

PENERAPAN ANALISIS KOMPONEN UTAMA DALAM MEREDUKSI FAKTOR-FAKTOR PENYEBAB DIARE DI PROVINSI MALUKU

THE APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO REDUCE DIARRHEA FACTORS IN MALUKU PROVINCE

Gabriella Haumahu^{1§}, Norisca Lewaherilla²

¹ Program Studi Statistika, Universitas Pattimura, Ambon, Indonesia [Email: gbhaumahu@gmail.com]

² Program Studi Statistika, Universitas Pattimura, Ambon, Indonesia [Email: lewaherillanorisca@gmail.com]

[§]Corresponding Author

Received Mei 2020; Accepted Juni 2020; Published Juni 2020;

Abstrak

Analisis komponen utama merupakan suatu analisis untuk menjelaskan struktur varians-kovarians melalui sejumlah kecil kombinasi linier dari segugus variabel asal. Tujuan umumnya adalah untuk reduksi data dan interpretasi. Penelitian ini bertujuan untuk mereduksi faktor-faktor penyebab diare di Provinsi Maluku. Variabel yang digunakan dalam penelitian ini adalah jumlah penduduk miskin (X1), jumlah rumah sakit umum (X2), jumlah puskesmas (X3), jumlah apotek (X4), jumlah rumah tangga yang menggunakan fasilitas tempat buang air sendiri/not shared (X5), jumlah rumah tangga yang tidak menggunakan dan tidak mempunyai fasilitas tempat buang air besar (X6), dan jumlah rumah tangga yang memiliki akses sumber air minum layak (X7). Hasil penelitian menunjukkan variabel-variabel direduksi menjadi satu komponen utama dengan proporsi varians kumulatifnya sebesar 82%.

Kata Kunci: analisis multivariat, analisis komponen utama, diare

Abstract

Principal component analysis is an analysis to explain the structure of variance-covariance through a small number of linear combinations of a set of original variables. The general purpose is to reduce data and interpretation. This study aims to reduce the causes of diarrhea in Maluku Province. The variables used in this study are the number of poor people (X1), the number of public hospitals (X2), the number of Puskesmas (X3), the number of pharmacies (X4), the number of households that use their own not shared toilet facilities (X5), the number of households that do not use and do not have defecation facilities (X6), and the number of households that have access to improved drinking water sources (X7). The results showed the variables were reduced to one main component with a cumulative proportion of 82%.

Keywords: *multivariate analysis, principal component analysis, diarrhea*

1. Pendahuluan

Hingga saat ini penyakit diare masih merupakan masalah kesehatan masyarakat di

Indonesia. Hal ini dapat dilihat dari meningkatnya angka keseakitan dari tahun ke tahun. Di dunia, sebanyak 6 juta anak meninggal setiap tahun karena diare, yang mana sebagian kematian tersebut terjadi di negara berkembang. Berdasarkan laporan WHO, kematian karena diare di Indonesia sudah menurun tajam. Berdasarkan survei rumah tangga, kematian karena diare diperkirakan menurun. Penyakit diare masih merupakan masalah kesehatan masyarakat di negara berkembang seperti di Indonesia, karena masih sering timbul dalam bentuk kejadian luar biasa (KLB) dan disertai dengan kematian yang tinggi, terutama di Indonesia bagian timur [1].

Faktor-faktor yang berhubungan dengan diare antara lain sanitasi lingkungan, ketersediaan air bersih, *hygiene* perorangan, sanitasi makanan, ketersediaan jamban, dan perilaku buang tinja [2]. Faktor-faktor inilah yang menyebabkan naiknya jumlah penderita daerah di suatu wilayah.

Analisis komponen utama (AKU) bertujuan untuk mereduksi faktor-faktor tersebut kedalam komponen utama. Dalam penelitian [3] AKU digunakan untuk mereduksi faktor-faktor inflasi di Kota Ambon, berkaitan dengan penyakit, pada penelitian [4] AKU digunakan untuk mereduksi faktor-faktor yang mempengaruhi penyakit jantung koroner, dan dalam penelitian ini AKU digunakan untuk mereduksi faktor-faktor penyebab diare di Provinsi Maluku.

2. Landasan Teori

2.1 Analisis Komponen Utama

Analisis komponen utama merupakan suatu analisis untuk menjelaskan struktur varians-kovarians melalui sejumlah kecil kombinasi linier dari segugus variabel asal. Tujuan umumnya adalah untuk reduksi data dan interpretasi. Komponen utama yang dihasilkan dapat menjelaskan sebanyak mungkin variasi total dalam data melalui sedikit faktor. Komponen utama diekstrak sedemikian rupa sehingga komponen utama ke- p yang dinyatakan dengan Y_p merupakan kombinasi linier dari variabel pengamatan X_j untuk $j = 1, 2, \dots, p$ yaitu:

$$Y_p = w_{1p}X_1 + w_{2p}X_2 + \dots + w_{pp}X_p$$

Dengan bobot w dipilih untuk memaksimalkan rasio dari varians Y [5].

