Biplot *Multivariate Regression* dan Penerapannya dalam Menganalisis Faktor-Faktor yang Mempengaruhi Indikator Derajat Kesehatan di Jawa Barat

Thania Indra Mutiara*, Suliadi Suliadi

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*thania.indr@gmail.com, suliadi@gmail.com

Abstract. Multivariate regression analysis is a statistical tool that is concerned with describing and evaluating the relationship between a given set of responses and a set of predictors. This analysis involves many variables, making difficulty to analyze and interpret. Therefore, we need a method that can visualize the regression model in a comprehensive manner to facilitate interpretation. The multivariate regression biplot aims to graphically visualize the effect of the predictor variable on the response variable and to display the regression result. This multivariate regression biplot is well suited for solving problems in multivariate regression analysis. This research uses biplot multivariate regression to analyze data on factors that influence health status indicators in West Java. The results of applying multivariate regression biplot show that the Y_1 variable has a positive correlation with the variables X_2, X_3, and X_4 and has a negative correlation with the variables X_1, and X_5. The variable Y_2 has a positive correlation with the variables X_1, X_2, X_3, and X_5 and has a negative correlation with the variable X_4. The variable Y_3 has a positive correlation with the variable X_4 and has a negative correlation with the variable X_1, X_2, X_3, and X_5. The overall quality of the multivariate regression biplot visualization result is 98.7%. This means that the information provided by the multivariate regression biplot in the form of visualization can explain 98.7% of all the information contained in the data.

Keywords: *Multivariate Regression, Health Degree, Health Degree Indicator*.

Abstrak. Analisis regresi multivariat adalah alat statistik yang berkaitan dengan menggambarkan dan mengevaluasi hubungan antara satu set respon yang diberikan dan satu set prediktor. Analisis tersebut melibatkan banyak variabel, sehingga dalam menganalisis dan menginterpretasikannya menjadi sulit. Oleh karena itu diperlukan metode yang dapat memvisualisasikan model regresi secara komprehensif, sehingga memudahkan interpretasi. Biplot regresi multivariat bertujuan untuk memvisualisasikan pengaruh variabel prediktor terhadap variabel respon secara grafis dan untuk menampilkan hasil regresi. Biplot regresi multivariat ini sangat cocok digunakan untuk mengatasi permasalahan dalam analisis regresi multivariat. Penelitian ini akan membahas biplot regresi multivariat dan penerapannya pada data faktor-faktor yang mempengaruhi indikator derajat kesehatan di Jawa Barat. Hasil penerapan biplot regresi multivariat menunjukkan bahwa variabel Y_1 berkorelasi positif dengan variabel X_2, X_3, dan X_4 serta berkorelasi negatif dengan variabel X_1, dan X_5. Variabel Y_2 berkorelasi positif dengan variabel X_1, X_2, X_3, dan X_5 serta berkorelasi negatif dengan variabel X_4. Variabel Y_3 berkorelasi positif dengan variabel X_4 dan berkorelasi negatif dengan variabel X_1, X_2, X_3, dan X_5. Kualitas keseluruhan dari hasil visualisasi biplot regresi multivariat sebesar 98,7%. Artinya informasi yang diberikan oleh biplot regresi multivariat dalam bentuk visualisasi mampu menerangkan 98,7% dari seluruh informasi yang terkandung dalam data.

Kata Kunci: *Regresi Multivariat, Derajat Kesehatan, Indikator Derajat Kesehatan.*

Corresponding Author
Email: suliadi@gmail.com

A. Pendahuluan

Analisis regresi multivariat adalah alat statistik yang berkaitan dengan menggambarkan dan mengevaluasi hubungan antara satu set respon yang diberikan dan satu set prediktor [3]. Regresi multivariat mencoba menemukan formula yang dapat menjelaskan bagaimana faktor-faktor dalam variabel merespon secara bersamaan terhadap perubahan pada variabel lain. Pada analisis regresi multivariat, karena melibatkan banyak variabel, sehingga dalam menganalisis dan menginterpretasikannya menjadi sulit. Oleh karena itu, diperlukan metode yang dapat memvisualisasikan model regresi secara komprehensif, sehingga memudahkan interpretasi.

Oyedele (2021) memperkenalkan suatu metode analisis biplot yang diperuntukkan untuk analisis regresi multivariat yang disebut dengan "Biplot Regresi Multivariat". Tujuan dari biplot regresi multivariat ini adalah untuk memvisualisasikan pengaruh variabel prediktor terhadap variabel respon secara grafis dan untuk menampilkan hasil regresi [6]. Biplot regresi multivariat ini sangat cocok digunakan untuk mengatasi permasalahan analisis regresi multivariat, yaitu cukup sulit dalam menganalisis dan menginterpretasikan data yang digunakan karena memerlukan proses yang cukup panjang dalam analisisnya.

