

**APLIKASI ANALISIS REGRESI KOMPONEN UTAMA
TERHADAP FAKTOR-FAKTOR YANG MEMPENGARUHI
PENYAKIT DIABETES MELLITUS
(Studi Kasus di Rumah Sakit Umum Dr. M. Haulussy Ambon)**

Y. A. Lesnussa dan Jenifer Pentury

Staff Pengajar Jurusan Matematika FMIPA Universitas Pattimura, Ambon
E-mail: yopi_a_lesnussa@yahoo.com

Diterima 10 Agustus 2015/Disetujui 17 September 2015

Abstract

Diabetes mellitus is disease characterized by increased of blood sugar levels that are excessive and occur in the long term. Factors that influence diabetes mellitus were age, body weight, systolic blood pressure, diastolic blood pressure, cholesterol, uric acid, hemoglobin, and pulse. The analysis was performed on 40 samples of patients with Diabetes Mellitus by using principal component regression analysis and meet all testing regression coefficients simultaneously, testing regression coefficients individually and classical assumption test. Then, based on the analysis of data, will obtain the best regression model and the dominant factors that cause diabetes mellitus hemoglobin.

Keyword: Diabetes Mellitus, Multiple Linear Regression Analysis, Multicollinearity, Principal Component Regression.

Abstrak

Penyakit Diabetes Mellitus ditandai dengan peningkatan kadar gula darah yang berlebihan dan occur in the jangka panjang. Faktor-faktor yang mempengaruhi diabetes mellitus adalah usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, kolesterol, asam urat, hemoglobin, dan pulsa. Analisis ini dilakukan pada 40 sampel dari pasien dengan Diabetes Mellitus dengan menggunakan analisis komponen utama regresi dan memenuhi semua koefisien regresi pengujian secara bersamaan, pengujian koefisien regresi secara individual dan uji asumsi klasik. Kemudian, berdasarkan analisis data, akan mendapatkan model regresi terbaik dan faktor dominan yang menyebabkan diabetes mellitus hemoglobin.

Kata Kunci: Diabetes Mellitus, Analisis Regresi Linear Berganda, Multikolinearitas, Principal Component Regression.

PENDAHULUAN

Diabetes Mellitus (DM) atau dalam kehidupan sehari-hari dikenal sebagai penyakit kencing manis adalah suatu

penyakit yang ditandai dengan meningkatnya kadar gula dalam darah yang berlebihan dan terjadi dalam jangka waktu yang panjang. Penderita diabetes tidak bisa

memproduksi insulin dalam jumlah yang cukup, sehingga terjadi kelebihan gula dalam tubuh. Ada beberapa faktor yang mempengaruhi penyakit diabetes mellitus diantaranya usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, kolesterol LDL, kolesterol HDL, asam urat, hemoglobin, dan denyut nadi.

Untuk mengetahui pengaruh faktor-faktor tersebut terhadap tingkat kadar gula dalam darah maka dapat digunakan suatu metode statistik yakni analisis regresi berganda. Tetapi untuk mengidentifikasi adanya kasus multikolinieritas pada faktor-faktor tersebut maka digunakan analisis regresi komponen utama.

Analisis regresi komponen utama juga merupakan suatu analisis regresi yang digunakan untuk mengetahui hubungan antara usia, berat badan, tekanan darah sistolik, tekanan darah diastolik, kolesterol LDL, kolesterol HDL, asam urat, hemoglobin, dan denyut nadi dengan penyakit DM yang tidak saling berkorelasi atau tidak terjadi multikolinieritas.

Analisis ini bertujuan untuk mendapatkan model terbaik dari data pasien penderita diabetes mellitus dengan menggunakan analisis regresi komponen utama dan mengetahui koefisien terbesar dari model regresi komponen utama untuk memperoleh faktor dominan penyebab penyakit diabetes mellitus.

METODE PENELITIAN

Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data arsip rekam medik pasien penderita diabetes mellitus sebanyak 40 data di RSUD Dr. M. Haulussy Ambon pada tahun 2013 sampai 2015.

Bahan atau Materi Penelitian

Variabel terikat Y yang digunakan dalam penelitian ini adalah kadar gula darah

pasien penderita Diabetes Mellitus. Sedangkan variabel bebas (X_j) yang digunakan antara lain: usia (x_1), berat badan (x_2), tekanan darah sistolik (x_3), tekanan darah diastolik (x_4), kolesterol (x_5), asam urat (x_6), hemoglobin (x_7), dan denyut nadi (x_8).