2.2 Pengujian Normal Multivariat

Pengujian normal multivariat dilakukan dengan uji Henze Zirkler's dengan rumusan hipotesis dan statistik uji sebagai berikut [6].

$$H_0: X \text{ berdistribusi } N_p(\mu, \Sigma)$$

$$H_1: X \text{ tidak berdistribusi } N_p(\mu, \Sigma)$$

$$HZ = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n e^{-\frac{\beta^2}{2} D_{ij}} - 2(1 + \beta^2)^{-\frac{p}{2}} \frac{1}{n} \sum_{i=1}^n e^{-\frac{\beta^2}{2(1+\beta^2)} D_i} + (1 + 2\beta^2)^{-\frac{p}{2}}$$

dengan

$$\beta = \frac{1}{\sqrt{2}} \left(\frac{n(2p + 1)}{4} \right)^{\frac{1}{p+4}}$$

$$D_{ij} = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j)$$

$$D_i = (x_i - \bar{x})^T S^{-1} (x_i - \bar{x})$$

Apabila nilai *p-value* yang dihasilkan lebih besar dari taraf signifikansi yang ditentukan, maka dapat disimpulkan data berdistribusi normal multivariat.

2.3 Bartlett Test of Sphericity dan Nilai Keiser-Meyers-Oklin(KMO)

Analisis komponen utama didasarkan pada matriks korelasi. Matriks korelasi digunakan untuk mendapatkan nilai kedekatan hubungan antar variabel penelitian [7].

a. Uji Bartlett

Pengujian dengan uji Bartlett digunakan untuk melihat apakah matriks korelasinya merupakan matriks identitas [8].

H_0 : matriks korelasi merupakan matriks identitas

H_1 : matriks korelasi bukan matriks identitas

Statistik uji:
$$\chi^2_{obs} = - \left[(N - 1) - \frac{(2p+5)}{6} \right] \ln|R|$$

dengan:

- N : jumlah observasi
- p : jumlah variabel
- |R| : determinan matriks korelasi

Keputusan:

H_0 : diterima jika $\chi^2_{obs} < \chi^2_{\sigma,p(p-1)/2}$

H_0 : ditolak jika $\chi^2_{obs} \geq \chi^2_{\sigma,p(p-1)/2}$

b. Uji Kaiser Meyer Olkin (KM)

Uji KMO ini digunakan untuk melihat apakah data tersebut layak dan dapat

dianalisis dengan analisis komponen utama [9]. Nilai KMO dan *Measure of Sampling Adequacy* (MSA) dianggap layak digunakan apabila besaran KMO > 0,5 dan kriteria MSA>0,5. Nilai statistik KMO sebagai berikut.

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum \sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} a_{ij}^2}, i = 1, 2, \dots, p; j = 1, 2, \dots, p$$

dengan:

r_{ij} = koefisien korelasi sederhana antara variabel ke-*i* dan ke-*j*

a_{ij} = koefisien korelasi parsial antara variabel ke-*i* dan ke-*j*

Nilai KMO yang kecil mengindikasikan bahwa penggunaan analisis faktor harus dipertimbangkan kembali, karena korelasi antar variabel tidak dapat diterangkan oleh variabel lain.

2.4 Penentuan Komponen Utama Berdasarkan Nilai Eigen

Nilai eigen merupakan suatu nilai yang menunjukkan seberapa besar pengaruh suatu variabel terhadap pembentukan karakteristik yang dinotasikan dengan λ . Faktor penentuannya adalah dengan mempertahankan nilai eigen yang lebih dari 1. Selanjutnya dengan melihat nilai variansi yang dapat dijelaskan lebih dari 80%.

3. Hasil Dan Pembahasan

Sebagai langkah awal dalam melakukan analisis komponen utama adalah menguji

normalitas data multivariate dengan menggunakan Uji Henze Zirkler's. Data mentah yang diteliti adalah data sekunder yang diperoleh dari BPS (Badan Pusat Statistik) Provinsi Maluku melalui publikasi oleh lembaga-lembaga terkait tahun 2019 per kabupaten/kota dan diolah dengan *R software* [10]. Variabel yang digunakan dalam penelitian ini terdiri dari 7 variabel, yaitu jumlah penduduk miskin (X1), jumlah rumah sakit umum(X2), jumlah puskesmas (X3), jumlah apotek (X4), jumlah rumah tangga yang menggunakan fasilitas tempat buang air sendiri/*not shared* (X5), jumlah rumah tangga yang tidak menggunakan dan tidak mempunyai fasilitas tempat buang air besar (X6), dan jumlah rumah tangga yang memiliki akses sumber air minum layak (X7). Datanya sebagai berikut.

