

PERBANDINGAN *PRINCIPAL COMPONENT REGRESSION* DAN REGRESI *RIDGE* PADA ANALISIS FAKTOR-FAKTOR INDEKS PEMBANGUNAN MANUSIA DI PROVINSI SUMATERA BARAT

COMPARISON OF *PRINCIPAL COMPONENT REGRESSION* AND *RIDGE REGRESSION* IN ANALYSIS OF HUMAN DEVELOPMENT INDEX FACTORS IN WEST SUMATERA PROVINCE

Ismi Khoiro¹, Ezhari Asfa'ani^{2§}, Syarto Musthofa³

¹Jurusan Matematika, Universitas Islam Negeri Imam Bonjol Padang, Indonesia [ismik6650@gmail.com]

²Jurusan Matematika, Universitas Islam Negeri Imam Bonjol Padang, Indonesia [ezhariasfaani@uinib.ac.id]

³Jurusan Matematika, Universitas Islam Negeri Imam Bonjol Padang, Indonesia [syartom@uinib.ac.id]

[§]Corresponding Author

Received 14th Nov 2024; Accepted 02nd Des 2024; Published 10th Des 2024;

Abstrak

Penelitian ini membahas tentang perbandingan metode *principal component regression* dan regresi *ridge* dalam mengatasi masalah multikolinearitas pada data Indeks Pembangunan Manusia (IPM) di Provinsi Sumatera Barat. Penelitian menggunakan data yang diambil dari Badan Pusat Statistika Sumatera Barat pada tahun 2023. Kedua metode tersebut akan dibandingkan berdasarkan nilai R^2 , R_{adj} , dan RSE. Nilai R^2 yang dihasilkan *principal component regression* (93,2%) > R^2 regresi *ridge* (84,89%), begitu juga dengan nilai R_{adj} *principal component regression* (92,8%) > R_{adj} regresi *ridge* (84,89%). Sedangkan nilai RSE *principal component regression* (0,2683) < RSE regresi *ridge* (0,4). Dapat disimpulkan bahwa model terbaik yang diperoleh untuk mengatasi masalah multikolinearitas pada data IPM Provinsi Sumatera Barat adalah *principal component regression*.

Kata Kunci: IPM, *Principal Component Regression*, Regresi *Ridge*

Abstract

This research discusses the method of measuring *principal component regression* and *ridge regression* in overcoming the problem of multicollinearity in Human Development Index (HDI) data in West Sumatra Province. The research uses data taken from the West Sumatra Central Statistics Agency in 2023. The two methods will be compared based on the R^2 , R_{adj} , and RSE values. The R^2 value generated by *principal component regression* (93.2%) > R^2 *ridge regression* (84.89%), as well as the R_{adj} value of *principal component regression* (92.8%) > R_{adj} *ridge regression* (84.89%). While the RSE value of *principal component regression* (0.2683) < RSE *ridge regression* (0.4). It can be concluded that the best model was obtained to overcome the problem of multicollinearity in the data West Sumatra Province's HDI is the main component of the regression.

Keywords: HDI, *Principal Component Regression*, *Ridge Regresion*

1. Pendahuluan

Indeks Pembangunan Manusia (IPM) merupakan suatu ukuran yang digunakan untuk mengukur keberhasilan pembangunan. Meskipun tidak mengukur semua dimensi dari pembangunan manusia, namun IPM dinilai mampu mengukur dimensi pokok dari pembangunan manusia. Meningkatnya nilai IPM menunjukkan bahwa terjadi peningkatan sumber daya manusia yang berkualitas pada suatu daerah. Sumber daya manusia merupakan subjek dan objek pembangunan, hal ini mengindikasikan bahwa manusia selain sebagai pelaku dari pembangunan juga merupakan sasaran pembangunan. Dengan demikian, peningkatan sumber daya manusia yang berkualitas merupakan unsur pembangunan yang sangat penting [1].

Pembangunan manusia di Provinsi Sumatera Barat terus mengalami kemajuan. Sejak tahun 2020, status pembangunan manusia Sumatera Barat sudah berada di level “tinggi”. Selama 2020–2023, IPM Sumatera Barat rata-rata meningkat sebesar 0,60 persen per tahun, dari 74,29 pada tahun 2020 menjadi 75,64 pada tahun 2023. Peningkatan IPM tahun 2023 didukung oleh semua dimensi penyusunnya. Pada tahun 2023 tiga dimensi pendukung IPM mengalami perlambatan pertumbuhan dibanding tahun 2022 [2]. Berdasarkan pada penelitian yang telah dilakukan sebelumnya, maka pada penelitian ini menggunakan delapan faktor yang berpengaruh terhadap indeks pembangunan manusia yaitu

jumlah penduduk, tingkat pengangguran terbuka, penduduk miskin, PDRB, angka harapan hidup, penduduk usia 15 tahun ke atas yang bekerja, fasilitas kesehatan dan rata-rata lama sekolah.

