, Springer, New York.

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W., 2005. Applied Linier Statistical Models (Fifth Edition ed.). McGraw-Hill.

Mattjik, A. A., & Sumertajaya, I. M., 2011. Sidik Peubah Ganda, IPB PRESS, Bogor.

Montgomery, D. C., Peck, E. A., & Vining, G. G., 2012. Introduction to linear regression analysis (Fifth Edition ed.). Wiley.

Younker, J., 2012. Ridge Estimation and its Modifications for Linear Regression for Linear Regression with Deterministic or Stochastic Predictors, University of Ottawa, Canada.