

## METODE PRINCIPAL COMPONENT ANALYSIS (PCA) SEBAGAI PENANGANAN ASUMSI MULTIKOLINEARITAS (STUDI KASUS: DATA PRODUKSI TAPIOKA)

### *Principal Component Analysis (PCA) Method to Handle Multicollinearity Assumptions (Case Study: Tapioca Production Data)*

**Dwi Retno Puspita Sari**

Institut Agama Islam Negeri Metro

Jl. K.H Dewantara 15A Iringmulyo, Kota Metro, 34111, Lampung, Indonesia

E-mail Correspondence Author: [dwiretno.ps@gmail.com](mailto:dwiretno.ps@gmail.com)

#### **Abstrak**

Salah satu metode analisis yang banyak digunakan adalah analisis regresi linier. Sedangkan pada kasus regresi berganda, pola hubungan linier diterapkan untuk satu variabel bebas dengan beberapa variabel terikat. Pada tahapan analisis regresi, beberapa asumsi yakni asumsi normalitas, linearitas, heterokedastisitas, autokorelasi, dan multikolinearitas. Metode principal component analysis atau PCA merupakan suatu teknik multivariat yang bertujuan untuk mereduksi faktor atau variabel dalam jumlah besar menjadi beberapa faktor yang lebih sedikit. Selain digunakan untuk mereduksi jumlah variabel, metode PCA juga dapat digunakan untuk menangani masalah multikolinearitas dengan mereduksi jumlah variabelnya. Tujuan penulisan jurnal ini yakni untuk melakukan penanganan pada pelanggaran asumsi multikolinearitas tanpa melakukan reduksi jumlah variabel. Hasil dari penelitian ini adalah bahwa metode PCA layak untuk digunakan dalam menangani masalah pelanggaran asumsi multikolinearitas tanpa melakukan reduksi terhadap jumlah variabel awal. Sehingga keseluruhan informasi yang terkandung pada masing-masing variabel dapat tetap dipertahankan.

**Kata Kunci:** Metode PCA, Multikolinearitas, Regresi Berganda, Reduksi Data.

#### **Abstract**

One of the most analytical methods used is linear regression analysis. Where as in the case of multiple regression, the linear relationship pattern is applied to one independent variable with several dependent variables. At the regression analysis stage, there are several assumptions are normality, linearity, heteroscedasticity, autocorrelation, and multicollinearity. Principal component analysis or PCA method is a multivariate technique that aims to reduce a large number of factors or variables into a number of fewer factors. Besides being used to reduce the number of variables, the PCA method used to be solve of multicollinearity problems by reducing the number of variables. The purpose of writing this journal is to be solve the violations of multicollinearity assumption without reducing the number of variables. The result of this research is that the PCA method is feasible to be used in solving with the problem of violating the multicollinearity assumption without reducing the number of initial variables. So that the overall information contained in each variable can be maintained.

**Keywords:** PCA Method, Multicollinearity, Multiple Regression, Data Reduction.



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

## 1. PENDAHULUAN

Dalam ilmu statistika, salah satu metode analisis yang banyak digunakan dalam penelitian adalah analisis regresi linier, baik sederhana maupun berganda. Metode analisis regresi bertujuan untuk menentukan pola hubungan linier sebab akibat antara variabel bebas dengan variabel terikat. Pola hubungan linier ini nantinya akan dibentuk dalam persamaan matematik, sehingga dapat digunakan untuk memprediksi nilai variabel terikat jika diketahui nilai variabel bebasnya [1].

Penentuan variabel bebas dan terikat membutuhkan pengkajian terlebih dahulu berdasarkan teori-teori pendukungnya. Tujuannya untuk meminimalisir kesalahan dalam penentuan variabel terikat dan bebas yang kadang sedikit sulit jika penelitian dilakukan oleh orang awam. Selain untuk memperoleh persamaan matematis antara variabel bebas dan terikat, tujuan lain ketika metode analisis regresi digunakan adalah untuk meminimumkan kesalahan antara data asli dengan data prediksi dari variabel terikat, serta untuk memaksimalkan korelasi antara data asli dengan data prediksi variabel terikat berdasarkan data yang ada [2].