Di era globalisasi saat ini, kesejahteraan penduduk menjadi aspek yang sangat penting dalam kehidupan masyarakat. Selain itu, kualitas sebuah negara juga diukur dari tingkat kesejahteraan penduduk. Kesejahteraan penduduk sangat bergantung pada derajat kesehatan. Derajat kesehatan terdiri dari dua hal yaitu indikator derajat kesehatan dan faktor-faktor yang mempengaruhi indikator derajat kesehatan [1]. Berbagai upaya dalam meningkatkan derajat kesehatan masyarakat juga berarti berupaya untuk pembangunan negara. Maka, untuk mewujudkan peningkatan derajat kesehatan masyarakat, pemerintah dan masyarakat harus mengetahui indikator dan faktor yang mempengaruhi derajat kesehatan, agar hak kesehatan seluruh masyarakat dapat terjamin [8].

Berdasarkan latar belakang yang telah diuraikan, maka tujuan yang ingin dicapai dalam penelitian ini adalah untuk mengetahui hasil analisa penerapan metode biplot regresi multivariat pada data faktor-faktor yang mempengaruhi indikator derajat kesehatan di Jawa Barat.

B. Metodologi Penelitian

Analisis Regresi Multivariat

Analisis regresi multivariat merupakan model linear dengan q buah variabel respon $Y_1, Y_2, \ldots, Y_q$ yang saling berkorelasi dengan p buah variabel prediktor $X_1, X_2, \ldots, X_p$ dan diasumsikan adanya hubungan ketergantungan di antara variabel respon [2]. Model regresi multivariat untuk pengamatan ke-i, variabel respon ke-j adalah sebagai berikut:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{i1} + \beta_{2j}X_{i2} + \cdots + \beta_{kj}X_{ik} + \varepsilon_{ij} \quad (1)$$

dimana $i = 1, 2, \ldots, n, j = 1, 2, \ldots, q, k = 1, 2, \ldots, p$, Y_{ij} merupakan nilai pengamatan ke-i untuk variabel respon ke-j, X_{ik} merupakan nilai pengamatan ke-i untuk variabel prediktor ke-k, β_{kj} merupakan parameter-parameter regresi variabel prediktor ke-k dan respon ke-j, dan ε_{ij} merupakan residual pengamatan ke-i untuk variabel respon ke-j.

Model regresi multivariat yang terdiri dari q variabel respon dan p variabel prediktor dengan jumlah pengamatan sebesar n secara simultan dapat ditunjukkan dengan bentuk matriks seperti pada persamaan berikut:

$$\mathbf{Y}_{(n\times q)} = \mathbf{X}_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times q} + \boldsymbol{\varepsilon}_{(n\times q)} \quad (2)$$

Adapun tahapan dalam analisis regresi multivariat adalah:

1. Pengujian asumsi-asumsi analisis regresi multivariat. (1) Pengujian kebebasan antar variabel respon yang dapat dilakukan dengan uji *Bartlett Sphericity*, asumsi terpenuhi jika antar variabel respon tidak berkorelasi [5]. (2) Distribusi normal multivariat pada variabel respon yang dapat dilakukan dengan menggunakan nilai *square distance* (d_i^2), asumsi terpenuhi jika variabel respon berdistribusi normal multivariat [7]. (3) Deteksi multikolinieritas yang dapat dilakukan dengan menggunakan nilai VIF (*Variance Inflation Factor*). Jika nilai VIF kurang dari 10 maka dapat disimpulkan bahwa tidak terdapat masalah multikolinieritas [4].

2. Pengujian signifikansi model. (1) Pengujian serentak yang dilakukan untuk mengetahui pengaruh semua variabel prediktor secara bersama-sama terhadap variabel respon. Pengujian ini dilakukan dengan uji *Wilk's Lambda* [7] (2) Pengujian parsial yang dilakukan untuk melihat pengaruh signifikan setiap variabel prediktor terhadap variabel respon secara parsial. Pengujian ini dilakukan dengan uji *Wilk's Lambda* [7].
3. Uji Asumsi Sisaan. (1) Uji asumsi ragam galat identik yang dapat dilakukan dengan menggunakan statistik uji *Box's M*, asumsi terpenuhi jika kovarian ragam galat identik [7]. (2) Uji asumsi galat *independent* yang dapat dilakukan dengan menggunakan statistik uji *Bartlett Sphericity*, asumsi terpenuhi jika galat bersifat *independent* [5]. (3) Uji asumsi galat berdistribusi normal multivariat yang dapat dilakukan dengan menggunakan nilai *square distance* (d_i^2), asumsi terpenuhi jika galat berdistribusi normal multivariat [7].
4. Ukuran kebaikan model regresi pada analisis regresi multivariat dapat dihitung dengan menggunakan nilai *eta square lambda* [7], dengan persamaan sebagai berikut:

$$\eta_{\Lambda}^2 = 1 - \Lambda \tag{3}$$

Nilai η_{Λ}^2 berada pada interval $0 \leq \eta_{\Lambda}^2 \leq 1$. Artinya apabila nilai η_{Λ}^2 semakin mendekati 1, maka ukuran kebaikan model proporsi variabilitas variabel respon yang dijelaskan semakin besar.

Biplot Regresi Multivariat

Biplot regresi multivariat bertujuan untuk memvisualisasikan pengaruh variabel prediktor terhadap variabel respon secara grafis dan untuk menampilkan hasil regresi yang diperkenalkan oleh Oyedele (2021) [6]. Persamaan biplot regresi multivariat, dapat ditulis sebagai matriks:

$$\widehat{\mathbf{K}}_{\mathbf{MVR}} = \mathbf{Z}_{[r]}[\mathbf{V}_{[r]} \quad \mathbf{Q}_{+}]' \tag{4}$$

di mana $\mathbf{Z}_{[r]} = n \times r$ adalah matriks penanda baris dan $\begin{bmatrix}\mathbf{V}_{[r]} \\ \mathbf{Q}_{+}\end{bmatrix} = (p + q) \times r$ adalah matriks penanda kolom.

Biplot sering dibangun dalam dua dimensi, karena kemudahannya, maka $\mathbf{Z}_{[r]}$, $\mathbf{V}_{[r]}$ dan $\mathbf{Q}^{+}$ akan memiliki dua kolom ($r = 2$). Matriks $\mathbf{Z}_{[r]}$ berisi informasi tentang n data, sedangkan matriks $\mathbf{V}_{[r]}$ dan $\mathbf{Q}^{+}$ masing-masing berisi informasi tentang p variabel-X dan q variabel-Y. Oleh karena itu, untuk biplot MVR, titik data diwakili oleh n baris dari $\mathbf{Z}_{[r]}$, sedangkan arah sumbu biplot masing-masing dihitung oleh baris p dan q dari $\mathbf{V}_{[r]}$ dan $\mathbf{Q}^{+}$.

Kualitas keseluruhan dari biplot regresi multivariat dapat diketahui dengan melakukan perhitungan nilai prediktivitas sumbu (*Axis Predictivity*) dan kualitas secara keseluruhan dari biplot (*Overall quality*). Persamaan dari nilai prediktivitas sumbu (*Axis Predictivity*) dapat ditulis sebagai berikut:

$$\text{Axis Predictivity} = \text{diag}(\widehat{\mathbf{K}}'_{\mathbf{MVR[r]}}\ \widehat{\mathbf{K}}_{\mathbf{MVR[r]}})[\text{diag}(\mathbf{K}'_{\mathbf{MVR[r]}}\ \mathbf{K}_{\mathbf{MVR[r]}})]^{-1} \tag{5}$$

Nilai prediksi terletak antara 0 dan 1. Prediktivitas sumbu bernilai 1 berarti bahwa semua nilai dapat dibaca dari sumbu yang tepat. Semakin rendah nilai prediksi sumbu, semakin kurang akurat sumbu mendekati nilai yang diamati di bawah variabel tersebut.

Persamaan dari nilai kualitas secara keseluruhan dari biplot (*Overall quality*) dapat ditulis sebagai berikut:

$$\text{Overall quality} = \text{tr}(\widehat{\mathbf{K}}'_{\mathbf{MVR[r]}}\ \widehat{\mathbf{K}}_{\mathbf{MVR[r]}})[\text{tr}(\mathbf{K}'_{\mathbf{MVR[r]}}\ \mathbf{K}_{\mathbf{MVR[r]}})]^{-1} \tag{6}$$

Data

Data yang digunakan merupakan data sekunder yang diperoleh dari *website* instansi resmi yaitu Badan Pusat Statistika Jawa Barat (BPS Jawa Barat) dan Open Data Jawa Barat. Data tersebut berisikan data indikator derajat kesehatan dan faktor-faktor yang mempengaruhi indikator derajat kesehatan. Data tersebut merupakan data pada tahun 2021 yang terdiri dari 8 variabel,

yaitu, persentase angka kematian bayi (Y_1), persentase angka harapan hidup (Y_2), persentase status gizi buruk (Y_3), persentase rumah tangga yang memiliki jamban dengan anti septik (X_1), persentase rumah yang menggunakan air bersih untuk minum/masak (X_2), persentase jumlah posyandu aktif (X_3), persentase rata-rata lamanya bayi diberi ASI Eksklusif (X_4), dan persentase kepadatan penduduk (X_5).

