

PENERAPAN ANALISIS REGRESI RIDGE PADA DATA PASIEN HIPERTENSI DI RUMAH SAKIT UMUM DAERAH SIDIKALANG TAHUN 2014

Medis Pasaribu¹, Drs. Abdul Jalil A.A., M.Kes², dr. Ria Masniari Lubis, M.Si²

1. Mahasiswa FKM USU peminatan Biostatistika dan Informasi Kesehatan

2. Staf Pengajar FKM USU

ABSTRACT

Multicollinearity is the correlation between some independent variables which makes coefficient regression yielded by multiple regression analysis cannot be used in estimating. The problem of multicollinearity which is indicated by the value of Variance Inflation Factor (VIF) > 10 occurs when the influence of age, obesity (IMT), and cholesterol were modeled toward systole and diastole blood pressure.

The objective of the research was to obtain the regression model of age, obesity (IMT), and cholesterol with systole blood pressure and with diastole blood pressure of patients suffered from hypertension at the Inpatient Wards of the RSUD (Regional General Hospital) Sidikalang, in 2014. The samples consisted of 105 patients. The data were secondary data and using $\alpha=0,1$. Analyzed by using ridge regression in order to obtain correct regression model.

The result of ANOVA test showed that $p < 0.000$ which indicated that, simultaneously, there was the influence of age, obesity (IMT), and cholesterol on systole and diastole blood pressure. The equation for modeling blood pressure was obtained at $k = 0.07$; they were $Y_{\text{systole}} = 21.2828 + 0.7169 \text{ Age} + 1.7986 \text{ IMT} + 0.2222 \text{ Cholesterol}$, and $Y_{\text{diastole}} = 24.5290 + 0.5025 \text{ Age} + 0.6648 \text{ IMT} + 0.0946 \text{ Cholesterol}$.

It is recommended that the next researchers solve the problems of multicollinearity during the modeling; one of them is by using ridge regression analysis.

Keywords: Ridge Regression, Multicollinearity, Hypertension

PENDAHULUAN

Analisis regresi merupakan salah satu metode statistik yang sering digunakan dalam kehidupan sehari-hari. Tujuan analisis regresi yaitu mengetahui sejauh mana hubungan sebuah variable bebas dengan beberapa variable tak bebas. Bila dalam analisisnya hanya melibatkan sebuah variable bebas, maka analisis yang digunakan adalah analisis regresi sederhana. Sedangkan, apabila analisisnya melibatkan lebih dari satu variable bebas, maka digunakan analisis regresi linier berganda.

Analisis model regresi dinyatakan baik jika memenuhi asumsi-asumsi klasik antara lain yaitu tidak ada autokorelasi, tidak ada

heteroskedastisitas, dan tidak ada multikolinier. Saat menentukan model regresi populasi ada kemungkinan beberapa atau semua variable bebas (variable X) membentuk hubungan antara satu sama lain, kejadian ini dapat disebut juga sebagai multikolinier (Ryan, 1997).

Sembiring (1995) berpendapat bahwa salah satu akibat dari adanya multikolinearitas adalah variansi estimator β sangat besar mendekati tak hingga, bahkan untuk keadaan multikolinearitas sempurna tidak lagi sangat besar melainkan tak hingga. Jika variansi estimator suatu model sangat besar, menandakan perkiraan koefisien regresi sangat lemah.

Banyak metode yang untuk mengatasi multikolinearitas, salah satunya regresi ridge. Metode ini merupakan modifikasi dari metode kuadrat terkecil dengan cara menambahkan tetapan bias k yang kecil pada diagonal matriks $X^T X$. Walaupun memiliki penduga yang bias, penduga ini memiliki variansi estimator koefisien minimum (Ryan, 1977).