Pada kasus regresi sederhana, pola hubungan linier diterapkan untuk satu variabel bebas dan satu variabel terikat. Sedangkan pada kasus regresi berganda, pola hubungan linier diterapkan untuk satu variabel bebas dengan beberapa variabel terikat. Pada tahapan analisis regresi, terdapat beberapa asumsi yang wajib untuk dipenuhi. Hal ini dikarenakan analisis regresi merupakan salah satu ilmu statistika parametrik. Selain itu, pengujian asumsi diperlukan untuk mengetahui apakah data yang digunakan betul-betul memiliki sebaran data tertentu atau tidak. Beberapa asumsi tersebut yakni asumsi normalitas, linearitas, heterokedastisitas, autokorelasi, dan multikolinearitas.

Metode principal component analysis atau PCA merupakan suatu teknik multivariat yang bertujuan untuk mereduksi faktor atau variabel dalam jumlah besar menjadi beberapa faktor yang lebih sedikit. Permasalahan yang sering muncul dalam proses reduksi faktor atau variabel ini yakni bagaimana memperkecil jumlah variabel namun tetap mempertahankan informasi atau karakter penting yang terkandung pada data [1]. Beberapa penelitian terdahulu terkait metode principal component analysis atau PCA pernah ditulis oleh Wangge [3] yakni dengan melakukan reduksi faktor-faktor yang mempengaruhi lamanya penyelesaian skripsi mahasiswa Prodi Matematika di FKIP UNDANA. Data yang digunakan dalam penelitian tersebut adalah data primer hasil wawancara terhadap 50 mahasiswa dengan 13 variabel dan diperoleh hasil 10 variabel terpilih. Khikmah [4] juga pernah melakukan penelitian dengan menggunakan metode PCA dalam menentukan faktor dominan cuaca terhadap penyebaran wabah Covid-19 di Surabaya. Hasil penelitian tersebut yakni terdapat 3 komponen yang mempengaruhi penyebaran virus, yakni suhu, kelembaban, dan lama sinar matahari. Penelitian berikutnya terkait metode PCA juga pernah dilakukan oleh Fitriainingsih dan Sugiyarto [5] dalam mereduksi variabel yang mempengaruhi perbaikan pada fungsi ginjal tikus dari 8 variabel menjadi 3 variabel.

Selain pada data dengan jumlah variabel banyak, metode PCA ini dapat digunakan juga pada data yang memiliki korelasi besar antar variabelnya. Sehingga diharapkan persamaan regresi yang diperoleh bisa terbebas dari masalah adanya korelasi antar variabel tanpa menghilangkan informasi penting pada data yang digunakan. Berdasarkan hal tersebut, tujuan dari penulisan jurnal ini adalah untuk mengetahui tahapan dalam penerapan metode PCA dalam menangani pelanggaran

asumsi multikolinearitas pada analisis regresi linier berganda tanpa harus mereduksi jumlah variabel yang digunakan dalam penelitian.

## 2. METODE PENELITIAN

### 2.1. Skala Penelitian

Hal utama yang paling dibutuhkan pada analisis statistik adalah data. Data merupakan kumpulan angka yang mewakili karakteristik dari objek yang diamati. Berdasarkan skala pengukurannya, skala data dibedakan dalam empat kelompok mulai dari yang terendah hingga tertinggi berdasarkan tingkatan pembandingnya, yakni skala nominal, ordinal, interval, dan rasio.