Untuk mendukung proses penelitian digunakan program komputer yaitu software SPSS versi 16 serta literatur pendukung dalam bentuk buku cetak dan informasi ilmiah lainnya yang diperoleh dari materi-materi kuliah maupun media elektronik seperti internet.

Teknik Analisis Data

Teknik analisis data dalam penelitian ini menggunakan metode analisis Regresi Komponen Utama. Langkah-langkahnya adalah sebagai berikut:

1. Melakukan analisis regresi linear berganda.
2. Interpretasi hasil analisis regresi linear berganda.
 - a. R^2
 - b. Uji koefisien regresi linear berganda secara serentak
 - c. Uji koefisien regresi linear berganda secara individual
3. Identifikasi multikolinieritas dengan melihat R^2 dan uji koefisien regresi linier secara individu.
4. Untuk memastikan adanya kasus multikolinieritas gunakan VIF dan *Tolerance*.
5. Melakukan analisis komponen utama
 - a. Menentukan nilai eigen dan mengambil akar cirri yang lebih dari 1 ($\lambda_j > 1$)
 - b. Menentukan komponen matriks
6. Melakukan analisis regresi komponen utama.
7. Menentukan persamaan regresi yaitu mensubstitusikan komponen utama yang

merupakan kombinasi linear yang terbobot dari variabel asal.

8. Interpretasi hasil analisis regresi komponen utama
 - a. R^2
 - b. Uji signifikansi parameter

HASIL DAN PEMBAHASAN

Pada penelitian ini, data yang digunakan merupakan data arsip rekam medik pasien penderita Diabetes Mellitus sebanyak 40 data (Lampiran 1) di RSUD Dr. M. Haulussy Ambon pada tahun 2013 sampai 2015.

Analisis Regresi Linear Berganda

Dari data rekam medik pasien penderita Diabetes Mellitus, diperoleh model Regresi Linear Berganda dengan menggunakan metode kuadrat terkecil untuk menduga koefisien regresi adalah:

$$\hat{Y} = -464,230 + 0,304x_1 - 0,371x_2 + 0,811x_3 + 1,490x_4 - 0,334x_5 + 20,348x_6 + 4,021x_7 + 5,403x_8$$

Dari hasil analisis regresi linier berganda diperoleh nilai R^2 sebesar 97,3%

menunjukkan bahwa hubungan antara kadar gula darah dengan variable bebasnya sebesar 97,3% sedangkan sisanya 2,7% disebabkan oleh faktor-faktor lainnya untuk selanjutnya dilakukan uji serentak dan uji individual untuk melihat pengaruh secara serentak dan individual antara variabel bebas dan variabel terikatnya. Dari hasil analisis diperoleh nilai signifikan $F = 0,000 < \alpha = 0,05$ dapat dikatakan bahwa variabel bebas secara bersama-sama mempengaruhi variabel terikat. Untuk pengujian koefisien regresi secara individu diperoleh nilai signifikan $t < \alpha = 0,05$ hanya terdapat pada x_6 (asam urat) dan x_8 (denyut nadi), sehingga dapat dikatakan secara individual x_6 dan x_8 mempengaruhi variable terikat sedangkan variable bebas lainnya tidak mempengaruhi variable terikat. Sehingga diduga terjadi kasus multikolinieritas pada data ini. Untuk memastikan hal tersebut pada uji multikolinieritas akan digunakan pemeriksaan nilai VIF dan *Tolerance*.

Uji Multikolinieritas

Tabel 1. Nilai *Tolerance* dan VIF

Prediktor	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
<i>Tolerance</i>	0,036	0,022	0,048	0,053	0,062	0,048	0,021	0,074
VIF	28,049	44,825	20,936	18,858	16,116	20,923	46,715	13,476

Berdasarkan hasil Tabel 1 nilai *Tolerance* dan VIF diperoleh nilai VIF untuk semua variabel bebas lebih dari 10,0 dan nilai *Tolerance* nya kurang dari 0,1 sehingga dapat disimpulkan terjadi multikolinieritas antar variabel bebas.

Analisis Komponen Utama

Dalam ilmu statistika untuk mengatasi masalah multikolinieritas, dapat digunakan analisis komponen utama.

H_0 = tidak terdapat korelasi antar variabel bebas

H_1 = terdapat korelasi antar variabel bebas

Tabel 2. KMO and Bartlett's test

<i>Kaiser-Meyer-Olkin-Measure</i>		0,907
<i>Bartlett's Test of Sphericity</i>	<i>Approx. Chi-Square</i>	672,501
	<i>Df</i>	28
	<i>Sig. Bartlett</i>	0,000

Dari Tabel 2 dapat dilihat nilai KMO adalah 0,907 yang mendekati 1, artinya analisis dapat dilanjutkan. Nilai *chi-Square* adalah 672,501 dengan derajat bebas sebesar 28 dan nilai signifikan sebesar 0,000.