Faktor-faktor tersebut bisa dibentuk kedalam suatu model di bidang matematika dengan menggunakan metode analisis regresi. Hal ini dikarenakan analisis regresi merupakan alat yang digunakan untuk mengetahui hubungan antara variabel satu dengan variabel lainnya. Dalam analisis regresi dibutuhkan asumsi-asumsi klasik, asumsi-asumsi tersebut yaitu normalitas, heteroskedastisitas, tidak mengalami autokorelasi, dan tidak terjadi multikolinearitas [3]. Dalam proses menganalisa data bervariasi banyak biasanya terdapat permasalahan umum yang sering ditemui yaitu multikolinearitas [4]. Multikolinearitas yaitu adanya korelasi atau hubungan antar variabel bebas. Terindikasinya data dengan multikolinearitas akan mempengaruhi keakuratan prediksi model dan dapat membuat kesalahan dalam pengambilan keputusan. Metode yang menjadi alternatif dalam mengatasi permasalahan tersebut diantaranya *principal component regression* dan regresi *ridge* [5].

Metode *principal component regression* (PCR) merupakan salah satu analisis regresi yang menggunakan komponen utama untuk mengatasi masalah multikolinieritas pada regresi linear berganda [6]. Sedangkan regresi *ridge* merupakan metode estimasi koefisien regresi

yang diperoleh melalui penambahan konstanta bias c pada diagonal $X'X$, sehingga diperoleh persamaan regresi linear yang baru dan tidak mengandung multikolinearitas [7].

Berdasarkan penjelasan di atas penulis tertarik menganalisis “Perbandingan *Principal Component Regression* dan Regresi *Ridge* pada Analisis Faktor-Faktor Indeks Pembangunan Manusia di Provinsi Sumatera Barat tahun 2023”.

2. Landasan Teori

2.1 Regresi Linear Berganda

Analisis regresi merupakan satu metode statistik yang mempelajari tentang hubungan secara matematis antara variabel terikat (Y) dengan variabel bebas (X) [8][8]. analisis regresi linear berganda yang mengkaji lebih dari satu variabel bebas [3]. Regresi linear berganda dinyatakan dalam bentuk persamaan berikut.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (1)$$

Keterangan:

Y_i : Variabel terikat dengan $i = 1, 2, \dots, n$

X_{ij} : Variabel bebas dengan $i = 1, 2, \dots, n$; $j = 1, 2, \dots, k$

β_j : Koefisien regresi ke- j dengan $j = 1, 2, \dots, k$

ε_i : Error ke- i dengan $i = 1, 2, \dots, n$

Jika dinyatakan dalam notasi matriks menjadi:

$$Y = X\beta + \varepsilon \quad (2)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

2.2 Metode *Ordinary Least Square* (OLS)

Metode *Ordinary Least Square* (OLS) merupakan metode regresi yang meminimumkan jumlah kuadrat error [9]. Berdasarkan Persamaan (1) maka diperoleh bentuk matriks pada Persamaan (2), kemudian persamaan tersebut dapat disederhanakan menjadi:

$$Y = X\hat{\beta} + \varepsilon \quad (3)$$

atau dapat ditulis:

$$\varepsilon = Y - X\hat{\beta} \quad (4)$$

diperoleh jumlah kuadrat error yaitu:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 \quad (5)$$

Apabila diubah dalam bentuk matriks menjadi:

$$S = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_n] \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} = \varepsilon' \varepsilon \quad (6)$$

Maka berdasarkan Persamaan (4) dan (6) Sehingga didapat:

$$X'X\hat{\beta} = X'Y$$

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'Y \quad (7)$$

2.3 Uji Asumsi Klasik

a) Uji Normalitas

Uji normalitas digunakan untuk memeriksa apakah data yang diamati berdistribusi normal atau tidak [9]. Dalam penelitian ini digunakan uji *Kolmogorov-Smirnov* [10]:

Hipotesis:

H_0 : Error data berdistribusi normal

H_1 : Error data tidak berdistribusi normal

Statistik Uji:

Statistik Uji:

$$D = \text{Max}|F_0(e_i) - S_N(e_i)| \quad (10)$$

dengan,

$F_0(e_i)$: Frekuensi relatif kumulatif , $i = 1, \dots, n$

$S_N(e_i)$: Frekuensi teoritis (ekspektasi), $i = 1, \dots, n$

Keputusan:

Jika $p\text{-value} > \alpha$ maka H_0 diterima.

b) Uji heteroskedastisitas

Uji heteroskedastisitas digunakan untuk mengetahui apakah residual dari model yang diteliti mempunyai varians yang konstan atau tidak [9]. Berikut adalah pengujian hipotesis menggunakan uji Glejser:

H_0 : Tidak ada heteroskedastisitas

H_1 : Ada heteroskedastisitas

Statistik Uji :

$$t = \frac{\hat{\beta}_j}{Se(\hat{\beta}_j)} \quad (11)$$

dengan,

$\hat{\beta}_j$: Koefisien regresi ke- j , $j = 1, 2, \dots, k$

$Se(\hat{\beta}_j)$: Standar error dari koefisien regresi $\hat{\beta}_j$,
 $j = 1, 2, \dots, k$

Keputusan:

Jika $p\text{-value} > \alpha$ maka H_0 diterima.

c) Uji Autokorelasi

Uji autokorelasi digunakan untuk memeriksa apakah terdapat hubungan dalam model regresi antara residual pada periode t dengan residual pada periode $t - 1$ [11]. Pada penelitian ini menggunakan uji Durbin Watson. Berikut merupakan pengujian hipotesis [12].

Hipotesis:

H_0 : Tidak terdapat autokorelasi

H_1 : Terdapat autokorelasi

Statistik Uji:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (12)$$

dengan,

e_i : Nilai residual ke- i , $i = 1, 2, \dots, n$

e_{i-1} : Nilai residual satu periode sebelumnya

Keputusan:

Jika $0 < d < dL$ atau $4 - dL < d < 4$ artinya terdapat autokorelasi.

Jika $dU < d < 4 - dU$ artinya tidak terdapat autokorelasi.

Jika $dL \leq d \leq dU$ atau $4 - dU \leq d \leq 4 - dL$ artinya tidak bisa ditarik kesimpulan.

dengan,

dL : Batas bawah tabel Durbin Watson

dU : Batas atas tabel Durbin Watson

Apabila keputusan tidak bisa ditarik kesimpulan maka dilakukan uji lain, yakni *run test*. *Run test* merupakan bagian dari statistik non-parametrik yang digunakan untuk menguji apakah antar residual terdapat korelasi [13].

d) Uji Multikolinearitas

Multikolinearitas membuat varians dari parameter yang diestimasi lebih besar dari seharusnya sehingga tingkat akurasi dari estimasi akan menurun [14]. Terdapat beberapa akibat apabila terjadi multikolinearitas tidak sempurna antara lain:

- Standar error cenderung semakin meningkat dengan meningkatnya korelasi antar variabel bebas meskipun estimator OLS bisa ditentukan.
- Dengan nilai standar error yang besar maka interval kepercayaan parameter populasi cenderung besar.
- Apabila multikolinearitas tinggi maka kemungkinan untuk menerima hipotesis yang salah juga meningkat.

Terdapat beberapa cara untuk mendeteksi

multikolinearitas antara lain:

- Apabila R^2 tinggi, akan tetapi hanya sedikit variabel bebas yang signifikan.
- Menganalisis matrik korelasi variabel-variabel independen. Jika antar variabel independen ada korelasi yang cukup tinggi (umumnya di atas 0,90), maka hal ini merupakan indikasi adanya masalah multikolinearitas.
- Apabila nilai dari *Varian Inflation Factor* (*VIF*) lebih besar dari 10.

Dalam penelitian ini untuk mendeteksi multikolinearitas menggunakan nilai *VIF*. Adapun pengujian hipotesis sebagai berikut:

Hipotesis:

$H_0 : VIF < 10$ (Tidak terdapat multikolinearitas)

$H_1 : VIF \geq 10$ (Terdapat multikolinearitas)

Statistik Uji:

$$VIF_j = \frac{1}{1-R_j^2} \quad (13)$$

dengan,

R_j^2 : Koefisien determinasi ke- j , $j = 1, 2, \dots, k$

Keputusan:

Jika $VIF \geq 10$ maka H_0 ditolak.

2.4 Principal Component Analysis

Principal component analysis merupakan suatu teknik statistik untuk mengubah dari sebagian besar variabel asli yang saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru yang lebih kecil dan saling bebas [15]. *Principal component analysis* bergantung pada matriks kovarian (Σ) atau matriks korelasi (ρ). Matriks kovarian digunakan

jika variabel yang diamati mempunyai satuan yang sama. Akan tetapi jika variabel yang diamati mempunyai satuan berbeda maka menggunakan matriks korelasi [16].

2.5 Principal Component Regresion

Principal component regresion merupakan analisis regresi untuk mengetahui hubungan antara variabel terikat dengan sejumlah skor komponen utama sebagai variabel bebas. Penggunaan skor komponen utama sebagai variabel bebas pada PCR dilakukan untuk menyelesaikan adanya kasus multikolinieritas antar variabel bebas [15]. komponen utama merupakan kombinasi linear dari semua variabel bebas.

Pemilihan komponen utama dapat dilakukan dengan cara yaitu [17]:

- Memilih komponen utama yang mempunyai keragaman kumulatif sebesar 70 % sampai 80%.
- Memilih *eigenvalue* yang lebih besar dari satu. Kriteria ini digunakan jika berdasarkan matriks korelasi.
- Melihat *scree plot*. *Scree plot* adalah plot antara *eigenvalue* λ_j dengan j . Untuk menentukan banyaknya komponen utama dengan melihat patahan siku dari *scree plot*.