#### a. Skala Nominal

Skala nominal merupakan skala pengukuran data yang paling sederhana karena hanya dapat digunakan pada data berupa kategori atau kualitatif saja, seperti nama dan jenis kelamin. Dalam tahapan analisis pun, data kategori selanjutnya akan diberi label berupa angka. Penggunaan angka pada skala nominal tidak menjadikan data dapat diolah secara matematis seperti penambahan, pengurangan, perkalian dan pembagian. Oleh karena itu, data pada skala nominal hanya dapat diterapkan pada prosedur statistik deskriptif [6]. Ciri-ciri dari data berskala nominal antara lain [7]:

1. Hanya dapat digunakan pada data yang bersifat membedakan. Sehingga data tidak dapat diurutkan atau dibandingkan.
2. Kategori data bersifat homogen. Artinya bahwa setiap datanya hanya dapat dikategorikan pada satu kategori saja dan setiap kategori harus mencakup seluruh data.

#### b. Skala Ordinal

Skala ordinal merupakan skala pengukuran yang dapat digunakan jika data penelitian dapat dibandingkan antara satu dengan yang lainnya, misalnya lebih banyak, sama atau lebih sedikit. Selain pada data perbandingan, skala ordinal juga dapat diterapkan pada data berperingkat seperti pada tingkat kepuasan pelanggan dari sangat tidak suka sampai sangat suka, atau pada peringkat kelas pertama, kedua, dan seterusnya. Pemberian urutan menjadi hak sepenuhnya dari peneliti, apakah akan dimulai dari peringkat terendah sampai tertinggi atau dari peringkat tertinggi sampai terendah dengan jarak antar peringkat tidak harus sama. Namun begitu, angka yang digunakan hanya akan menentukan posisi pada urutan, sehingga tidak dapat dioperasikan matematis seperti ditambah, kurang, kali maupun bagi [7].

#### c. Skala Interval

Tingkatan skala pengukuran berikutnya diantara skala nominal dan ordinal adalah skala interval. Pada skala interval, terdapat satuan pengukuran standar dan jarak antar kategori. Sehingga dapat dikatakan pada skala interval ini terdapat batas bawah dan batas atasnya. Karakteristik lainnya yakni pada skala interval tidak terdapat nilai 0 mutlak, sehingga nilai-nilai pada skala interval tidak dapat dibandingkan, akan tetapi dapat dijumlah atau dikurangkan [7].

#### d. Skala Rasio

Skala pengukuran yang terakhir yakni skala interval. Pada skala interval secara

garis besar memiliki karakteristik yang sama dengan skala interval, akan tetapi pada skala rasio memiliki nilai 0 mutlak. Hal tersebut menjadikan data pada skala rasio dapat dibandingkan serta dapat dilakukan operasi matematis. Sebagai contoh tinggi pohon A adalah 2 meter sedangkan tinggi pohon B adalah 4 meter, sehingga dapat dikatakan bahwa tinggi pohon B adalah dua kali tinggi pohon A [7].

## 2.2. Analisis Regresi Berganda

Analisis regresi merupakan salah satu metode statistik yang banyak digunakan dalam penelitian. Keunggulan metode regresi adalah relatif mudah dalam penerapan pada kasus data aktual. Metode regresi bertujuan untuk mengetahui pengaruh dari satu atau beberapa variabel bebas terhadap variabel terikat. Penentuan variabel bebas dan terikat membutuhkan pemahaman dari peneliti. Variabel yang bersifat mempengaruhi disebut variabel bebas, sedangkan variabel yang sifatnya dipengaruhi disebut variabel terikat. Hubungan sebab akibat tersebut nantinya akan dimodelkan dalam persamaan matematis yang disebut persamaan regresi.