C. Hasil Penelitian dan Pembahasan

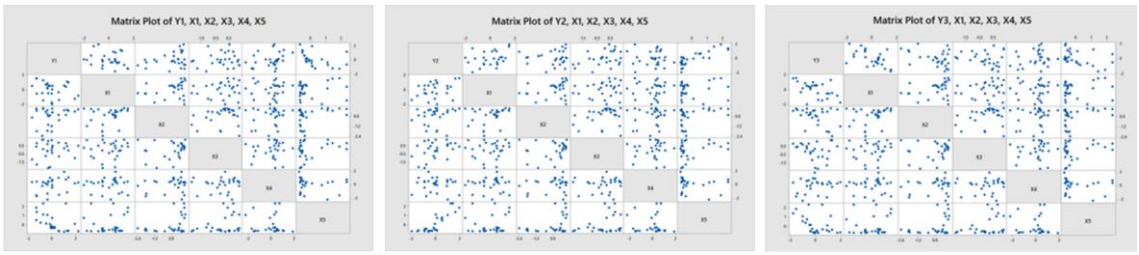
Statistika Deskriptif

Dalam penelitian ini statistika deskriptif yang digunakan adalah nilai rata-rata, nilai maksimum, nilai minimum, dan nilai standar deviasi. Statistika deskriptif dari data dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Ringkas Data Penelitian (%)

Variabel	N	Minimal	Maksimal	Rata-Rata	Standar Deviasi
Y1 (angka kematian bayi)	27	0.02	0.90	0.40	0.22
Y2 (angka harapan hidup)	27	69.69	75.19	72.45	1.40
Y3 (status gizi buruk)	27	12.30	35.20	23.06	6.09
X1 (Jamban)	27	39.64	97.54	72.59	16.20
X2 (Air)	27	77.35	100.00	93.42	6.76
X3 (Posyandu)	27	48.47	99.62	81.08	14.69
X4 (ASI)	27	42.52	90.69	68.57	11.53
X5 (Kepadatan Penduduk)	27	0.40	13.91	3.70	4.36

Data variabel prediktor dan variabel respon juga disajikan dalam bentuk *scatterplot* yang dapat dilihat pada Gambar 1.



Gambar 1. Gambar Kiri: *Scatterplot* antara variabel Y_1 dengan variabel X_1, X_2, X_3, X_4, dan X_5. Gambar Tengah: *Scatterplot* antara variabel Y_2 dengan variabel X_1, X_2, X_3, X_4, dan X_5. Gambar Kanan: *Scatterplot* antara variabel Y_3 dengan variabel X_1, X_2, X_3, X_4, dan X_5.

Pada Gambar 1 dapat diketahui bahwa data variabel respon Y1, Y2, dan Y3 dengan keseluruhan variabel prediktor cenderung menyebar secara linear dan tidak membentuk pola tertentu, sehingga data dapat dikatakan linear. Pada *scatterplot* tersebut juga dapat dilihat korelasi antara variabel prediktor, di mana terlihat bahwa titik-titik cenderung menyebar secara merata tanpa ada pola yang jelas, hal ini menunjukan bahwa tidak terdapat korelasi antar variabel prediktor.

Analisis Regresi Multivariat

1. Pengujian asumsi-asumsi analisis regresi multivariat. (1) Pengujian kebebasan antar variabel respon yang dilakukan dengan menggunakan uji *Bartlett Sphericity*, hasil yang diperoleh yaitu sebagai berikut:

$$\chi^2_{\text{hitung}} = -(24.1667)(-0.3756) = 9.077$$

Dengan nilai $\chi^2_{\text{tabel}} = \chi^2_{0.05;\frac{1}{2}3(3-1)} = \chi^2_{0.05,3} = 7.185$

Karena $\chi^2_{\text{hitung}} > \chi^2_{\text{tabel}}$ (9.077 > 7.185), maka H_0 ditolak. Sehingga dapat disimpulkan bahwa antar variabel respon bersifat *dependent* (berkorelasi) dan hasil analisis dapat dilanjutkan dengan metode regresi multivariat. (2) Distribusi normal multivariat pada variabel respon yang dilakukan dengan menggunakan nilai *square distance* (d_i^2), hasil yang diperoleh yaitu sebagai berikut:

Tabel 2. Statistik Uji d_i^2 untuk variabel respon

i	d_i^2	i	d_i^2	i	d_i^2	i	d_i^2	i	d_i^2	i	d_i^2
1	4.0034	6	4.3973	11	2.3344	16	2.9725	21	3.5532	26	5.6139
2	1.3938	7	2.2262	12	4.6381	17	1.4606	22	2.2821	27	1.4414
3	6.9289	8	1.3078	13	0.8765	18	1.2310	23	4.7917		
4	4.4628	9	0.7487	14	1.6693	19	1.4251	24	4.5648		
5	4.0782	10	3.1370	15	0.4356	20	0.4567	25	5.0892		

Dengan nilai $\chi^2_{\text{tabel}} = \chi^2_{3,0.5} = 2.366$

Karena ada 14 pengamatan atau 51.85% dari 27 pengamatan dengan $d_i^2 < \chi^2_{3,0.5}$. Maka, Dapat disimpulkan bahwa H_0 diterima. Sehingga dapat dikatakan bahwa variabel respon berdistribusi normal multivariat. (3) Deteksi multikolinieritas yang dilakukan dengan menggunkana nilai VIF (*Variance Inflation Factor*), hasil yang diperoleh yaitu sebagai berikut:

Tabel 3. Nilai *Variance Inflation Factor* X_1-X_5

Variabel	VIF
X1 (Jamban)	1.163
X2 (Air)	1.402
X3 (Posyandu)	1.101
X4 (ASI	1.102
X5 (Kepadatan Penduduk)	1.267

Tabel 3 menunjukkan bahwa masing-masing variabel prediktor memiliki nilai VIF yang kurang dari 10, maka dapat disimpulkan bahwa tidak terdapat masalah multikolinieritas.

2. Mengestimasi parameter regresi multivariat, sehingga didapatkan model regresi multivariat sebagai berikut:
   $Y_1 = -0.341X_1 + 0.486X_2 + 0.189X_3 + 0.340X_4 - 0.453X_5$
   $Y_2 = 0.289 + 0.249X_2 + 0.297X_3 - 0.208X_4 + 0.564X_5$
   $Y_3 = -0.458X_1 - 0.290X_2 - 0.045X_3 + 0.056X_4 - 0.267X_5$
3. Pengujian signifikansi model. (1) Pengujian serentak yang dilakukan dengan menggunakan uji *Wilk's Lambda*, hasil yang diperoleh yaitu sebagai berikut:
   $$\Lambda = 0.173$$
   Dengan nilai $\Lambda_{\text{tabel}} = \Lambda_{0.05,3,5,21} = 0.310$
   Karena $\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$ (0.173 < 0.310), maka H_0 ditolak. Sehingga dapat disimpulkan bahwa paling sedikit terdapat satu parameter yang berpengaruh signifikan terhadap model. (2) Pengujian Parsial yang dilakukan dengan menggunakan uji *Wilk's Lambda*, hasil yang diperoleh yaitu sebagai berikut:

Tabel 4. Pengaruh Kelima Variabel Prediktor terhadap Variabel Respon

Variabel Prediktor	Λ_{hitung}	Λ_{tabel}	Kriteria Uji	Kesimpulan
X1	-0.071	0.374	$\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$	H0 ditolak
X2	0.109	0.374	$\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$	H0 ditolak
X3	0.070	0.374	$\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$	H0 ditolak
X4	0.031	0.374	$\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$	H0 ditolak
X5	-0.367	0.374	$\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$	H0 ditolak

Karena pada semua variabel prediktor $\Lambda_{\text{hitung}} < \Lambda_{\text{tabel}}$, maka H0 ditolak. Sehingga dapat disimpulkan bahwa variabel prediktor Xk berpengaruh minimal terhadap satu variabel respon.

4. Uji asumsi sisaan. (1) Uji asumsi ragam galat identik yang dilakukan dengan menggunakan uji *Box's M*, hasil yang diperoleh yaitu sebagai berikut:

$$u = -2(1 + 0.09117)(-2.51027) = 4.5628$$

Dengan nilai $\chi^2_{\text{tabel}} = \chi^2_{0.05,\frac{1}{2}(3-1)5(5+1)} = \chi^2_{0.05,30} = 43.775$