Karena nilai signifikan *bartlett* 0,000 < $\alpha = 0,05$ maka H_0 ditolak artinya terdapat

korelasi antar variabel bebas. Selanjutnya adalah melihat tabel *communalities* yang menunjukkan berapa varians yang dapat dijelaskan oleh komponen yang terbentuk. Hasil varians tersebut dapat ditunjukkan dalam tabel dibawah ini:

Tabel 3. Communalities

	Usia (x_1)	BB (x_2)	TDS (x_3)	TDD (x_4)	Kolesterol (x_5)	AU (x_6)	HB (x_7)	DN (x_8)
<i>Extraction</i>	0,905	0,935	0,941	0,936	0,939	0,911	0,974	0,915

Hasil yang diperoleh dari tabel 3 memperlihatkan nilai variabel usia sebesar 0,905 yang berarti sekitar 90,5% variansi variabel usia dapat dijelaskan oleh komponen yang terbentuk, variabel berat badan (BB) sebesar 0,935 yang berarti sekitar 93,5% variansi variabel BB dapat dijelaskan oleh komponen yang terbentuk, variabel tekanan darah sistolik (TDS) sebesar 0,941 yang berarti sekitar 94,1% variansi variabel TDS dapat dijelaskan oleh komponen yang terbentuk, variabel tekanan darah diastolik (TDD) sebesar 0,936 yang berarti sekitar 93,6% variansi variabel TDD dapat dijelaskan oleh komponen yang

terbentuk, variabel kolesterol sebesar 0,939 yang berarti 93,9% variansi variabel TDD dapat dijelaskan oleh komponen yang terbentuk, variabel asam urat (AU) sebesar 0,911 yang berarti sebesar 91,1% variansi variabel AU dapat dijelaskan oleh komponen yang terbentuk, variabel hemoglobin sebesar 0,974 yang berarti sebesar 97,4% variansi variabel dapat dijelaskan oleh komponen yang terbentuk, dan variabel denyut nadi (DN) sebesar 0,915 yang berarti sebesar 91,5% variansi variabel DN dapat dijelaskan oleh komponen yang terbentuk.

Tabel 4. Nilai Eigen Berdasarkan Analisis Komponen Utama

Total komponen	Nilai Eigen	Keragaman Total (%)	Keragaman Kumulatif (%)
1	7,456	93,199	93,199
2	0,240	3,002	96,201
3	0,128	1,604	97,805
4	0,062	0,774	98,579
5	0,044	0,549	99,128
6	0,031	0,391	99,519
7	0,027	0,343	99,862
8	0,011	0,138	100,000

Dari tabel 4 memperlihatkan hanya terdapat satu komponen utama, komponen 1 memiliki nilai eigen sebesar 7,456 artinya

komponen 1 ini dapat menjelaskan variansi dari variabel bebas sebesar 93,199%.

Tabel 5. Komponen Matriks

Prediktor	Usia	BB	TDS	TDD	Kolesterol	AU	HB	DN
Komponen I	0,128	0,130	0,130	0,130	0,130	0,128	0,132	0,128

Berdasarkan hasil yang diperoleh pada Tabel 5, maka kombinasi linier dari komponen utama yang terbentuk adalah sebagai berikut:

$$K_1 = 0,128x_1 + 0,130x_2 + 0,130x_3 + 0,130x_4 + 0,130x_5 + 0,128x_6 + 0,132x_7 + 0,128x_8$$

Regresi Komponen Utama

Hubungan antara variabel – variabel bebas dengan variabel terikat dapat diketahui dengan melakukan analisis *Principal Component Regression (PCR)*. Dalam hal ini variabel terikat (kadar gula darah) diregresikan dengan komponen utama yang terbentuk. Dengan demikian,

model regresi komponen utama yang dibangun untuk kasus ini adalah:

$$\hat{Y} = w_0 + w_1K_1 + \varepsilon$$

Uji Asumsi Klasik

Sebelum melakukan analisis regresi komponen utama perlu dilakukan pengujian terhadap asumsi-asumsi klasik sebagai berikut:

1. Multikolinearitas

Terdapat 1 komponen utama yang terbentuk sehingga tidak mungkin terjadi multikolinearitas. Hal ini dapat dilihat pada tabel berikut:

Tabel 6. Hasil Statistik Kolinieritas

Model	t	Sig.	Kolinieritas	
			TOL	VIF
(constant)	77,659	0,000		
K1	28,514	0,000	1,000	1,000

Dari Tabel 6 menyatakan nilai $VIF = TOL = 1$ yang berarti tidak terjadi multikolinieritas.