Tabel 1. Data Penelitian

X1	X2	X3	X4	X5	X6	X7
30760	4	12	3	16933	1726	12757
22690	3	19	2	14674	2077	13211
74800	4	35	10	61770	10956	46464
23890	1	13	10	21765	5450	14286
25620	1	30	2	8863	3802	11799
43140	1	17	5	20337	9657	14684
26440	2	22	5	11165	7356	12850
21490	1	30	3	8819	3141	9726
10170	1	12	2	6726	3176	6341
21660	10	21	20	84677	2999	39977
17030	1	15	1	11873	902	4763

Sumber: BPS Provinsi Maluku

Selanjutnya akan diuji kenormalan data dengan menggunakan *Henze-Zirkler's Multivariate Normality Test*, dengan hasil sebagai berikut.

Tabel 2. Uji Normalitas Data

<i>Henze-Zirkler's Multivariate Normality Test (p-value)</i>
0.005845

Sumber: Hasil Penelitian (Software R)

Berdasarkan nilai *p-value* dari uji Henze Zirkler's menunjukkan data tidak berdistribusi normal karena *p-value* <0,05 sehingga akan dilakukan proses transformasi data menggunakan logaritma, dengan $Y_i = \log X_i$ dan hasilnya terlihat pada tabel berikut.

Tabel 3. Uji Normalitas Data Transformasi

<i>Henze-Zirkler's Multivariate Normality Test (p-value)</i>
0.1574

Sumber: Hasil Penelitian (Software R)

Berdasarkan Tabel 3, nilai *p-value* dari uji Henze Zirkler's pada data hasil transformasi menggunakan fungsi logaritma menunjukkan data transformasi berdistribusi normal, karena *p-value* =0.1574. Uji selanjutnya yakni uji Kaiser Meyer Olkin (KMO), Bartlett dan MSA. Hasilnya disajikan dalam tabel berikut.

Tabel 4. Uji KMO, Bartlett, dan MSA

Variabel	KMO & Bartlett Test	MSA
Y1	KMO = 0.6028159 Bartlett's Test (X-squared) = 56.2069	0.6729799
Y2		0.5393110
Y3		0.4328951
Y4		0.6846136
Y5		0.6776200
Y6		0.4729242
Y7		0.6253256

Sumber: Hasil Penelitian (Software R)

Pada Tabel 4, hasil analisis menunjukkan nilai KMO yaitu 0.6028159 atau lebih dari 0.5 yang menandakan kumpulan variabel dapat diproses lebih lanjut. Untuk nilai MSA, pada tabel terlihat untuk data Y3 dan Y6 memiliki nilai MSA < 0,5 sehingga dua variabel tersebut harus dikeluarkan dari pengujian dan dilakukan pengujian ulang tanpa mengikutsertakan dua variabel dimaksud.

Hasil pengujian ulang disajikan dalam tabel berikut.

Tabel 5. Uji KMO, Bartlett, dan MSA (Pengujian Ulang Tahap 1)

Variabel	KMO & Bartlett Test	MSA
Y1	KMO = 0.6287303	0.4735263
Y2		0.5586095
Y4		0.6009442
Y5		0.8462501
Y7	Bartlett's Test (X-squared) = 39.2143	0.6387925

Sumber: Hasil Penelitian (Software R)

Pada Tabel 5, nilai KMO menandakan data dapat diproses lebih lanjut, tetapi masih terdapat satu variabel, yakni variabel Y1 yang nilai MSA <0.5, sehingga akan dilakukan pengujian ulang tanpa melibatkan variabel Y1.

Tabel 6. Uji KMO, Bartlett, dan MSA (Pengujian Ulang Tahap 2)

Variabel	KMO & Bartlett Test	MSA
Y2	KMO = 0.7908954	0.7889410
Y4		0.7685613
Y5		0.8232256
Y7		0.7809702
	Bartlett's Test (X-squared) = 31.33	

Sumber: Hasil Penelitian (Software R)

Tabel 6 menunjukkan nilai MSA keempat variabel >0,5, maka data dapat dilakukan pengujian lanjut.