Karena $u < \chi^2_{\text{tabel}}$ (4.5628 < 43.775), maka H_0 diterima. Sehingga dapat disimpulkan bahwa matriks varian-kovarian galat homogen atau ragam galat identik. (2) Uji asumsi galat *independent* yang dilakukan dengan menggunakan uji *Bartlett Sphericity*, hasil yang diperoleh yaitu sebagai berikut:

$$\chi^2{}_{\text{hitung}} = -(24.1667)(-0.1723) = 4.1646$$

Dengan nilai $\chi^2{}_{\text{tabel}} = \chi^2{}_{0.05;\frac{1}{2}3(3-1)} = \chi^2{}_{0.05,3} = 7.185$

Karena $\chi^2{}_{\text{hitung}} < \chi^2{}_{\text{tabel}}$ (4.1646 < 7.185), maka H_0 diterima. Sehingga dapat disimpulkan bahwa galat bersifat *independent* (tidak berkorelasi). (3) Uji asumsi galat berdistribusi normal multivariat yang dilakukan dengan menggunakan nilai *square distance* (d_i^2), hasil yang diperoleh yaitu sebagai berikut:

Tabel 5. Statistik Uji d_i^2 untuk Galat

i	d_i^2	i	d_i^2	i	d_i^2	i	d_i^2	i	d_i^2	i	d_i^2
1	1.7492	6	3.3770	11	2.9222	16	5.8158	21	0.8041	26	0.2177
2	1.6014	7	3.9588	12	3.0715	17	0.6266	22	8.8732	27	2.0286
3	7.7397	8	4.9063	13	1.4996	18	0.3651	23	0.4288		
4	3.0586	9	2.5557	14	0.5659	19	0.7084	24	2.2694		
5	1.3689	10	5.5731	15	0.6438	20	2.6774	25	8.2933		

Karena ada 14 pengamatan atau 51.85% dari 27 pengamatan dengan $d_i^2 < \chi^2{}_{3,0.5}$. Maka, Dapat disimpulkan bahwa H_0 diterima. Sehingga dapat dikatakan bahwa galat berdistribusi normal multivariat.

5. Ukuran kebaikan model regresi yang dilakukan dengan menggunakan nilai *eta square lambda*, hasil yang diperoleh yaitu sebagai berikut:

$$\eta^2_\Lambda = 1 - 0.173$$
$$= 0.827$$
$$= 82.7\%$$

Nilai $\eta^2_\Lambda = 0.827$ berarti bahwa variabel prediktor yang meliputi persentase rumah tangga yang memiliki jamban dengan anti septik (X_1), persentase rumah yang

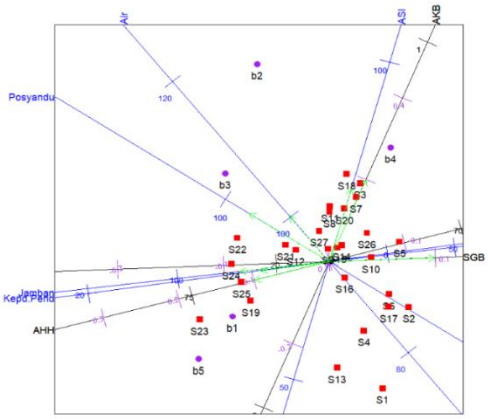
menggunakan air bersih untuk minum/masak (X_2), persentase jumlah posyandu aktif (X_3), persentase rata-rata lamanya bayi diberi ASI Eksklusif (X_4), dan kepadatan penduduk (X_5) mampu menjelaskan informasi proporsi variabilitas variabel respon yaitu angka kematian bayi (Y_1), angka harapan hidup (Y_2), dan status gizi buruk (Y_3) pada model regresi multivariat sebesar 82.7%.

Biplot Regresi Multivariat

Dari hasil analisis regresi multivariat di atas dapat dilakukan analisis menggunakan analisis biplot regresi multivariat yang diterapkan pada data faktor-faktor yang mempengaruhi indikator derajat kesehatan di Jawa Barat. Data ini dapat dilihat sebagai matriks data dengan matriks prediktor **X**: 27 × 5 dan matriks respon **Y**: 27 × 3.

Langkah pertama dalam melakukan analisis biplot regresi multivariat adalah mencari nilai matriks $\mathbf{V}_{[r]}$ yang dimana merupakan nilai PCA dari matriks **X**, nilai matriks $\mathbf{Z}_{[r]}$ dan nilai matriks $\mathbf{Q}_{+}$. Matriks $\mathbf{V}_{[r]}$, $\mathbf{Z}_{[r]}$, dan $\mathbf{Q}_{+}$ masing-masing memiliki dua kolom ($r = 2$). Dimana $\mathbf{Z}_{[r]} = 27 \times 2$ adalah matriks penanda baris dan $\begin{bmatrix}\mathbf{V}_{[r]}\\ \mathbf{Q}_{+}\end{bmatrix} = (5 + 3) \times 2$ adalah matriks penanda kolom. Selanjutnya memplotkan nilai-nilai tersebut sehingga membentuk sebuah biplot regresi multivariat.