Setelah diperoleh komponen utama melalui *principal component analysis*, selanjutnya adalah meregresi komponen utama terpilih dengan variabel terikat.

2.6 Regresi Ridge

Regresi *ridge* merupakan salah satu metode untuk menstabilkan nilai koefisien regresi karena

adanya multikolinearitas yang merupakan modifikasi dari metode kuadrat terkecil. Modifikasi tersebut dilakukan dengan menambahkan tetapan bias c pada diagonal matriks yang mempengaruhi besarnya koefisien penduga ridge dan penduga yang dihasilkan adalah penduga yang bias [15]. Hoerl dan R. W. Kennard pada tahun 1962 pertama kali memperkenalkan regresi *ridge*. Regresi *ridge* bertujuan mengatasi akibat korelasi yang tinggi antara beberapa variabel bebas dalam model regresi yang membuat matriks $X'X$ hampir singular sehingga menghasilkan estimasi parameter model regresi yang tidak stabil [7]. Didefinisikan $W = XQ$ dan $\alpha = Q'\beta$ dengan Q adalah matriks berukuran $k \times k$ sedemikian sehingga $Q'Q = QQ' = I_k$ dan $(XQ)'(XQ) = \Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$, $\lambda_1, \dots, \lambda_k$ adalah *eigenvalue* dari matriks $X'X$.

sehingga diperoleh estimator regresi *ridge* dalam persamaan berikut:

$$\hat{\beta}_R = (X'X + cI)^{-1}X'Y \quad (14)$$

dan diperoleh

$$\hat{\alpha}_R = (W'W + cI)^{-1}W'Y \quad (15)$$

a) Pemusatan dan Penskalaan

Pemusatan dan penskalaan merupakan bagian dari *standardized* variabel. Pemusatan dilakukan dengan menghilangkan $\hat{\beta}_0$ sehingga membuat perhitungan model regresi menjadi lebih sederhana [18]. Sedangkan penskalaan dilakukan dengan mentransformasi variabel terikat dan variabel bebas dalam bentuk berikut:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{S_{Xj}} \quad (16)$$

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \quad (17)$$

$$S_{Xj} = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1}} \quad (18)$$

$$S_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n-1}} \quad (19)$$

dengan,

Z_{ij} : Bentuk standar dari variabel X_{ij} , $i = 1, 2, \dots, n$; $j = 1, 2, \dots, k$

Y_i^* : Bentuk standar dari variabel Y_i , dengan $i = 1, 2, \dots, n$

$\bar{Y}$: Nilai *mean* dari variabel terikat

$\bar{X}_j$: Nilai *mean* dari variabel bebas ke- j , $j = 1, 2, \dots, k$

S_{Xj} : *Standard deviation* dari X_j untuk $j = 1, 2, \dots, k$

S_Y : *Standard deviation* dari variabel terikat

Sehingga diperoleh model regresi standar yaitu

$$Y_i^* = \beta_1^* Z_{1i} + \beta_2^* Z_{2i} + \beta_3^* Z_{3i} + \dots + \beta_k^* Z_{ki} + \varepsilon_i$$

b) Estimator Parameter *Hoerl and Kennard*

Konstanta bias didefinisikan sebagai berikut [19]:

$$c_{HK} = \frac{k\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} \quad (20)$$

dengan,

c_{HK} : Konstanta bias *Hoerl and Kennard*

$\hat{\sigma}^2$: MSE dari metode OLS

$$\hat{\sigma}^2 = \frac{(Y - W\hat{\alpha})'(Y - W\hat{\alpha})}{n - k - 1} \quad (21)$$

$\hat{\alpha}$: Estimasi parameter dari metode OLS

2.7 Uji Signifikansi Parameter

Adapun uji hipotesis menggunakan uji F sebagai berikut:

Hipotesis:

$$H_0 : \beta_{R(j)}^* = 0, j = 1, \dots, 8 \quad (\text{Model regresi})$$

tidak signifikan)

$H_1 : \exists \beta_{R(j)}^* \neq 0, j = 1, \dots, 8$ (Model regresi signifikan)

Statistik uji:

$$F_{hitung} = \frac{SSR/k}{SSE/(n-k-1)} = \frac{MSR}{MSE} \quad (22)$$

dengan,

SSR : Sum of Squares Regression

SSE : Sum of Squares Error

MSE : Mean Square Error

MSR : Mean Square Regression

n : Banyaknya observasi

Keputusan:

Jika $F_{hitung} > F_{tabel}$ atau $p\text{-value} < \alpha$ maka H_0 ditolak.

