p-ISSN 2085-7829
e-ISSN 2798-3455

Mengatasi Multikoliniearitas Dalam Regresi Linier Berganda Menggunakan Principal Component Analysis

Overcoming Multicollinearity in Multiple Linear Regression Using Principal Component Analysis

Niken Harel Chairunnisa[1a)], Darnah Andi Nohe[2], Syaripuddin[3]
[1,2] Program Studi Statistika, FMIPA Universitas Mulawarman, Indonesia
[3] Program Studi Matematika, FMIPA Universitas Mulawarman, Indonesia
[1,2] Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman, Indonesia

[a)]Corresponding author: nikenharel@gmail.com

ABSTRACT
Multiple linear regression analysis has assumptions that must be met, one of which is multicollinearity. Multicollinearity occurs when the independent variables correlate with each other, resulting in the regression coefficient produced by multiple linear regression analysis being very weak or unable to provide analysis results that represent the nature or influence of the independent variable concerned. The detection of multicollinearity can be known through the VIF value. In this study, human development index data on Kalimantan Island in 2019 detected multicollinearity because some independent variables have a VIF value of more than 10 so that the method used to overcome multicollinearity in this study is Principal Component Analysis (PCA). Based on the results of research using the Principal Component Regression method, There are five independent variables that influence the IPM that is Percentage of Poor Population, Number of Health Workers, Number of Workforce, Number of High Schools, and Number of High School Teachers.

Keywords*: multicollinearity, multiple linear regression, PCA*

1. Pendahuluan

Analisis regresi merupakan suatu metode analisis statistik yang bertujuan untuk mengetahui pengaruh antara variabel bebas dengan variabel terikat. Variabel bebas disimbolkan dengan *X*, sedangkan variabel terikat disimbolkan dengan *Y* (Alfigari, 2000). Model regresi dalam statistika dikatakan cocok atau baik jika garis regresi yang dihasilkan untuk melakukan estimasi atau prediksi dari sebaran data menghasilkan *error* terkecil. Syarat yang diperlukan untuk melakukan analisis regresi harus dipenuhi berbagai asumsi klasik, antara lain data tidak mengalami autokorelasi, heteroskedastisitas, dan multikolinearitas.

Multikolinearitas terjadi apabila terdapat hubungan atau korelasi diantara beberapa atau seluruh variabel bebas (Soemartini, 2008). Efek dari multikolinearitas ini dapat mengakibatkan estimasi parameter regresi yang dihasilkan dari analisis regresi linear berganda menjadi tidak efisien karena dapat menyebabkan regresi berganda mempunyai bias dan varians yang besar. multikolineritas memberikan pengaruh terhadap standar *error* estimasi koefisien regresi sehingga hasil estimasi dimungkinkan tidak akurat (Sungkono, 2017). Masalah multikolinearitas dapat menyebabkan uji T menjadi tidak signifikan, hal tersebut yang menyebabkan hasil analisis yang dilakukan menjadi tidak sejalan atau bertentangan. Jika ada data yang terdapat multikolineritas berarti salah satu asumsi klasik regresi linier berganda dilanggar maka kesimpulan yang didapat dari hasil pengujian untuk model regresi linier berganda maupun untuk masing-masing variabel yang ada dalam model seringkali tidak tepat (Montgomery & Peck, 1991).

Ada beberapa metode yang dapat digunakan untuk mengatasi masalah multikolinearitas, salah satunya adalah metode *Principal Component Analysis* (PCA). Metode PCA pada dasarnya bertujuan untuk menyederhanakan variabel yang diamati dengan cara menyusutkan (mereduksi) dimensinya. Hal ini dilakukan dengan cara menghilangkan korelasi diantara variabel bebas melalui transformasi variabel bebas asal ke variabel baru yang tidak berkorelasi sama sekali atau yang biasa disebut dengan komponen utama. Sriningsih, dkk (2018) dalam Penanganan Multikolinieritas dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras di Provinsi Sulut, diperoleh lima dari sepuluh variabel bebas terdeteksi multikolinieritas dan setelah dilakukan PCA seluruh variabel bebas memiliki nilai VIF kurang dari 10 (tidak terdeteksi multikolinieritas).

2. Tinjauan Pustaka

2.1 Analisis Regresi

Analisis regresi adalah metode analisis yang bertujuan untuk mengetahui pengaruh variabel bebas terhadap variabel terikat. Menurut Drapper dan Smith (1992), analisis regresi merupakan metode yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan tentang hubungan ketergantungan variabel terhadap variabel lainnya.