2. Heteroskedastisitas
Model regresi yang baik tidak terjadi heteroskedastisitas, untuk mendeteksi heteroskedastisitas digunakan Uji Glejser yang hasilnya sebagai berikut:

Tabel 7. Hasil Uji Glejser

Prediktor	S.E	t	Sig.
K_1	2,671	0,410	0,684

Dari tabel 7, nilai signifikan $0,684 > \alpha = 0,05$ sehingga disimpulkan tidak terjadi heteroskedastisitas dalam model regresi.

3. Normalitas

Tabel 8. One – sample kolmogorov

Kolmogorov – Smirnov Z	0,824
Sig.	0,506

Dari hasil Tabel 8, nilai sig = $0,506 > \alpha = 0,05$. Hal ini berarti data terdistribusi normal.

4. Autokorelasi
Pendektesian autokorelasi menggunakan *Run test*. Jika antar residu tidak terjadi hubungan korelasi maka dikatakan residu adalah acak atau tidak terjadi autokorelasi.

Tabel 9. Run Test

Nilai test	1,38024
Sig.	1,000

Dari Tabel 9, dapat dilihat nilai signifikan = $1,000 > \alpha = 0,05$ sehingga dapat disimpulkan tidak terjadi autokorelasi

dalam model regresi. Karena semua asumsi regresi telah terpenuhi maka dapat

dilanjutkan ke analisis regresi komponen Koefisien Determinasi (R^2) utama.

Tabel 10. Koefisien Determinasi Hasil Regresi

R^2	Adjusted R^2	S.E
0,955	0,954	24,031

Dari Tabel 10, dapat dilihat nilai R^2 sebesar 0,955 ini berarti 95,5% variasi kadar gula darah dapat dijelaskan oleh variabel – variabel bebas yang telah ditentukan dalam hal ini komponen 1,

sedangkan 0,045 atau 4,5% dijelaskan oleh faktor lainnya. Sehingga diperoleh model regresi komponen utama sebagai berikut:

$$\hat{Y} = 295,075 + 109,722 K_1$$

Tabel 11. Uji Signifikansi Parameter

	β_i	t	Sig.
Konstanta	295,075	77,659	0,000
Komponen 1	109,722	28,514	0,000

Dari Tabel 11, nilai sig < $\alpha = 0,05$ maka dapat dikatakan komponen 1 berpengaruh signifikan terhadap kadar gula darah.

Pendugaan terhadap parameter koefisien regresi dari variabel asli (X) dapat menggunakan hubungan yang ada di antara parameter model regresi komponen utama (w) dan parameter regresi baku. Pendugaan parameter b dilakukan dengan jalan mensubstitusikan komponen utama K_1 ke dalam persamaan regresi utama. Persamaannya adalah sebagai berikut:

$$\begin{aligned} \hat{Y} &= 295,075 + 109,722K_1 \\ &= 295,075 + 109,722(0,128x_1 \\ &\quad + 0,130x_2 \\ &\quad + 0,130x_3 \\ &\quad + 0,130x_4 \\ &\quad + 0,130x_5 \\ &\quad + 0,128x_6 + 0,132x_7 \\ &\quad + 0,128x_8) \end{aligned}$$

Sehingga persamaan regresinya adalah sebagai berikut:

$$\begin{aligned} \hat{Y} &= 295,075 + 14,044x_1 \\ &\quad + 14,264x_2 \\ &\quad + 14,264x_3 \\ &\quad + 14,264x_4 \\ &\quad + 14,264x_5 \\ &\quad + 14,044x_6 + 14,483x_7 \\ &\quad + 14,044x_8 \end{aligned}$$

Interpretasi model data penderita penyakit diabetes mellitus berdasarkan faktor – faktor yang mempengaruhinya dengan menggunakan analisis regresi komponen utama sebagai berikut:

Semakin bertambahnya usia tiap tahunnya maka semakin besar kandungan kadar gula darah karena setiap kenaikan 1 tahun usia mengakibatkan kadar gula bertambah sebesar 14,044 mg/dl.