Hasil dari analisis biplot regresi multivariat dapat dilihat pada Gambar 2. Pada gambar tersebut data diwakili oleh titik merah, sedangkan titik ungu untuk titik koefisien regresi multivariat. Variabel prediktor dari data diwakili oleh sumbu biru, sedangkan variabel respon diwakili sumbu hitam.



Gambar 2. Biplot Regresi Multivariat antara Faktor-Faktor yang Mempengaruhi Indikator Derajat Kesehatan dengan Indikator Derajat Kesehatan.

Pada Gambar 2 dapat dilihat hubungan antar variabel, dimana:

1. Variabel angka kematian bayi (AKB) berkorelasi negatif lemah dengan variabel persentase rumah tangga yang memiliki jamban dengan anti septik (X_1), berkorelasi positif sedang dengan variabel persentase rumah yang menggunakan air bersih untuk minum/masak (X_2), berkorelasi positif lemah dengan variabel persentase jumlah posyandu aktif (X_3), berkorelasi positif kuat dengan variabel persentase rata-rata lamanya bayi diberi ASI Eksklusif (X_4), dan berkorelasi negatif sedang dengan variabel kepadatan penduduk (X_5). Variabel angka kematian bayi (AKB) memiliki pengaruh yang kuat dengan variabel persentase rata-rata lamanya bayi diberi ASI Eksklusif (X_4).
2. Variabel angka harapan hidup (AHH) berkorelasi positif kuat dengan variabel persentase rumah tangga yang memiliki jamban dengan anti septik (X_1), berkorelasi positif lemah dengan variabel persentase rumah yang menggunakan air bersih untuk minum/masak (X_2), berkorelasi positif sedang dengan variabel persentase jumlah posyandu aktif (X_3), berkorelasi negatif lemah dengan variabel persentase rata-rata lamanya bayi diberi ASI

Eksklusif (X_4), dan berkorelasi positif kuat dengan variabel kepadatan penduduk (X_5). Variabel angka harapan hidup (AHH) memiliki pengaruh yang kuat dengan variabel persentase rumah tangga yang memiliki jamban dengan anti septik (X_1) dan kepadatan penduduk (X_5).

3. Variabel status gizi buruk (SGB) berkorelasi negatif kuat dengan variabel persentase rumah tangga yang memiliki jamban dengan anti septik (X_1), berkorelasi negatif lemah dengan variabel persentase rumah yang menggunakan air bersih untuk minum/masak (X_2), berkorelasi negatif sedang dengan variabel persentase jumlah posyandu aktif (X_3), berkorelasi positif lemah dengan variabel persentase rata-rata lamanya bayi diberi ASI Eksklusif (X_4), dan berkorelasi negatif kuat dengan variabel kepadatan penduduk (X_5). Variabel status gizi buruk (SGB) memiliki pengaruh yang kuat dengan variabel persentase rumah tangga yang memiliki jamban dengan anti septik (X_1) dan kepadatan penduduk (X_5).

Biplot regresi multivariat juga menunjukkan representasi varian dari masing-masing variabel, yang diwakili oleh tanda panah (vektor) berwarna hijau. Dari biplot tersebut dapat dilihat bahwa variabilitas dari variabel persentase rumah yang menggunakan air bersih untuk minum/masak (X_2) lebih kecil dibandingkan dengan variabel lainnya. Sedangkan variabilitas dari variabel persentase jumlah posyandu aktif (X_3) lebih besar dibandingkan dengan variabel lainnya. Hal ini dapat dilihat berdasarkan panjang vektor dari setiap sumbu.

Selain itu, dalam prosesnya biplot regresi multivariat juga menampilkan hasil dari koefisien regresi. Hasil koefisien regresi dari biplot regresi multivariat dari data yang distandarisasi sebagai berikut:

$$Y_1 = -0.041X_1 + 0.059X_2 + 0.007X_3 + 0.301X_4 - 0.144X_5$$
$$Y_2 = 0.137X_1 + 0.215X_2 + 0.214X_3 - 0.184X_4 + 0.353X_5$$
$$Y_3 = -0.108X_1 - 0.168X_2 - 0.157X_3 + 0.089X_4 - 0.239X_5$$

Meskipun biplot regresi multivariat juga menghasilkan koefisien regresi, namun koefisien regresi yang dihasilkan tidak sebaik yang dihasilkan melalui analisis regresi multivariat. Secara keseluruhan kualitas representasi hubungan sebab akibat variabel dalam bentuk matriks koefisien regresi yang dihasilkan biplot regresi multivariat yaitu $0.515/1.670 = 30.86\%$, yang merupakan kualitas yang rendah.