Analisis regresi linier berganda merupakan suatu metode yang mengindentifikasi pengaruh antara dua atau lebih variabel bebas dengan satu variabel terikat. Model regresi umum dengan k variabel *bebas* dapat dituliskan sebagai berikut:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

Estimasi parameter model regresi berganda pada persamaan (1) dapat diperoleh dengan menggunakan metode kuadrat terkecil (*Ordinary Least Square/ OLS*) mendapatkan $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$,..., $\hat{\beta}_k$ yang *Best Linear Unbiased Estimator* (BLUE) sehingga menyebarnya persamaan regresi sedekat mungkin pada data aktualnya (Widarjono, 2007). Prinsip dasar metode OLS adalah meminimumkan jumlah kuadrat *error*, untuk mendapatkan nilai minimum dari fungsi maka syaratnya adalah differensiasi atau turunan pertama dari fungsi tersebut harus sama dengan nol. Nilai parameter $\hat{\beta}$ didapatkan dengan meminimumkan jumlah kuadrat *error*, yaitu sebagai berikut:

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \quad (2)$$

2.2 Multikolinieritas

Salah satu asumsi dari model regresi linier berganda adalah tidak ada hubungan linier antara variabel bebas. Jika ada satu atau lebih hubungan tersebut antara variabel bebas maka disebut multikolinearitas. Ketika terdapat multikolinearitas pada variabel bebas maka keputusan secara statistiknya menjadi lemah (Gujarati, dkk., 2015).

1. Nilai R^2 yang tinggi tapi hanya ada sedikit nilai *t* yang signifikan. Nilai *t* yang tidak signifikan dapat terjadi dikarenakan adanya multikolinearitas pada variabel bebas.
2. Korelasi tinggi diantara variabel bebas.
3. Nilai *Tolerance* (*TOL*) mendekati nol menandakan terdapat multikolinearitas. Nilai *TOL* adalah invers dari nilai *Variance Inflation Factor* (*VIF*).
4. Nilai *VIF* yang lebih dari 10.

Koefisien determinasi R^2 biasanya dituliskan dalam bentuk persentase (100%) yang menyatakan variansi total nilai-nilai variabel terikat yang dapat dijelaskan oleh variabel bebas melalui hubungan linier. Semakin besar nilainya semakin baik persamaan regresi itu dalam menjelaskan kovariansi data. Nilai koefisien determinasi R^2 diperoleh dengan menggunakan persamaan sebagai berikut:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}-\bar{y})^2}{\sum_{i=1}^{n}(y-\bar{y})^2} \quad (3)$$

Rumus VIF dapat dituliskan sebagai berikut.

$$VIF = \frac{1}{1-R^2} \quad (4)$$

2.3 *Principal Component Analysis* (PCA)

PCA merupakan suatu teknik statistik untuk mengubah dari sebagian besar variabel asli yang digunakan yang saling berkorelasi satu dengan yang lainnya menjadi satu set variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi lagi). Komponen utama tergantung sepenuhnya pada matriks varian kovarian yang di simbolkan dengan $\sum$ (atau matriks korelasi ρ) dari komponen utama peubah-peubah $X_1, X_2, X_3, ..., X_k$. Matriks kovarian digunakan jika variabel yang diamati mempunyai satuan yang sama. Akan tetapi, jika variabel yang diamati mempunyai satuan yang berbeda maka menggunakan matriks korelasi.

Komponen utama yang didasarkan pada matriks kovarian lalu dibentuk variabel baru yaitu W adalah komponen utama yang merupakan kombinasi linier dari $X_1, X_2, X_3, ..., X_k$ dengan *eigen value* $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$ Pengembangan komponen utama tidak memerlukan asumsi multivariat normal. Apabila variansi dari variabel-variabel yang diamati mempengaruhi besarnya bobot atau koefisien komponen utamanya maka analisis komponen utama dapat dilakukan menggunakan matriks variansi-kovariansi. Matriks Varians Kovarian berisi variansi dan kovariansi dari data yang berukuran m x m. Diagonal Matriks Varian Kovarian merupakan variansi dari data, sedangkan elemen lainnya merupakan kovarian data.

Varian Kovarian dapat diperoleh menggunakan persamaan sebagai berikut

$$S_{rk} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ri} - \bar{X}_r)(X_{ki} - \bar{X}_k) \quad (4)$$

Selain matriks varian kovariansi, komponen utama juga dapat dibentuk berdasarkan matriks korelasi dengan persamaan sebagai berikut:

$$\rho = \left(V^{\frac{1}{2}}\right)^{-1} \Sigma \left(V^{\frac{1}{2}}\right)^{-1} \quad (5)$$

Ada tiga metode umum yang digunakan untuk menentukan banyaknya komponen utama yang dapat digunakan sebagai variabel baru yaitu:

1. Berdasarkan proporsi kumulatif total keragaman yang mampu di jelaskan oleh k komponen utama minimal 80%, dan proporsi total variansi populasi bernilai cukup besar. Metode ini di terapkan pada

p-ISSN 2085-7829
e-ISSN 2798-3455

matriks korelasi ataupun matriks kovarian.

2. Berdasarkan nilai eigen dari komponen utama. Tapi hanya bisa diterapkan pada matriks korelasi. Jika nilai eigen lebih atau sama dengan satu.
3. Berdasarkan *scree plot*. Dengan menggunakan metode ini banyaknya komponen utama yang di pilih yaitu k, adalah jika pada titik k tersebut plotnya curam ke kiri tetapi tidak curam ke kanan. Ide yang ada di belakang metode ini adalah bahwa banyaknya komponen utama yang dipilih sedemikian rupa sehingga selisih antara akar ciri yang berurutan sudah tidak besar lagi.