Pada kasus hubungan sebab akibat antara satu variabel terikat dengan satu variabel bebas disebut regresi sederhana. Sebagai contoh regresi sederhana yakni pengaruh banyaknya intensitas air yang digunakan terhadap tinggi tumbuhan. Contoh lain yakni intensitas belajar mandiri terhadap nilai hasil ujian mahasiswa. Sedangkan regresi berganda merupakan metode analisis untuk mengetahui sebab akibat antara beberapa variabel bebas terhadap satu variabel terikat. Salah satu contoh regresi berganda adalah pengaruh kecepatan pelayanan dan kualitas produk terhadap keputusan konsumen dalam memilih produk [2]. Bentuk persamaan regresi berganda dapat dituliskan sebagai berikut:

$$Y_i = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (1)$$

Dimana:

$Y_i$  = variabel terikat yang dipengaruhi

$b_1, b_2, \dots, b_n$  = parameter regresi

$X_1, X_2, \dots, X_n$  = variabel bebas yang mempengaruhi

Dalam analisis regresi, penduga parameter regresi diperoleh dengan menggunakan metode kuadrat terkecil (MKT). Hal tersebut bertujuan untuk memperoleh penduga parameter regresi terbaik dengan tingkat kesalahan terkecil.

## 2.3. Asumsi Regresi

Dalam penerapan metode analisis regresi, terdapat beberapa asumsi yang wajib dipenuhi terlebih dahulu sebelum terbentuk persamaan regresi matematis. Hal tersebut dikarenakan analisis regresi merupakan salah satu metode statistika parametrik, sehingga memerlukan asumsi-asumsi tersendiri yang harus dipenuhi. Beberapa asumsi yang wajib dipenuhi antara lain asumsi normalitas, heteroskedastisitas, autokorelasi dan multikolinearitas.

Asumsi normalitas merupakan asumsi utama yang wajib untuk dipenuhi sebelum melangkah ke asumsi yang lain. Tujuan pengujian asumsi normalitas ini untuk mengetahui apakah sisaan sebaran data yang digunakan berdistribusi normal atau

tidak. Dalam analisis regresi, pengujian yang digunakan adalah uji F dan atau uji t yang keduanya mengasumsikan sisaan berdistribusi normal. Sehingga, adanya pelanggaran asumsi normalitas akan mengakibatkan persamaan yang diperoleh memiliki parameter yang tidak valid. Terdapat beberapa cara yang dapat digunakan untuk mendeteksi asumsi normalitas, yakni menggunakan grafik sebaran sisaan data, analisis statistik melalui nilai kurtosis dan skewness, serta menggunakan uji *Kolmogorov-smirnov*.

Asumsi heteroskedastisitas adalah asumsi yang bertujuan untuk menguji apakah sisaan data yang digunakan memiliki perbedaan ragam antara satu dengan yang lainnya. Model regresi yang baik jika persamaan yg dihasilkan memiliki ragam yang tetap atau homoskedastisitas. Suatu data bisa saja memiliki kemungkinan mengalami masalah heteroskedastisitas jika data pada variabel yang digunakan cenderung beragam, sehingga sisaan yang dihasilkan juga tidak konstan. Hal yang dapat dilakukan untuk menguji ada atau tidaknya kondisi heteroskedastisitas yakni dapat dilihat dari sebaran sisaan apakah membentuk pola tertentu atau tidak. Apabila sisaan memiliki sebaran dengan pola tertentu, maka data dapat dikatakan cenderung memiliki kondisi heteroskedastisitas.

Asumsi autokorelasi bertujuan untuk mengetahui apakah pada sisaan amatan ke  $i$  memiliki pengaruh dari amatan ke  $i-1$ . Penyebab adanya autokorelasi biasanya karena data yang digunakan berupa data runtun waktu, sehingga ada hubungan antara data yang satu dengan yang lainnya. Beberapa cara yang dapat digunakan untuk menguji ada tidaknya autokorelasi pada data dengan menggunakan uji *durbin-watson*. Pengambilan keputusan terkait uji *durbin-watson* dapat menggunakan pedoman pada [Tabel 1 \[8\]](#).