Adapun pengujian hipotesis menggunakan uji t sebagai berikut [5]:

Hipotesis:

$H_0 : \beta_{R(j)}^* = 0$ untuk $j = 1, \dots, 8$ (Koefisien regresi ke- j tidak signifikan)

$H_1 : \exists \beta_{R(j)}^* \neq 0$ untuk $j = 1, \dots, 8$ (Koefisien regresi ke- j signifikan)

Statistik uji:

$$t_{hitung} = \frac{\hat{\beta}_{R(j)}^*}{Se(\hat{\beta}_{R(j)}^*)} \quad (23)$$

dengan,

$$Se(\hat{\beta}_{R(j)}^*) = \sqrt{\frac{s_d^2}{n}} \quad (24)$$

dan

$$s_d^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2 \quad (25)$$

di mana,

$\hat{\beta}$: Estimasi parameter

Se : Standar error

s_d^2 : Variansi sampel

n : Banyaknya observasi

z_i : Variabel acak untuk $i = 1, 2, \dots, n$

$\bar{z}$: Rata-rata sampel

Keputusan:

Jika $|t_{hitung}| > t_{tabel}$ atau $p\text{-value} < \alpha$ maka H_0 ditolak.

2.8 Kebaikan Model

a) Koefisien Determinasi

Adapun rumus koefisien determinasi yaitu:

$$R^2 = 1 - \frac{SSE}{SST} \quad (26)$$

dengan,

SST : Sum Of Squaes Total

SSE : Sum of Squares Error

Menambahkan banyak variabel bebas kedalam model maka akan meningkatkan (R^2). Karena (R^2) sering dibuat besar dengan menambahkan variabel bebas maka disarankan agar ukuran ini dimodifikasi. Untuk mengoreksi R^2 digunakan koefisien determinasi ganda terkoreksi (*adjusted coefficient multiple determination*). Adapun rumus R_{adj} sebagai berikut:

$$R_{adj} = 1 - \frac{SSE/(n-k)}{SST/(n-1)} \quad (27)$$

b) Residual Standard Error (RSE)

Residual Standard Error (RSE) merupakan estimasi standar deviasi dari ε . ε adalah selisih dari nilai yang diprediksi dengan nilai yang diamati. RSE didefinisikan sebagai berikut [20].

$$RSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-k-1}} = \sqrt{MSE} \quad (28)$$

3 Hasil Dan Pembahasan

3.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik Provinsi Sumatera Barat Tahun 2023 dengan delapan variabel bebas yang berpengaruh terhadap IPM. Adapun deskripsi data sebagai berikut:

Tabel 1. Deskripsi Statistika

Variabel	Mean	Standar deviasi	Maks	Min
Y	75,15	4,62	92500 0	5872 9,00
X ₁	30003 0,16	222154,5 2	10,86	1,33
X ₂	5,11	1,86	13,72	2,27
X ₃	5,90	2,39	55377 ,42	2019 2,61
X ₄	34798 ,32	10205,82	75,13	65,10
X ₅	70,95	2,75	47900 0	2860 6,00
X ₆	15918 0,53	115179,9 2	25,00	4,00
X ₇	14,68	7,41	11,94	7,76
X ₈	9,50	1,44	83,98	65,79

Nilai *mean* dan standar deviasi dari masing-masing variabel pada Tabel 1 akan digunakan untuk perhitungan standarisasi data dan perhitungan metode pemusatan dan penskalaan. Berdasarkan Tabel 1 diketahui bahwa satuan variabel yang digunakan berbeda sehingga rentang data yang diperoleh cukup jauh sehingga dilakukan standarisasi.

1. Uji Asumsi Klasik

a) Uji Normalitas

Dengan menggunakan taraf signifikan $\alpha = 5\%$ dan kriteria uji yaitu jika $p\text{-value} > \alpha$ maka H_0 diterima. Hasil uji Kolmogorov

Smirnov dengan $p\text{-value}$ diperoleh sebesar 0,793. Hal ini menunjukkan bahwa $p\text{-value} > \alpha$ maka H_0 diterima, artinya *error* data berdistribusi normal. Dengan demikian asumsi normalitas terpenuhi.

b) Uji Heteroskedastisitas

hasil uji Glejser yang menunjukkan bahwa semua variabel bebas memiliki $p\text{-value} > \alpha$ maka dapat disimpulkan bahwa H_0 diterima yang artinya tidak ada heteroskedastisitas.

c) Uji Autokorelasi

Pengujian autokorelasi pada penelitian ini menggunakan uji Durbin Watson. Dari hasil uji autokorelasi diketahui bahwa $dL \leq d \leq dU$ artinya tidak bisa ditarik kesimpulan. Karena tidak dapat ditarik kesimpulan maka dilakukan uji lain yaitu *run test*. Adapun hipotesisnya sebagai berikut:

H_0 : Tidak ada autokorelasi

H_1 : Ada autokorelasi

Dengan menggunakan taraf signifikan $\alpha = 5\%$ dan kriteria uji yaitu jika nilai R_{obs} berada diantara nilai kritis yang diberikan tabel F_1 dan F_2 ($F_1 < R_{obs} < F_2$), maka H_0 diterima. Hasil dari uji *run test* dengan $F_1 = 5 < R_{obs} = 10 < F_2 = 16$, sehingga H_0 diterima artinya tidak ada autokorelasi.

d) Uji Multikolinearitas

Pengujian multikolinearitas pada penelitian ini dengan melihat nilai *VIF*. Adapun hipotesis yang digunakan sebagai berikut:

H_0 : $VIF < 10$ (Tidak ada multikolinearitas)

H_1 : $VIF \geq 10$ (Ada multikolinearitas)

Tabel 2. Uji Multikolinearitas

Variabel	VIF	Keputusan
Z_1	215,617	Ada Multikolinearitas
Z_2	5,412	Tidak Ada Multikolinearitas
Z_3	3,097	Tidak Ada Multikolinearitas
Z_4	5,441	Tidak Ada Multikolinearitas
Z_5	4,442	Tidak Ada Multikolinearitas
Z_6	193,308	Ada Multikolinearitas
Z_7	10,498	Ada Multikolinearitas
Z_8	16,885	Ada Multikolinearitas

Pada Tabel 2 diketahui bahwa variabel jumlah penduduk (Z_1), penduduk usia 15 tahun ke atas yang bekerja (Z_6), fasilitas kesehatan (Z_7) dan rata-rata lama sekolah (Z_8) terdapat multikolinearitas di mana nilai $VIF \geq 10$, maka dapat dikatakan model regresi memiliki multikolinearitas. Untuk mengatasi masalah tersebut maka digunakan metode *principal component regression* dan regresi *ridge*.

3.2 *Principal Component Regression*

a) Standarisasi Data

Dalam penelitian ini data yang digunakan memiliki satuan yang berbeda maka langkah pertama yaitu mentransformasi variabel bebas kedalam variabel standar.

b) Menghitung *eigenvalue* dari matriks korelasi

Untuk menentukan jumlah komponen utama yang terbentuk adalah dengan melihat *eigenvalue* dari matriks korelasi yang lebih besar dari satu.

c) Menghitung Skor Komponen Utama

Tabel 3. Skor Komponen Utama

Variabel	K_1	K_2
Z_1	0,631	0,758
Z_2	-0,002	0,919
Z_3	0,622	-0,507
Z_4	-0,750	0,348
Z_5	-0,604	0,617
Z_6	0,646	0,744
Z_7	0,863	0,425
Z_8	-0,860	0,435

Berdasarkan Tabel 3, diperoleh persamaan *principal component regression* sebagai berikut:

$$K_1 = 0,631 Z_1 - 0,002 Z_2 + 0,622 Z_3 - 0,750 Z_4 - 0,604 Z_5 + 0,646 Z_6 + 0,863 Z_7 - 0,860 Z_8$$

$$K_2 = 0,758 Z_1 + 0,919 Z_2 - 0,507 Z_3 + 0,348 Z_4 + 0,617 Z_5 + 0,744 Z_6 + 0,425 Z_7 + 0,435 Z_8$$

selanjutnya K_1 dan K_2 atau variabel baru komponen utama sudah bisa digunakan untuk analisis regresi sehingga diperoleh model *principal component regression* sebagai berikut:

$$Y = 75,154 + 4,464 K_1 + 0,323 K_2$$

Setelah diperoleh model Persamaan selanjutnya, dilakukan uji F dan uji t untuk mengetahui signifikansi model regresi.

Sehingga diperoleh persamaan *principal component regression* sebagai berikut:

$$\hat{Y} = 165,1391 + 0,00001 X_1 - 0,0049 X_2 + 1,1951 X_3 - 0,00033 X_4 - 1,0088 X_5 + 0,000025 X_6 + 0,5342 X_7 - 2,7409 X_8.$$

3.3 Regresi Ridge

a) Estimasi Parameter *Hoerl and Kennard*

Untuk menentukan tetapan bias c dengan menggunakan estimator parameter *Hoerl and Kennard* pada Persamaan (2.46). Sehingga diperoleh $\hat{\alpha}_R$, $\hat{\beta}_R^*$, dan VIF yang disajikan pada tabel dibawah ini.

Tabel 4. Estimasi Parameter *Hoerl and Kennard*

Nilai c_{HK}	$\hat{\alpha}_R$	$\hat{\beta}_R^*$	VIF
0,5876	-0,1019	-0,0445	1,5648
	-0,4060	0,1909	5,2853
	0,0439	-0,1333	3,0027
	0,1551	0,0137	4,1079
	-0,0931	0,2171	3,7127
	-0,0044	0,0094	4,9867
	0,3711	0,0699	9,4133
	0,3773	0,6176	1,6780

Pada Tabel 4 diperoleh nilai $c_{HK} = 0,5876$ dan nilai VIF kurang dari 10. Dengan demikian, dapat disimpulkan bahwa estimasi dengan metode *Hoerl and Kennard* tidak mengandung multikolinearitas.