Untuk mengetahui seberapa baik kualitas keseluruhan dari biplot MVR, diperlukan perhitungan kualitas secara keseluruhan dari biplot (*Overall quality*) dan nilai prediktivitas sumbu (*Axis Predictivity*). Nilai prediktivitas sumbu (*Axis Predictivity*) pada setiap sumbu biplot pada Gambar 2 dapat dilihat pada Tabel 6.

Tabel 6. Nilai Prediktivitas Sumbu Biplot Regresi Multivariat

Jamban	Air	Posyandu	ASI	Kepadatan Penduduk	AKB	AHH	SGB
0.972	0.998	0.980	0.986	0.796	0.850	0.999	0.954

Berdasarkan Tabel 6 dapat dilihat bahwa masing-masing sumbu mewakili data asli dengan cukup baik. Variabel angka harapan hidup (AHH) memiliki nilai prediktivitas sumbu tertinggi, yaitu sebesar 0.999, nilai tersebut hampir mendekati 1. Artinya, bahwa sumbu variabel AHH mewakili data asli dengan sangat baik secara visualisasi. Namun, variabel kepadatan penduduk memiliki nilai prediktivitas sumbu terendah, yaitu sebesar 0.796, nilai tersebut hampir mendekati 1. Artinya, bahwa sumbu variabel kepadatan penduduk mewakili data asli dengan baik secara visualisasi, tetapi tidak sebaik sumbu lainnya.

Berdasarkan perhitungan nilai kualitas secara keseluruhan dari biplot (*Overall quality*) diperoleh bahwa kualitas keseluruhan dari biplot regresi multivariat dalam bentuk visualisasi adalah 98.7%. Artinya, informasi yang diberikan biplot regresi multivariat dalam bentuk visualisasi mampu menerangkan 98.7% dari keseluruhan informasi yang terkandung dalam data.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Berdasarkan model regresi dari kedua analisis dapat dilihat bahwa representasi hubungan antar variabel secara keseluruhan sama. Namun, antar kedua model regresi memiliki koefisien yang berbeda. Hal ini disebabkan karena pada analisis regresi multivariat hasil koefisien regresi diperoleh berdasarkan nilai asli antara variabel prediktor dan variabel respon, sedangkan pada analisis biplot regresi multivariat hasil koefisien regresi diperoleh melalui nilai-nilai pendekatan biplot. Oleh karena itu, koefisien regresi pada biplot regresi multivariat memiliki kualitas yang lebih rendah jika dibandingkan dengan koefisien regresi pada analisis regresi multivariat.
2. Berdasarkan hasil analisis biplot regresi multivariat maupun analisis regresi multivariat dihasilkan bahwa semua variabel faktor-faktor yang mempengaruhi indikator derajat kesehatan berpengaruh secara signifikan terhadap variabel indikator derajat kesehatan.
3. Diperoleh kualitas keseluruhan dari biplot regresi multivariat dalam bentuk visualisasi adalah 98.7%.

Acknowledge

Penelitian ini merupakan bagian dari tugas akhir dari penulis pertama di bawah bimbingan penulis kedua.

Daftar Pustaka

[1] Depkes, RI. (2003). *Penyelenggaraan Makanan Institusi*. Jakarta: Departemen Kesehatan RI.

[2] Johnson R. A., & Wichern D. W. (2002). *Applied Multivariate Statistical Analysis Six Edition*. New Jersey.

[3] Martens, H., & Naes, T. (1989). *Multivariate calibration*. Wiley, Chichester, U. K.

[4] Myers, R. H. (1990). *Clasical and Modern Regression With Application*. PWSKENT publishing Company, Boston.

[5] Morrison, D. F. (2016). *Multivariate Statistical Methods Fourth Edition*. Pennsylvania: The Wharton School University of Pennsylvania.

[6] Oyedele, O. F. (2021). *Extension of Biplot Methodology to Multivariate Regression Analysis*. *Journal of Applied Statistics*, 48(10), 1816-1832.

[7] Rencher A. C. (2002) *Method of Multivariat Analysis Second Edition*. John Wiley and Sons Inc, New York.

[8] Tawangki Sri Fadilah and Abdul Kudus, "Penerapan Metode Regresi Kernel Smoothing untuk Imputasi Data Lama Waktu Terinfeksi Covid-19," *Jurnal Riset Statistika*, pp. 51–60, Jul. 2023, doi: 10.29313/jrs.v3i1.1802.