**Tabel 1. Kriteria Keputusan Uji Durbin-Watson**

Kriteria	Keputusan
$0 < dw < dL$	Ada masalah autokorelasi positif
$dL < dw < dU$ atau $4 - dU < dw < 4 - dL$	Tidak ada kesimpulan, wilayah ragu-ragu
$dU < dw < 4 - dU$	Tidak ada masalah autokorelasi
$4 - dL < dw < 4$	Ada masalah autokorelasi negatif
$2 < dw < 4 - dU$ atau $dU < dw < 2$	Tidak ada masalah autokorelasi positif atau negatif

Sumber : Veni Sujarweni, buku "Kupas Tuntas Penelitian Akuntansi dengan SPSS", 2016.

Asumsi multikolinearitas merupakan uji asumsi yang dilakukan dengan tujuan untuk mengetahui apakah antar variabel bebas yang digunakan dalam persamaan saling ada keterkaitan atau kemiripan. Untuk mengetahui ada tidaknya gejala multikolinearitas yakni dengan membandingkan nilai *variance inflation factor* (VIF) jika lebih dari 10, maka terdapat masalah multikolinearitas antar variabel bebas.

## 2.4. Principal Component Analysis (PCA)

Metode *principal component analysis* (PCA) atau analisis komponen utama merupakan salah satu metode analisis data multivariat. Secara umum, metode PCA digunakan untuk mereduksi data atau variabel dalam jumlah besar menjadi lebih sedikit tanpa perlu menghilangkan informasi penting yang terkandung dalam data [9]. Seringkali informasi tersebut dikelompokkan menjadi variabel baru yang lebih sederhana berdasarkan kemiripan informasi yang diperoleh dari data awal.

Pada Tahun 1901 seorang peneliti bernama Karl Pearson pertama kali

memperkenalkan metode PCA. Secara bersamaan, Harold Hotelling juga melakukan analisa untuk variabel stokastik. Dengan menggunakan pendekatan PCA, Harold memperkenalkan metode component sebagai variabel baru yang dihasilkan dari metode PCA. Hingga kini, metode PCA masih terus mengalami perkembangan. Beberapa fungsi dari penerapan metode PCA antara lain [10]:

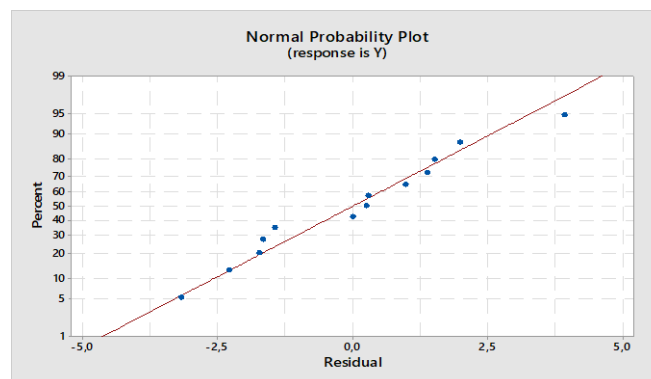
1. Mengidentifikasi variabel baru dari data multivariat,
2. Mereduksi jumlah variabel bebas menjadi variabel bebas dengan jumlah yang lebih sedikit tanpa menghilangkan informasi dari data asli,
3. Mereduksi variabel asli yang memiliki sedikit informasi.

### 3. HASIL DAN PEMBAHASAN

Diberikan contoh studi kasus sebagai berikut, yakni terdapat data produksi tapioka yang dipengaruhi oleh data jumlah modal, jumlah debit air, jumlah tenaga kerja, dan jumlah bahan baku produksi. Banyak data yang digunakan sebanyak 13 data dengan data sebagai berikut:

**Tabel 2. Data Contoh Studi Kasus**

Y	X1	X2	X3	X4
78.5	7	26	6	60
74.3	1	29	15	52
104.3	11	56	8	20
87.6	11	31	8	47
.....	.....	.....	.....	.....
83.8	1	40	23	34
113.3	11	66	9	12
109.4	10	68	8	12