Semakin bertambahnya 1 kg berat badan maka semakin besar kandungan kadar gula darah karena setiap kenaikan 1 kg berat

badan mengakibatkan kadar gula darah bertambah sebesar 14,264 mg/dl

Semakin tingginya tekanan darah sistolik maka semakin tinggi kandungan kadar gula darah karena setiap kenaikan 1 mmHg tekanan darah sistolik mengakibatkan kadar gula darah bertambah sebesar 14,264 mg/dl.

Semakin tingginya tekanan darah diastolik maka semakin tinggi kandungan kadar gula darah karena setiap kenaikan 1 mmHg tekanan darah diastolik mengakibatkan kadar gula darah bertambah sebesar 14,264 mg/dl.

Semakin tingginya kandungan kolesterol maka semakin tinggi kandungan kadar gula darah karena setiap kenaikan 1 mg/dl kolesterol mengakibatkan kadar gula darah bertambah sebesar 14,264 mg/dl.

Semakin tingginya tingkat asam urat maka semakin tinggi kandungan kadar gula darah karena setiap kenaikan 1mg/dl asam urat mengakibatkan kadar gula darah bertambah sebesar 14,044 mg/dl.

Semakin tinggi kandungan hemoglobin maka semakin tinggi kandungan kadar gula darah karena setiap kenaikan 1 mg/dl hemoglobin mengakibatkan kadar gula darah bertambah sebesar 14,483 mg/dl.

Semakin cepat denyut nadi maka semakin besar kandungan kadar gula darah karena setiap kenaikan 1 satuan denyut nadi mengakibatkan kadar gula darah bertambah sebesar 14,044mg/dl.

KESIMPULAN DAN SARAN

Dari data penderita diabetes mellitus terdapat kasus multikolinieritas sehingga dengan menggunakan analisis regresi komponen utama diperoleh model terbaik untuk data tersebut adalah:

$$\begin{aligned} \hat{Y} = & 295,075 + 14,044x_1 \\ & + 14,264x_2 \\ & + 14,264x_3 \\ & + 14,264x_4 \\ & + 14,264x_5 + 14,044x_6 + 14,483x_7 \\ & + 14,044x_8 \end{aligned}$$

Model regresi komponen utama memenuhi semua pengujian koefisien regresi secara serentak, pengujian koefisien regresi secara individual dan pengujian asumsi klasik.

Berdasarkan model persamaan regresi diperoleh koefisien terbesar yaitu pada x_7 (hemoglobin), yang berarti bahwa faktor dominan penyebab diabetes mellitus adalah Hemoglobin.

Bagi penderita diabetes mellitus agar menghindari makanan yang mengakibatkan peningkatan hemoglobin dalam tubuh karena hemoglobin merupakan faktor dominan penyebab Diabetes Mellitus.

Bagi masyarakat diharapkan lebih mengutamakan pola hidup sehat agar dapat terhindar dari penyakit.

DAFTAR PUSTAKA

- Djalal, N, *et al.* 2002. Penggunaan Teknik Ekonometrika. Edisi Revisi. Jakarta: PT. Raja Grafindo Persada
- Drapper, N.R dan Smith, H. 1992. *Analisis Regresi Terapan*, edisi kedua. Alih bahasa: sumantri b. Jakarta : gramedia
- Farhan, M.Q. 2012. *Analisis Regresi Terapan*, Penerbit Andi, Yogyakarta
- Ghozali, Imam, 2005, *Aplikasi Analisis Multivariat dengan program SPSS*, Badan Penerbit Universitas Diponegoro, Semarang
- Imam, G. (2013). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 21 Update PLS Regresi Edisi 7*. Semarang: UNDIP.

- Johnson, R, A. & Wichem, D, W. 2002. *Applied Multivariate Statistical Analysis. 5th edition.* Pearson education International.
- Rendi, A, P. 2013. *Hidup Manis dengan Diabetes.* Yogyakarta: PT. Buku Seru
- Sudjana. 2004. *Pengantar Statistika untuk penelitian.* Bandung : CV.Alfabet
- Sunarto, H. & Riduwan. 2010. *Pengantar Statistika untuk Penelitian Pendidikan, Sosial, Ekonomi, Komunikasi, dan Bisnis.* Bandung : CV.Alfabet
- Supranto, J. 2008. *Statistik Teori dan Aplikasi,* edisi ketujuh. Jakarta : PT. Gelore Aksara Pratama
- Sembiring, R.K. 1995. *Analisis Regresi.* Bandung: ITB Bandung
- Widharto. 2007. *Kencing Manis (DIABETES).* Jakarta: Sunda Kelapa Pustaka.