Tabel 7. Importance of components

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.3773722	0.5322774	0.27571812	0.21321318
Proportion of Variance	0.8241495	0.1230777	0.03302433	0.01974838
Cumulative Proportion	0.8241495	0.9472273	0.98025162	1.00000000

Sumber: Hasil Penelitian (Software R)

Tabel 7 menjelaskan proporsi varians kumulatif komponen pertama dapat menjelaskan 82% total varians dan bila ditambahkan dengan komponen 2 menjadi 95% [11]. Jika komponen

pertama saja yang diambil, itu berarti sudah mencukupi. Nilai standar deviasi yang ditampilkan dalam Tabel 7 diperoleh dari akar positif nilai eigen matriks kovariansi. Selanjutnya nilai loadings tiap komponen akan dijelaskan lewat tabel berikut [12].

Tabel 8. Nilai Loadings

	Comp.1	Comp.2	Comp.3	Comp.4
Y2	-0.464	0.783	0.389	-0.139
Y4	-0.557	-0.617	0.473	-0.291
Y5	-0.527		-0.788	-0.318
Y7	-0.443			0.892

Sumber: Hasil Penelitian (Software R)

Berdasarkan Tabel 8, nilai loadings adalah nilai vector eigen dari matriks kovariansi dan sesuai dengan nilai proporsi varians kumulatif yang ditampilkan pada Tabel 7, maka fungsi komponen utama $KU_1 = -0.464 \log(Y2) - 0.557 \log(Y4) - 0.527 \log(Y5) - 0.443 \log(Y7)$ dapat menjelaskan 82% total varians.

4. Kesimpulan Dan Saran

Berdasarkan pembahasan mengenai penggunaan analisis komponen utama dalam mereduksi faktor-faktor penyebab diare di Provinsi Maluku dapat disimpulkan dari 7 variabel yang diteliti terdapat 4 variabel hasil reduksi yaitu variabel jumlah rumah sakit umum(X2), jumlah apotek (X4), jumlah rumah tangga yang menggunakan fasilitas tempat buang air sendiri/not shared (X5), dan jumlah rumah tangga yang memiliki akses sumber air minum layak (X7) dengan total varians sebesar 82%.

Daftar Pustaka

- [1] Kementerian Kesehatan RI. 2012. Buletin Jendela Data Informasi Kesehatan - Situasi Diare di Indonesia. Kementerian Kesehatan RI.
- [2] H. F. Rahman, S. Widoyo, H. Siswanto and Biantoro. 2016. Faktor-faktor yang Berhubungan Dengan Kejadian Diare di Desa Solor Kecamatan Cermee Bondowoso, *NurseLine Journal*, pp. 24-35.
- [3] M. S. Noya Van Delsen, A. Z. Wattimena and S. D. Saputri. 2017. Penggunaan Metode Analisis Komponen Utama untuk Mereduksi Faktor-faktor Inflasi di Kota Ambon. *Barekeng: Jurnal Ilmu Matematika dan Terapan*, pp. 109-118.
- [4] G. H. M, T. B. Adji and N. A. Setiawan. 2012. Penggunaan Metodologi Analisa Komponen Utama (PCA) untuk Mereduksi Faktor-faktor yang Mempengaruhi Penyakit Jantung Koroner. *Seminar Nasional "Science, Engineering, and Technology"*, Yogyakarta.
- [5] N. Hajarisman. 2008. Seri Buku Ajar Statistika. Program Studi Statistika Universitas Islam Bandung. Bandung.
- [6] N. Henze and B. Zirkler. 1990. A Class of Invariant Consistent Tests for Multivariate Normality. *Communications in Statistics - Theory and Methods*, pp. 3595-3617.
- [7] I. T. Jolliffe. 2002. *Principal Component Analysis*. Second Edition. Springer. New York.
- [8] J. F. Hair, R. Anderson, R. I. Tatham and W. C. Black. 1995. *Multivariate Data Analysis With Readings*, 4th Edition. Prentice Hall: Englewood Cliffs, NJ
- [9] M. A. Supranto. 2004. *Analisis Multivariat (Anti dan Interpretasi)*. Rineka Cipta. Jakarta.
- [10] Badan Pusat Statistik. 2020. Provinsi Maluku Dalam Angka, Katalog: 11020001.81. BPS Provinsi Maluku. Ambon.
- [11] B. Tantular. 2011. "www.academia.edu," [Online]. Available: https://www.academia.edu/38570912/Praktikum_Analisis_Data_Multivariat_II_Menggunakan_Software_R_MODUL_I_ANALISIS_KOMPONEN_UTAMA. [Accessed 3 May 2020].
- [12] J. Maindonald and W. J. Braun. 2006. *Data Analysis and Graphics Using R – an Example – Based Approach*. Cambridge University. UK.