**Gambar 1. Uji Normalitas**

**Tabel 3. Uji Analisis Varian**

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	2667,9	666,975	111,48	0
X1	1	25,95	25,951	4,34	0,071
X2	1	2,97	2,972	0,5	0,501
X3	1	0,11	0,109	0,02	0,896
X4	1	0,25	0,247	0,04	0,844
Error	8	47,86	5,983		
Total	12	2715,76			



Berdasarkan plot sebaran data pada **Gambar 1**, terlihat bahwa residual data menyebar mengikuti garis diagonal. Hal tersebut menunjukkan bahwa data memiliki sebaran data distribusi normal, sehingga data telah memenuhi asumsi data menyebar normal. Pada tabel 2 hasil pengujian analisis varian menunjukkan bahwa nilai signifikan variabel bebas secara berurutan adalah 0.071, 0.501, 0.896, dan 0.844. Secara keseluruhan nilai signifikan telah melebihi 0.05, sehingga asumsi heteroskedastisitas telah terpenuhi.

**Tabel 4. Hasil Uji Durbin-Watson**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.991 <sup>a</sup>	.982	.974	2.44601	2.053

Berdasarkan **Tabel 4**, hasil uji durbin-watson sebesar 2,053. Jika mengacu pada jumlah data sebanyak 13 dan jumlah variabel bebas sebanyak 4, maka pada tabel durbin watson akan diperoleh nilai *durbin lower* sebesar 0,5745 dan *durbin upper* sebesar 2,0943. Jika mengacu pada kriteria keputusan pada **Tabel 1**, maka tidak ada kesimpulan yang dapat diambil.

**Tabel 5. Korelasi Antar Variabel Bebas dan Variabel Terikat**

	X1	X2	X3	X4
X2	0,229 0,453			
X3	-0,824 0,001	-0,139 0,65		
X4	-0,245 0,419	-0,973 0	0,03 0,924	
Y	0,731 0,005	0,816 0,001	-0,535 0,06	-0,821 0,001

Hasil di atas merupakan nilai korelasi antara variabel terikat dengan variabel bebas. Jika menggunakan tingkat signifikansi 5%, maka variabel X3 memberikan hasil pengujian korelasi yang tidak signifikan dengan variabel terikat, sedangkan variabel lain memiliki korelasi yang signifikan dengan variabel terikat. Hubungan korelasi yang signifikan antar variabel bebas terjadi pada variabel X1 dengan X3, kemudian antara X2 dengan X4.

**Tabel 6. Hasil Analisis Varian**

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	62,4	70,1	0,89	0,399	
X1	1,551	0,745	2,08	0,071	38,5
X2	0,51	0,724	0,7	0,501	254,42
X3	0,102	0,755	0,14	0,896	46,87
X4	-0,144	0,709	-0,2	0,844	282,51

Berdasarkan pengujian serentak untuk seluruh parameter dari di atas dapat disimpulkan bahwa paling tidak terdapat satu variabel yang berpengaruh terhadap variabel terikat. Selanjutnya, berdasarkan nilai korelasi sebesar 98,24% artinya keberagaman variabel terikat dapat dijelaskan oleh variabel yang ada dalam model sebesar 98,24% tentunya sudah tergolong besar. Akan tetapi, pada nilai p-value masing-masing peubah bebas tidak menunjukkan hasil yang signifikan. Hal tersebut terlihat

dari nilai p-value yang lebih besar dari 0,05.

Berdasarkan hal di atas, terdapat hasil yang bertolak belakang antara hasil uji serentak, nilai korelasi yang besar dengan nilai uji parsial. Sehingga terdapat indikasi adanya multikolinearitas yang perlu ditangani terlebih dahulu dengan metode principal component analysis. Hal tersebut diperkuat dengan adanya nilai VIF keseluruhan variabel yang melebihi batas angka 10.