2.4 Regresi Komponen Utama

Pada dasarnya regresi komponen utama merupakan teknik analisis regresi yang dikombinasikan dengan teknik analisis komponen utama, dimana analisis komponen utama dijadikan sebagai tahap analisis antara. Persamaan regresi komponen utama berdasarkan matriks kovarian pada dasarnya sama dengan persamaan regresi komponen utama berdasarkan matriks korelasi yaitu X_1, X_2, …, X_k diganti dengan variabel baku Z_1, Z_2, …, Z_k dengan rumus sebagai berikut:

$$Z = \left(V^{\frac{1}{2}}\right)^{-1}(X - \bar{X}) \tag{6}$$

Apabila diberikan notasi W_1, W_2, …, W_k sebagai banyaknya komponen utama yang dilibatkan dalam analisis regresi komponen utama, dimana k lebih kecil daripada banyaknya variabel penjelas asli X, yaitu sejumlah p ($k < p$). Maka bentuk umum persamaan regresi komponen utama adalah:

$$Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \varepsilon \tag{7}$$

2.5 Indeks Pembangunan Manusia

Indikator Pembangunan Manusia dapat diidentifikasi dan diinventarisasi berdasarkan definisi dari pembangunan manusia itu sendiri (Azahari, 2000). Dalam hal ini, definisi pembangunan manusia dapat merujuk pada terminologi yang beragam. Pembangunan manusia adalah sebuah proses dalam rangka meningkatkan kebebasan dan peluang seseorang serta meningkatkan kesejahteraan mereka, memungkinkan mereka menjalani hidup yang panjang dan sehat; untuk memiliki akses ke pengetahuan; untuk menikmati standar hidup yang layak; dan untuk berpartisipasi dalam keputusan yang mempengaruhi mereka (Khodabakhshi, 2011).

3. Metodologi Penelitian

Penelitian ini dilakukan dengan mengumpulkan data Indeks Pembangunan Manusia (IPM) di Pulau Kalimantan tahun 2019 yang diterbitkan oleh Badan Pusat Statistik (BPS).

Tahapan analisis data pada penelitian ini adalah sebagai berikut.

a. Melakukan analisis statistika deskriptif.
b. Menentukan model Regresi linear berganda 5 variabel bebas dengan menggunakan metode OLS.
c. Mendeteksi multikolinieritas dengan dengan melihat Nilai VIF.
d. Melakukan langkah-langkah metode PCA untuk mengatasi adanya Multikolinieritas;
e. Membentuk persamaaan regresi komponen utama berdasarkan persamaan (7).
f. Menarik kesimpulan dari hasil penelitian.

4. Hasil dan Pembahasan

Hasil Analisis statitika deskriptif data IPM (Y), Persentase Penduduk Miskin (X_1), Jumlah Tenaga Kesehatan (X_2), Jumlah Angkatan Kerja (X_3), Jumlah Sekolah Menengah Atas (X_4), dan Jumlah Guru Sekolah Menengah Atas (X_5) disajikan pada Tabel 1 berikut.

Tabel 1. Statistika Deskriptif

Variabel	Minimum	Rata-Rata	Maksimum	Standar Deviasi
$\boldsymbol{Y}$	65,49	71,53	80,20	3,90
$\boldsymbol{X_1}$	2,42	6,13	12,38	2,33
$\boldsymbol{X_2}$	265,00	1.150,20	3009,00	629,18
$\boldsymbol{X_3}$	13.203,00	145.660,00	428.353,00	93.851,03
$\boldsymbol{X_4}$	3,00	20,79	62,00	12,49

Berdasarkan Tabel 1. diperoleh bahwa rata-rata IPM adalah 71,53 dengan simpangan baku 3,90. IPM terendah berada di Kab. Hulu Sungai Utara Prov. Kalimantan Selatan yakni sebesar 65,49 dan IPM tertinggi berada di Kab. Sanggau Provinsi Kalimantan Barat yakni sebesar 80,20.

4.1 Pendeteksian Multikolinieritas

p-ISSN 2085-7829
e-ISSN 2798-3455

Pendeteksian multikolinieritas bertujuan untuk mengetahui apakah terdapat hubungan linier antar variabel bebas dalam model regresi. Berikut model regresi yang dihasilkan pada penelitian ini.