Setelah diperoleh estimasi parameter metode *Hoerl and Kennard*, selanjutnya dilakukan uji F dan uji t untuk mengetahui signifikansi model regresi.

b) Transformasi Parameter Regresi Ridge

Nilai estimasi parameter regresi ridge standar yang telah diperoleh ditransformasi ke bentuk awal sehingga diperoleh hasil sebagai berikut:

$$\begin{aligned}
 \hat{\beta}_{R(8)} &= \left(\frac{S_y}{S_x} \right) \hat{\beta}_{R(8)}^* \\
 &= \left(\frac{4,62367}{1,43899} \right) (0,9175) \\
 &= 2,9480 \\
 \hat{\beta}_{R(0)} &= \bar{Y} - \hat{\beta}_{R(8)} \bar{X}_8 = 47,1342
 \end{aligned}$$

Sehingga diperoleh model regresi ridge menggunakan metode *Hoerl and Kennard* sebagai berikut:

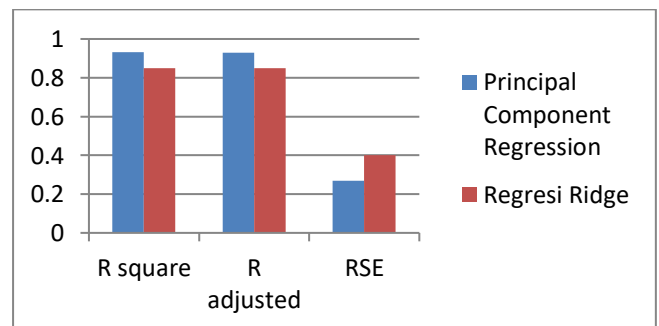
$$\hat{Y}_R = 47,1342 + 2,9480X_8$$

3.4 Kebaikan Model

Model yang telah dilakukan uji signifikansi parameter akan dilihat kebaikan model berdasarkan nilai R^2 , R_{adj} dan RSE.

Tabel 5. Kebaikan Model

Metode	R^2	R_{adj}	RSE
<i>Principal Component Regression</i>	0,932	0,928	0,2683
Regresi Ridge	0,8489	0,8489	0,400



Gambar 1. Kebaikan Model

4 Kesimpulan Dan Saran

a) Nilai R^2 yang dihasilkan *principal component regression* (93,2%) > R^2 regresi ridge (84,89%), begitu juga dengan nilai R_{adj} *principal component regression* (92,8%) > R_{adj} regresi ridge (84,89%). Sedangkan nilai RSE *principal component regression* (0,2683) < RSE regresi ridge (0,4).

b) Dapat disimpulkan bahwa metode *principal component regression* lebih baik dibandingkan metode regresi ridge dalam

menganalisis faktor-faktor indeks pembangunan manusia di Provinsi Sumatera Barat tahun 2023.

Saran

Disarankan untuk penelitian selanjutnya agar menambahkan variabel bebas lainnya seperti pertumbuhan ekonomi, anggaran pendapatan dan belanja daerah, penanaman modal dalam negeri, infrastruktur dan lainnya yang mungkin mempengaruhi IPM dan menggunakan perbandingan metode lain.

5 Ucapan Terima Kasih

Pada artikel ini penulis ucapkan terimakasih kepada BPS Sumatera Barat dan Prodi Matematika Fakultas Sains dan Teknologi UIN Imam Bonjol Padang.