**Tabel 7. Nilai Eigenvalue Data**

Eigenvalue	2,2357	1,5761	0,1866	0,0016
Proportion	0,559	0,394	0,047	0
Cumulative	0,559	0,953	1	1

Variable	PC1	PC2	PC3	PC4
z1	-0,476	-0,509	-0,676	-0,241
z2	-0,564	0,414	0,314	-0,642
z3	0,394	0,605	-0,638	-0,268
z4	0,548	-0,451	0,195	-0,677

Berdasarkan output di atas, karena eigenvalue cumulative yang proporsinya mencapai 75% diperoleh sampai eigenvalue kedua, maka akan diambil komponen sebanyak 2 komponen yakni PC1 dan PC2 dengan variabel baru Z dan nilai komponen utama W.

**Tabel 8. Nilai Komponen Utama**

	W1	W2
	1,46724	-1,903
	2,13583	-0,2384
	-1,1299	-0,1839
	0,6599	-1,5768
	.....	.....
	1,64018	1,29511
	-1,6926	0,39225
	-1,7457	0,43753

Berdasarkan tahapan analisis komponen utama di atas, diperoleh persamaan regresi baru sebagai berikut :

$$W_1 = 0,476 Z_1 + 0,564 Z_2 - 0,394 Z_3 - 0,548 Z_4$$

$$W_1 = 0,509 Z_1 - 0,414 Z_2 - 0,605 Z_3 + 0,451 Z_4$$

Pada tahapan analisis ini juga diperoleh nilai-nilai variabel baru yang akan disebut variabel W1 dan W2. Kedua variabel W tersebut selanjutnya dilakukan tahapan analisis regresi terhadap variabel Y. Tahapan regresi dilakukan sampai diperoleh persamaan regresi terbaik antara variabel Y dengan variabel W yang berpengaruh.

**Tabel 9. Uji Varian Variabel W Baru**

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	1	2620,48	2620,48	302,53	0
w1	1	2620,48	2620,48	302,53	0



Error	11	95,28	8,66
Total	12	2715,76	

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	95,423	0,816	116,9	0	
w1	-9,883	0,568	-17,39	0	1

Pada hasil ANOVA di atas, karena nilai p-value pada W2 tidak signifikan, maka dilakukan regresi ulang hanya dengan W1 dan sudah memberikan hasil yang signifikan, baik pada uji serentak maupun parsial. Selain itu, nilai VIF juga menunjukkan hasil bahwa tidak terdapat lagi multikolinearitas pada model. Sehingga unsur W1 yang digunakan dalam model sudah baik. Maka model yang diperoleh adalah:

$$Y = 95,423 - 9,88 W_1$$

$$= 95,423 - 9,88 (0,476 Z_1 + 0,564 Z_2 - 0,394 Z_3 - 0,548 Z_4)$$

Tahapan berikutnya dengan mentransformasi model dengan variabel Z kembali dalam bentuk variabel X:

$$Y = 95,423 - 9,88 (0,476 Z_1 + 0,564 Z_2 - 0,394 Z_3 - 0,548 Z_4)$$

$$= 95,423 - 4,704 Z_1 - 5,57 Z_2 + 3,893 Z_3 + 5,415 Z_4$$

**Tabel 10. Perhitungan Deskriptif Data**

Variable	N	Mean	SE Mean	StDev	Min	Q1	Median	Q3	Max
X1	13	7,46	1,63	5,88	1	1,5	7	11	21
X2	13	48,15	4,32	15,56	26	31	52	61	71
X3	13	11,77	1,78	6,41	4	7	9	17,5	23
X4	13	30	4,64	16,74	6	16	26	45,5	60

$$Y = 95,423 - 4,704 \left( X_1 - \frac{7,46}{5,88} \right) - 5,57 \left( X_2 - \frac{48,15}{15,56} \right) + 3,893 \left( X_3 - \frac{11,77}{6,41} \right) + 5,415 \left( X_4 - \frac{30}{16,74} \right)$$