1. Model Awal

Berikut model awal regresi linier berganda pada penelitian ini:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (8)$$

2. Estimasi Parameter

Estimasi parameter untuk model regresi berganda variabel bebas terhadap Variabel Terikat. Setelah dilakukan estimasi parameter didapatkan model sebagai berikut.

$$\hat{Y} = 6{,}857 \times 10^{1} + 8{,}723 \times 10^{-2} X_1 + 3{,}051 \times 10^{-4} X_2 + 2{,}102 \times 10^{-5} X_3 - 7{,}475 \times 10^{-2} X_4 + 1.347 \times 10^{-3} X_5 + \varepsilon \quad (9)$$

Berdasarkan model pada persamaan (9), diketahui bahwa tanpa dipengaruhi oleh variabel bebas, maka nilai IPM adalah sebesar $6{,}587 \times 10^{1}$. Setiap kenaikan persentase penduduk miskin sebesar 1% maka mengakibatkan IPM cenderung naik sebesar $8{,}723 \times 10^{-2}$. Setiap kenaikan jumlah tenaga kesehatan sebanyak 1 orang maka mengakibatkan IPM cenderung naik sebesar $3{,}051 \times 10^{-4}$. Setiap kenaikan jumlah angkatan kerja sebanyak 1 orang maka mengakibatkan IPM cenderung naik sebesar $2{,}102 \times 10^{-5}$. Setiap kenaikan jumlah sekolah menengah atas sebanyak 1 unit maka mengakibatkan IPM cenderung turun sebesar $7{,}475 \times 10^{-2}$. Setiap kenaikan persentase jumlah guru sekolah menengah atas sebanyak 1 orang maka mengakibatkan IPM cenderung naik sebesar $1{,}347 \times 10^{-3}$.

Setelah dilakukan pemodelan, maka bisa dilakukan perhitungan nilai VIF yang dihitung berdasarkan persamaan (3), dimanavariabel bebas dinyatakan tidak terdapat multikolinearitas apabila VIF<10. Hasil nilai VIF dapat dilihat pada Tabel 2.

Tabel 2. Nilai VIF Masing-Masing Variabel

Variabel	VIF
X_1	1,23274
X_2	4,96152
X_3	4,70483
X_4	11,07813
X_5	18,89664

Berdasarkan Tabel 2, dapat diindikasikan bahwa terdapat kasus multikolinieritas pada variabel jumlah sekolah menengah atas (X_4) dan jumlah guru sekolah menengah atas (X_5) karena memiliki nilai VIF lebih dari 10 yang termasuk dalam multikolinieritas tinggi. Setelah terdapat variabel yang terindikasi terjadi multikolinieritas, maka perlu diatasi terlebih dahulu sebelum masuk ke metode regresi linier berganda. Pada penelitian ini, data yang terindikasi multikolinieritas diatasi dengan menggunakan metode PCA.

4.2 Mengatasi Multikolinieritas dengan PCA

Dalam mengatasi kasus multikolinieritas pada regresi menggunakan metode PCA terdapat beberapa langkah-langkah, yakni sebagai berikut:

4.3 Membuat Matriks Variansi-Kovariansi

Matriks varian kovarian merupakan matriks yang berisikan varian dari setiap variabel terikat dan kovariansi antar variabel terikat. Berikut langkah-langkah perhitungan variansi kovariansi:

a. Menghitung nilai rata-rata.

$$\bar{X}_1 = \sum_{i=1}^{n} \frac{\bar{X}_{1i}}{n} = \frac{X_{1,1} + X_{1,2} + \cdots + X_{1,56}}{56}$$
$$= \frac{8{,}95 + 9{,}09 + \cdots + 4{,}91}{56}$$
$$= 6{,}13$$

$$\bar{X}_2 = \sum_{i=1}^{n} \frac{\bar{X}_{2i}}{n} = \frac{X_{2,1} + X_{2,2} + \cdots + X_{2,56}}{56}$$
$$= \frac{1.157 + 803 + \cdots + 1.980}{56}$$
$$= 1.150{,}18$$

p-ISSN 2085-7829
e-ISSN 2798-3455

$$
\begin{aligned}
\bar{X}_3 &= \sum_{i=1}^{n} \frac{\bar{X}_{3i}}{n} = \frac{X_{3,1} + X_{3,2} + \cdots + X_{3,56}}{56} \\
&= \frac{137.737 + 72.552 + \cdots + 106.963}{56} \\
&= 145.660,29 \\
\bar{X}_4 &= \sum_{i=1}^{n} \frac{\bar{X}_{4i}}{n} = \frac{X_{4,1} + X_{4,2} + \cdots + X_{4,56}}{56} \\
&= \frac{19,00 + 2,00 + \cdots + 17,00}{56} \\
&= 20,79 \\
\bar{X}_5 &= \sum_{i=1}^{n} \frac{\bar{X}_{5i}}{n} = \frac{X_{5,1} + X_{5,2} + \cdots + X_{5,56}}{56} \\
&= \frac{400 + 357 + \cdots + 340}{56} \\
&= 414,14
\end{aligned}
$$

b. Setelah mendapat nilai $\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$, $\bar{X}_4$, $\bar{X}_5$ maka dihitung V
$V = X - \bar{X}$