DAFTAR PUSTAKA

- [1] M. Muliza, T. Zulham, and C. Seftarita, "Analisis Pengaruh Belanja Pendidikan, Belanja Kesehatan, Tingkat Kemiskinan dan PDRB terhadap IPM di Provinsi Aceh," *J. Perspekt. Ekon. Darussalam*, pp. 51–69, 2017.
- [2] "Badan Pusat Statistika Provinsi Sumbar (2023). Indeks Pembangunan Manusia (IPM) Provinsi Sumatera Barat Tahun 2023," Badan Pusat Statistika Cianjur.
- [3] Irwan and Hasriani, "Perbandingan Regresi Ridge dan Principal Component Analysis dalam Mengatasi Multikolinearitas," *J. Teknosains*, pp. 125–135, 2016.
- [4] H. . Sinaga and D. Saputro, "Performa Metode Elastic-Net dalam Kasus Multikolinearitas pada Analisis Linear Berganda," *Pros. Semin. Pendidik. Mat. dan Mat.*, p. 3(2721), 2021.
- [5] I. Fiqriah, S. MArtha, and D. Kusnandar, *Penerapan Regresi Ridge Robust-M dalam Mengatasi Multikolinearitas dan Pencilan pada Data Stunting di Indonesia*. 2024.
- [6] E. Supriyadi, S. Mariani, and Sugiman, "Perbandingan Metode Partial Least Square dan Principal Component Regression untuk Mengatasi Multikolinearitas pada Model Regresi Linear Berganda," *Unnes J. Math.*, vol. 6, no. 2, pp. 117–128, 2017.
- [7] T. L. Wasilaine, M. W. Talakua, and Y. A. Lesnussa, "Model Regresi Ridge Untuk Mengatasi Model Regresi Linier Berganda Yang Mengandung Multikolinieritas," *J. Ilmu Mat. Dan Terap.*, vol. 8, no. 1, 2014.
- [8] L. H. Hasibuan, F. Yanuar, D. Devianto, and M. Maiyastri, "Quantile Regression Analysis; Simulation Study With Violation of Normality Assumption," *JOSTECH J. Sci. Technol.*, vol. 4, no. 2, pp. 133–142, 2024.
- [9] A. Karim, "Seminar Nasional Pendidikan, Sains dan Teknologi Perbandingan Metode Ordinary Least Square (Ols) Dan Metode Regresi Robust Pada Hasil Produksi Padi di Kabupaten Indramayu," *Semin. Nas. Pendidikan, Sains dan Teknol. ISBN 978-602-61599-6-0 Fak. Mat. dan Ilmu Pengetah. Alam Univ. Muhammadiyah Semarang*, pp. 219–223, 2017.
- [10] A. Quraisy, "Normalitas Data Menggunakan Uji Kolmogorov-Smirnov dan Saphiro-Wilk," *J-HEST J. Heal. Educ. Econ. Sci. Technol.*, vol. 3, no. 1, pp. 7–11, 2022.
- [11] N. Y. Aditiya, E. S. Evani, and S. Maghfiroh, "Konsep Uji Asumsi Klasik Pada Regresi Linier Berganda," *J. Ris. Akunt. Soedirman*, vol. 2, no. 2, pp. 102–110, 2023.
- [12] R. A. Indhova, S. Suparti, and A. R. Hakim, "Perbandingan Gulud Regression Dan Principal Component Regression Terhadap Pemodelan Indeks Pembangunan Manusia Provinsi Jawa Timur," *J. Gaussian*, vol. 12, no. 2, pp. 199–208, 2023.
- [13] Z. Istianto and M. Muji, "Pengaruh Corporate Social Responsibility, Ukuran Perusahaan, Leverage, Profitabilitas, dan Capital Intensity Terhadap Agresivitas Pajak (Pada Perusahaan Manufaktur yang terdaftar di BEI Periode 2015-2019)," *Bus. Econ. Conf. Util. Mod. Technol.*, pp. 22662–9404, 2020.
- [14] R. A. Gunawan, D. Prasetya Zulkarmain, and D. Arianto, S. T., Ekonomi, F., Bisnis, "Socius: Jurnal Penelitian Ilmu-Ilmu Sosial Perbandingan Metode Ordinary Least Square

- dan Metode Partial Least Square Untuk Mengatasi Multikolinearitas,” *J. Ilm. Bid. Sos. Ekon. Budaya, Ekol. Dan Pendidik.*, vol. 1, no. 6, pp. 795–808, 2023.
- [15] E. A. Fauzan and S. Soehardjoepri, “Perbandingan Metode Regresi Komponen Utama dan Regresi Ridge pada Analisis Pengaruh Rasio Keuangan Profitabilitas terhadap Harga Saham Perusahaan Perbankan,” *J. Sains Dan Seni ITS*, vol. 11, no. 6, 2023.
- [16] S. Sakdiyah and R. Subekti, “Optimalisasi Jumlah Kelompok Pada Metode Single Linkage dan Complete Linkage Menggunakan Indeks Davies Bouldin,” *J. Kaji. Dan Terap. Mat.*, vol. 7, no. 1, pp. 20–30, 2018.
- [17] P. Pendi, “Analisis Regresi Dengan Metode Komponen Utama Dalam Mengatasi Masalah Multikolinearitas,” *Bimaster Bul. Ilm. Mat. Stat. Dan Ter.*, vol. 10, no. 1, pp. 131–138, 2021.
- [18] N. A. D. K. Naomi Nessyana Debataraja, “Metode Generalized Ridge Regression Dalam Mengatasi Multikolinearitas,” *Bimaster Bul. Ilm. Mat. Stat. Dan Ter.*, vol. 8, no. 4, pp. 679–686, 2019.
- [19] N. N. Anggraeni, W. R., Debataraja and S. W. Rizki, “Estimasi Parameter Regresi Ridge Untuk Mengatasi Multikolinearitas,” *Bul. Ilm. Mat. Stat. Dan Ter.*, vol. 07, no. 4, pp. 295–302, 2018.
- [20] T. Faizia, A. Prahutama, and H. Yasin, “Pemodelan Indeks Pembangunan Manusia Di Jawa Tengah Dengan Regresi Komponen Utama Robust,” *J. Gaussian*, vol. 8, no. 2, pp. 253–271, 2019.