$$Y = 101,781 - 0,8X_1 - 0,35X_2 + 0,606X_3 + 0,32X_4$$

#### 4. KESIMPULAN

Berdasarkan rangkaian tahapan yang telah dilakukan, kesimpulan yang dapat diperoleh adalah bahwa metode *principal component analysis* (PCA) layak untuk digunakan sebagai metode dalam menangani pelanggaran asumsi multikolinearitas. Hal tersebut terlihat pada akhir tahapan, persamaan regresi yang di peroleh sudah tidak mengalami masalah asumsi multikolinearitas lagi. Penerapan metode PCA ini juga menjadikan keseluruhan variabel X yang digunakan dalam analisis regresi memberikan hasil yang signifikan terhadap variabel Y, sehingga seluruh variabel dapat digunakan dalam persamaan regresi.

## DAFTAR PUSTAKA

- [1] A. Mattjik and I. M. Sumertajaya, *Perancangan Percobaan Dengan Aplikasi SAS Dan Minitab*. Bogor: IPB Press, 2000.
- [2] R. Zelvia, "Penerapan Analisis Regresi Dummy Pada Data Kualitatif Kasus Ekonomi," *Adzkiya J. Huk. dan Ekon. Syariah*, vol. 5, no. 1, p. 45, 2017, doi: 10.32332/adzkiya.v5i1.804, 2017.
- [3] M. Wangge, "Penerapan Metode Principal Component Analysis (PCA) Terhadap Faktor-faktor yang Mempengaruhi Lamanya Penyelesaian Skripsi Mahasiswa Program Studi Pendidikan Matematika FKIP UNDANA," *J. Cendekia J. Pendidik. Mat.*, vol. 5, no. 2, pp. 974–988, 2021, doi: 10.31004/cendekia.v5i2.465, 2021.
- [4] K. N. Khikmah, "Penerapan Principal Component Analysis dalam Penentuan Faktor Dominan Cuaca Terhadap Penyebaran Covid-19 di Surabaya," *ESTIMASI J. Stat. Its Appl.*, vol. 2, no. 1, pp. 11–18, 2021, doi: 10.20956/ejsa.v2i1.11943, 2021.
- [5] F. Fitriyaningsih and S. Sugiyarto, "Implementasi Analisa Komponen Utama untuk Mereduksi Variabel yang Mempengaruhi Perbaikan pada Fungsi Ginjal Tikus," *J. Ilm. Mat.*, vol. 6, no. 2, p. 62, 2019, doi: 10.26555/konvergensi.v6i2.19549, 2019.
- [6] Syafril, *Statistika*, 1st ed. Padang: Sukabina Press, 2010.
- [7] I. N. Budiantara and Zulfikar, *Manajemen Riset Dengan Pendekatan Komputasi Statistika*, 1st ed. Yogyakarta: Deepublish, 2014. [Online]. Available: <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1005862#>, 2014.
- [8] V. Sujarweni, *Kupas Tuntas Penelitian Akuntansi Dengan SPSS*. Yogyakarta: Pustaka Baru Press, 2016.
- [9] I. Nyoman Radiarta, dan Akhmad Mustafa, P. Penelitian dan Pengembangan Perikanan Budidaya Jl Ragunan, P. Minggu, J. Selatan, and B. Penelitian dan Pengembangan Budidaya Air Payau, "Sitakka No. 129, Maros 90512," *Februari*, no. June 2011, p. 22, 2012.
- [10] Solimun;, A. Fernandes, and Nurjannah, *Metode statistika multivariat : pemodelan persamaan struktural (SEM) pendekatan WarpPLS*, 2nd ed. Malang: UB Press, 2017. [Online]. Available: <https://opac.perpusnas.go.id/DetailOpac.aspx?id=1141341>, 2017.