$$
V = \begin{bmatrix}
8,95 - 6,13 & 1.157,00 - 1.150,18 & \cdots & 400,00 - 414.14 \\
9,09 - 6,13 & 803,00 - 1.150,18 & \cdots & 357,00 - 414,14 \\
7,20 - 6,13 & 2.058,00 - 1.150,18 & \cdots & 989,00 - 414,14 \\
\vdots & \vdots & \ddots & \vdots \\
4,88 - 6,13 & 2.897,00 - 1.150,18 & \cdots & 1.112,00 - 414,14 \\
4,19 - 6,13 & 1.980,00 - 1.150,18 & \cdots & 340,00 - 414,14
\end{bmatrix}
$$

$$
V = \begin{bmatrix}
2,82 & 6,82 & \cdots & -14,14 \\
2,96 & -347,18 & \cdots & -57,14 \\
1,07 & 907,82 & \cdots & 574,86 \\
\vdots & \vdots & \ddots & \vdots \\
1,25 & 1.746,82 & \cdots & 697,86 \\
1,22 & 829,82 & \cdots & -74,14
\end{bmatrix}
$$

c. Menghitung nilai variansi kovariansi menggunakan persamaan (4), dimana $X - \bar{X}$ disederhanakan menjadi matriks V sehingga diperoleh matriks variansi-kovariansi sebagai berikut.

$$
S_{ip} = \frac{1}{55} \begin{bmatrix}
299,11 & -16.040,31 & \cdots & 1.986,07 \\
-16.040,30 & 21.772.898,21 & \cdots & 5.925.767,57 \\
-1.015.912,00 & 2.668.606.871,14 & \cdots & 971.762.858,61 \\
335,13 & 234.885,14 & \cdots & 142.266,71 \\
1.986,07 & 5.925.767,57 & \cdots & 2.783.406,40
\end{bmatrix}
$$

$$
S_{ip} = \begin{bmatrix}
5,44 & -291,64 & -18.471,13 & 6,09 & 36,11 \\
-291,64 & 395.870,88 & 48.520.124,93 & 4.270,64 & 107.741,23 \\
-18.471,13 & 48.520.124,93 & 8.757.408.995,04 & 826.124,43 & 17.668.42 \\
6,09 & 4.270,64 & 826.124,43 & 155,99 & 2.586,67 \\
36,11 & 107.741,23 & 17.668.415,61 & 2.586,67 & 50.607,40
\end{bmatrix}
$$

4.4 Menghitung Matriks Baku

Menghitung matriks baku, dengan asumsi $i \neq k$ dihasilkan $Cov(i,k) = 0$ sehingga didapat matriks sebagai berikut:

$$
\left(V^{\frac{1}{2}}\right) = \begin{bmatrix}
\sqrt{5,44} & 0 & 0 & 0 & 0 \\
0 & \sqrt{395.870,88} & 0 & 0 & 0 \\
0 & 0 & \sqrt{8.757.408.995,04} & 0 & 0 \\
0 & 0 & 0 & \sqrt{155,99} & 0 \\
0 & 0 & 0 & 0 & \sqrt{50.607,40}
\end{bmatrix}
$$

$$
\left(V^{\frac{1}{2}}\right) = \begin{bmatrix}
2,33 & 0 & 0 & 0 & 0 \\
0 & 629,18 & 0 & 0 & 0 \\
0 & 0 & 93.581,03 & 0 & 0 \\
0 & 0 & 0 & 12,49 & 0 \\
0 & 0 & 0 & 0 & 224,96
\end{bmatrix}
$$

$$
\left(V^{\frac{1}{2}}\right)^{-1} = \begin{bmatrix}
0,43 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0,08 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

p-ISSN 2085-7829
e-ISSN 2798-3455

4.5 Menghitung Matriks Korelasi

Matriks korelasi dihitung menggunakan persamaan (5), sehingga diperoleh hasil sebagai berikut:

$$\rho = \begin{bmatrix} 1 & -0{,}20 & -0{,}08 & 0{,}21 & 0{,}07 \\ -0{,}20 & 1 & 0{,}82 & 0{,}54 & 0{,}76 \\ -0{,}08 & 0{,}82 & 1 & 0{,}71 & 0{,}84 \\ 0{,}21 & 0{,}54 & 0{,}71 & 1 & 0{,}92 \\ 0{,}07 & 0{,}76 & 0{,}84 & 0{,}92 & 1 \end{bmatrix}$$

Berdasarkan matriks korelasi terlihat bahwa ada hubungan kolinieritas antar variabel bebas.

4.6 Melakukan Transformasi Variabel

Satuan pengukuran variabel X_1, X_2, X_3 , X_4, dan X_5 berbeda maka satuan itu perlu dibakukan dengan jalan melakukan transformasi X ke dalam variabel baku Z dan diperoleh hasil sebagai berikut:

a. Menghitung Z_1

$$Z_{1,i} = \frac{(X_{1,i} - \bar{X})}{\sqrt{S_{1,1}}}$$

$$Z_{1,1} = \frac{(X_{1,1} - \bar{X})}{\sqrt{S_{1,1}}} = \frac{(8{,}95 - 6{,}13)}{2{,}33} = 1{,}21$$

$$Z_{1,2} = \frac{(X_{1,2} - \bar{X})}{\sqrt{S_{1,1}}} = \frac{(9{,}09 - 6{,}13)}{2{,}33} = 1{,}27$$

$$\vdots$$

$$Z_{1,56} = \frac{(X_{1,56} - \bar{X})}{\sqrt{S_{1,1}}} = \frac{(4{,}91 - 6{,}13)}{2{,}33} = -0{,}52$$

b. Menghitung Z_2

$$Z_{2,i} = \frac{(X_{2,i} - \bar{X})}{\sqrt{S_{2,2}}}$$

$$Z_{2,1} = \frac{(X_{2,1} - \bar{X})}{\sqrt{S_{2,2}}} = \frac{(1.157 - 1.150{,}18)}{629{,}18} = 0{,}01$$

$$Z_{2,2} = \frac{(X_{2,2} - \bar{X})}{\sqrt{S_{2,2}}} = \frac{(803 - 1.150{,}18)}{629{,}18} = -0{,}55$$

$$\vdots$$

$$Z_{2,56} = \frac{(X_{2,56} - \bar{X})}{\sqrt{S_{2,2}}} = \frac{(1.980 - 1.150{,}18)}{629{,}18} = 1{,}32$$

c. Menghitung Z_3

$$Z_{3,i} = \frac{(X_{3,i} - \bar{X})}{\sqrt{S_{3,3}}}$$

$$Z_{3,1} = \frac{(X_{3,1} - \bar{X})}{\sqrt{S_{3,3}}} = \frac{(137.737 - 145.660{,}29)}{93.581{,}03} = -0{,}08$$

$$Z_{3,2} = \frac{(X_{3,2} - \bar{X})}{\sqrt{S_{3,3}}} = \frac{(72.552 - 145.660{,}29)}{93.581{,}03} = -0{,}78$$

$$\vdots$$

$$Z_{3,56} = \frac{(X_{3,56} - \bar{X})}{\sqrt{S_{3,3}}} = \frac{(106.963 - 145.660{,}29)}{93.581{,}03} = -0{,}41$$

d. Menghitung Z_4

$$Z_{4,i} = \frac{(X_{4,i} - \bar{X})}{\sqrt{S_{4,4}}}$$

$$Z_{4,1} = \frac{(X_{4,1} - \bar{X})}{\sqrt{S_{4,4}}} = \frac{(19 - 20{,}79)}{12{,}49} = -0{,}14$$

$$Z_{4,2} = \frac{(X_{4,2} - \bar{X})}{\sqrt{S_{4,4}}} = \frac{(22 - 20{,}79)}{12{,}49} = 0{,}10$$

$$\vdots$$

$$Z_{4,56} = \frac{(X_{4,56} - \bar{X})}{\sqrt{S_{4,4}}} = \frac{(17 - 20{,}79)}{12{,}49} = -0{,}30$$

p-ISSN 2085-7829
e-ISSN 2798-3455

e. Menghitung Z_5

$$Z_{5,i} = \frac{(X_{5,i} - \bar{X})}{\sqrt{S_{5,5}}}$$

$$Z_{5,1} = \frac{(X_{5,1} - \bar{X})}{\sqrt{S_{5,5}}} = \frac{(400 - 414{,}14)}{224{,}96} = -0{,}06$$

$$Z_{5,2} = \frac{(X_{5,2} - \bar{X})}{\sqrt{S_{5,5}}} = \frac{(357 - 414{,}14)}{224{,}96} = -0{,}25$$

$$\vdots$$

$$Z_{5,56} = \frac{(X_{5,56} - \bar{X})}{\sqrt{S_{5,5}}} = \frac{(340 - 414{,}14)}{224{,}96} = -0{,}33$$

4.7 Analisis Komponen Utama

Selanjutnya dicari nilai rata-rata dari $Z_1, Z_2, Z_3, Z_4,$ dan Z_5 sebagai berikut:

$$\bar{Z} = \sum_{i=1}^{n} \frac{Z_i}{n}$$

$$\bar{Z}_1 = \sum_{i=1}^{56} \frac{1{,}21 + 1{,}27 + \cdots + (-0{,}52)}{56} = -2{,}54 \times 10^{-16}$$

$$\bar{Z}_2 = \sum_{i=1}^{56} \frac{0{,}01 + (-0{,}55) + \cdots + 1{,}32}{56} = 1{,}43 \times 10^{-16}$$

$$\bar{Z}_3 = \sum_{i=1}^{56} \frac{(-0{,}08) + (-0{,}78) + \cdots + (-0{,}41)}{56} = 0$$

$$\bar{Z}_4 = \sum_{i=1}^{56} \frac{(-0{,}14) + 0{,}10 + \cdots + (-0{,}30)}{56} = 5{,}35 \times 10^{-17}$$

$$\bar{Z}_5 = \sum_{i=1}^{56} \frac{-0{,}06 + (-0{,}25) + \cdots + (-0{,}33)}{56} = -8{,}82 \times 10^{-17}$$

Skala pengukurannya belum sama, sehingga perlu dilakukan analisis komponen utama dan didapatkan hasil sebagai berikut:

Tabel 3. Hasil Analisis Komponen Utama

Variabel	Komponen Utama				
	I	**II**	**III**	**IV**	**V**
W_1	0,00	-0,90	-0,44	0,00	0,01
W_2	0,47	0,28	-0,59	-0,53	-0,26
W_3	0,51	0,12	-0,25	0,81	-0,04
W_4	0,48	-0,29	0,59	-0,11	-0,57
W_5	0,53	-0,10	0,23	-0,21	0,78
Nilai Eigen	3,31	1,15	0,36	0,15	0,03
Proporsi Total Variansi	0,66	0,23	0,07	0,03	0,01
Proporsi Variansi Kumulatif	0,66	0,89	0,96	0,99	1,00

Berdasarkan Tabel 3, komponen utama yang mewakili yaitu komponen utama W_1 dan W_2 utama sesuai dengan kriteria 80% dari proporsi total variansi yaitu 89%.

4.8 Perhitungan Skor Komponen Utama

Untuk meregresikan komponen W_1 dan W_2 maka perlu dihitung skor komponen utama**.** Berikut hasil perhitungan skor komponen utama W_1 dan W_2:

$$\begin{aligned} W_{1,1} &= 0{,}00Z_{11} + (-0{,}90)Z_{12} + \cdots + 0{,}01Z_{15} \\ &= 0{,}00(1{,}21) + (-0{,}90)(0{,}01) + \cdots + 0{,}01(-0{,}06) \\ &= 0{,}03 \end{aligned}$$

$$\begin{aligned} W_{1,2} &= 0{,}00Z_{21} + (-0{,}90)Z_{22} + \cdots + 0{,}01Z_{25} \\ &= 0{,}00(1{,}27) + (-0{,}90)(-0{,}55) + \cdots + 0{,}01(-0{,}25) \\ &= 0{,}83 \end{aligned}$$

p-ISSN 2085-7829
e-ISSN 2798-3455

$$
\begin{aligned}
&\vdots\\
W_{1,56} &= 0{,}00Z_{56,1} + (-0{,}90)Z_{56,2} + \cdots + 0{,}01Z_{56,5}\\
&= 0{,}00(-0{,}52) + (-0{,}90)(1{,}32) + \cdots + 0{,}01(-0{,}33)\\
&= -1{,}01\\
W_{2,1} &= 0{,}47Z_{11} + 0{,}28Z_{12} + \cdots + (-0{,}26)Z_{15}\\
&= 0{,}47(1{,}21) + 0{,}28(0{,}01) + \cdots + (-0{,}26)(-0{,}06)\\
&= 0{,}72\\
W_{2,2} &= 0{,}47Z_{21} + 0{,}28Z_{12} + \cdots + (-0{,}26)Z_{15}\\
&= 0{,}47(1{,}27) + 0{,}28(-0{,}55) + \cdots + (-0{,}26)(-0{,}25)\\
&= 0{,}92\\
&\vdots\\
W_{2,56} &= 0{,}47Z_{11} + 0{,}28Z_{12} + \cdots + (-0{,}26)Z_{15}\\
&= 0{,}47(-0{,}52) + 0{,}28(1{,}32) + \cdots + (-0{,}26)(-0{,}33)\\
&= 0{,}62
\end{aligned}
$$

Nilai-nilai dari W_1 dan W_2 kemudian disajikan dalam Tabel 4.

Tabel 4 Skor Komponen Utama

$\boldsymbol{i}$	$\boldsymbol{W_1}$	$\boldsymbol{W_2}$
1	0.03	0.72
2	0.83	0.92
3	-2.34	-2.87
4	-0.38	0.39
5	0,18	0,04
6	1,16	1,18
⋮	⋮	⋮
56	-1.01	0.62

4.9 Melakukan Analisis Regresi Komponen Utama

1. Model Awal

Berikut model awal regresi linier berganda pada penelitian ini:

$$Y = \beta_0 + \beta_1 W_1 + \beta_2 W_2 + \varepsilon \tag{10}$$

2. Estimasi Parameter

Estimasi parameter untuk model regresi berganda variabel bebas terhadap Variabel Terikat. Estimasi dilakukan menggunakan bantuan *Software R* dengan hasil berikut:

Tabel 5 Estimasi Parameter

Konstanta	Koefisien β
$\boldsymbol{\beta_0}$	71,53
$\boldsymbol{\beta_1}$	$-1{,}24$
$\boldsymbol{\beta_2}$	$-0{,}05$

Sehingga diperoleh persamaan regresi sebagai berikut.

$$\hat{Y} = 71{,}53 - 1{,}24W_1 - 0{,}05W_2 \tag{11}$$

3. Uji Signifikansi Parameter

a. Uji Simultan

Untuk mengetahui apakan variabel bebas memiliki pengaruh terhadap variabel terikat, maka dilakukan pengujian secara simultan. Pengujian dilakukan dengan menggunakan software R, dengan hasil sebagai berikut:

Tabel 6 Uji F

Uji F	
p-value	0,01

Berdasarkan pengujian signifikansi parameter secara simultan, didapatkan nilai *p-value* sebesar 0,01 lebih kecil dari α sebesar 0,05, sehingga dapat disimpulkan bahwa terdapat minimal satu parameter yang signifikan terhadap Y

b. Uji Parsial

Pengujian parsial dilakukan untuk mengetahui parameter mana saja yang berpengaruh secara signifikan

p-ISSN 2085-7829
e-ISSN 2798-3455

terhadap model regresi. Hipotesis pengujian parsial adalah sebagai berikut.

Tabel 7 Hasil Uji Signifikansi Parameter Parsial

Parameter	*P-value*	Keputusan	Kesimpulan
$\boldsymbol{\beta_1}$	0,05	H_0 ditolak	Signifikan
$\boldsymbol{\beta_2}$	0,95	H_0 diterima	Tidak Signifikan

Berdasarkan Tabel 7, terlihat bahwa dari 2 parameter yang diuji terdapat 1 parameter yang signifikan dan 1 parameter yang tidak signifikan.

4. Subtitusi Nilai *W* terhadap Model Regresi

Setelah diperoleh kemudian nilai $\hat{Y}$ ditransformasikan ke nilai *Z* sebagai berikut:

$$\hat{Y} = 71{,}53 - 1{,}25(0{,}00Z_1 - 0{,}90Z_2 - 0{,}44Z_3 + 0{,}00Z_4 + 0{,}01Z_5) - 0{,}05(0{,}47Z_1 + +0{,}28Z_2 - 0{,}59Z_3 - 0{,}53Z_4 - 0{,}26Z_5) \quad (12)$$

$$\hat{Y} = 71{,}53 + 0Z_1 - 0{,}02Z_1 + 1{,}12Z_2 - 0{,}01Z_2 + 0{,}54Z_3 + 0{,}03Z_3 - 0{,}00Z_4 + 0{,}02Z_4 - 0{,}02Z_5 + 0{,}01Z_5$$

Sehingga diperoleh nilai komponen utama adalah:

$$\hat{Y} = 71{,}53 - 0{,}02Z_1 + 1{,}11Z_2 + 0{,}57Z_3 + 0{,}02Z_4 - 0{,}00Z_5 \quad (13)$$

5. Kesimpulan

1. Pengidentifikasian multikolinieritas pada data IPM Kabupaten/Kota di Pulau Kalimantan adalah dengan menggunakan metode *Variansi Inflation Factor* (*VIF*), didapatkan bahwa terjadi multikoliniearitas pada variabel Jumlah Sekolah Menengah Atas dan Jumlah Guru Sekolah.
2. Multikolinieritas pada data IPM Kabupaten/Kota di Pulau Kalimantan diatasi dengan menggunakan metode PCA dan didapatkan tidak terjadi multikolinieritas terhadap model regresi.
3. Model regresi yang didapatkan setelah dilakukan analisis menggunakan metode PCA adalah sebagai berikut:

$$\hat{Y} = 71{,}53 - 0{,}02Z_1 + 1{,}11Z_2 + 0{,}57Z_3 + 0{,}02Z_4 - 0{,}00Z_5$$

6. Referensi

Alfigari. (2000). *Analisis Regresi Edisi Ke Dua.* Yogyakarta: BPFE.

Azahari, A. (2000). Pembangunan Sumber Daya Manusia dan Indeks Pembangunan Manusia Sektor Pertanian. *Jurnal Ekonomi dan Bisnis Indonesia*, 15(1), 56-69

Draper, Norman dan Smith, Harry. (1992). *Analisis Regresi Terapan Edisi Kedua*. Jakarta: PT Gramedia Pustaka Utama.

Gujarati, Damodar dan Dawn C, Porter. (2015). *Basics Ekonometrics*, Mc Graw Hill, Inc. New York.

Montgomery, D.C. & E.A. Peck. (1991). *Introduction to Linear Regression Analysis*, *Second Edition*. New York: John Wiley and Sons, Inc.

Soemartini. (2008). *Principal Component Analysis (PCA) Sebagai Salah Satu Metode Untuk Mengatasi Masalah Multikolinieritas*. Jurusan Statistika, F.MIPA, Universitas Padjadjaran.

Sriningsih, M., Hatidja, D., & Prang, J. D. (2018). Penanganan Multikolinearitas Dengan Menggunakan Analisis Regresi Komponen Utama Pada Kasus Impor Beras Di Provinsi Sulut. *Jurnal Ilmiah Sains*, 18(1), 18–24.

Sungkono, J. & Nugrahaningsih, T.K. (2017). Simulasi Dampak Multikolinearitas Pada Kondisi Penyimpangan Asumsi Normalitas. *Jurnal Magistra*. 1 (101), 45-50.