Pada penelitian ini, masalah kesehatan yang digunakan untuk mengaplikasikan penggunaan metode regresi ridge adalah penyakit hipertensi karena penyakit hipertensi memiliki faktor resiko yang digunakan sebagai variable bebas yang diindikasikan melanggar asumsi multikolinearitas. Selain dari pada itu, hipertensi merupakan penyakit tidak menular yang banyak mengakibatkan kematian. Dari 17 juta kematian akibat penyakit kardiovaskuler terdapat 9,4 juta kematian akibat hipertensi setiap tahun. Sebanyak 45% dari kematian akibat penyakit jantung disebabkan oleh hipertensi dan 51% kematian akibat stroke disebabkan oleh hipertensi (WHO, 2013).

Menurut Riskesdas Tahun 2007, prevalensi hipertensi yang didapat pada pengukuran mulai dari umur ≥ 18 tahun sebesar 31,7 % dan menurut Riskesdas 2013 menurun menjadi 25,8 %. Dan prevalensi hipertensi di Sumatera Utara sebesar 24,7 %.

Data yang diperoleh dari Sistem Informasi RumahSakit (SIRS) Tahun 2010-2011 menyebutkan bahwa penyakit hipertensi selalu masuk dalam peringkat 10 besar penyakit tidak menular yang menyebabkan rawat jalan dan rawat inap di seluruh Rumah Sakit di Indonesia tahun 2009-2010.

Berdasarkan data dari Dinas Kesehatan Dairi, hipertensi menjadi penyakit kedua terbesar diderita masyarakat Kabupaten Dairi pada tahun 2014 yaitu sebesar 9223 penderita. (Dinkes Dairi, 2014). Berdasarkan survey pendahuluan yang dilakukan di RSUD Sidikalang, hipertensi masuk dalam 10 penyakit terbesar untuk rawat inap dan rawat jalan. Untuk rawat jalan, hipertensi adalah penyakit kedua terbesar yaitu sebanyak 3148, Untuk rawat inap, hipertensi menjadi penyakit terbesar pertama yaitu sebanyak 1115 (RSUD Sidikalang, 2014).

Berdasarkan uraian diatas maka akan dilakukan penelitian untuk mencari model regresi dari tekanan darah pada pasien hipertensi yang berobat di RSUD Sidikalang tahun 2014 dengan menggunakan regresi ridge yang memodelkan hubungan antara umur, obesitas (yang diukur dengan IMT), dan kadar kolesterol dengan tekanan sistolik dan diastolik.

Yang menjadi rumusan masalah dalam penelitian ini adalah terjadinya multikolinearitas antara variabel-variabel yang mempengaruhi hipertensi sehingga regresi linear ganda tidak dapat digunakan untuk mendapatkan model regresi dari tekanan darah.

Tujuan umum dari penelitian ini adalah untuk mengaplikasikan analisis regresi ridge dalam mengatasi masalah multikolinearitas yang terjadi pada variabel - variabel yang mempengaruhi hipertensi pada pasien RSUD Sidikalang tahun 2014.

Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian deskriptif analitik. Sampel penelitian ada sebanyak 105 pasien dan cara penarikan sampel

yang digunakan dalam penelitian ini adalah dengan mengambil rekam medis pasien hipertensi dari bulan desember hingga bulan sebelumnya sampai terpenuhi sebanyak 105 data pasien. Data yang digunakan adalah data sekunder.

Hasil dan Pembahasan Mendeteksi Multikolinearitas

Untuk pendeteksian multikolinearitas digunakan dengan pendekatan nilai VIF yang dihasilkan dengan analisis regresi linier ganda. Berikut adalah nilai VIF pada tiap variable bebas:

Tabel 1. Nilai VIF variabel bebas

Variabel	VIF
UmurPasien	30,276
Indeksmassatubuh	6,225
Kadar kolesterol	39,258

Dari tabel diatas dapat dilihat bahwa terdapat dua variable bebas yang memiliki angka VIF lebih besar dari 10. Oleh karena itu dapat disimpulkan bahwa terdapat multikolinearitas pada data.

Menentukan Koefisien Regresi dengan Regresi Ridge

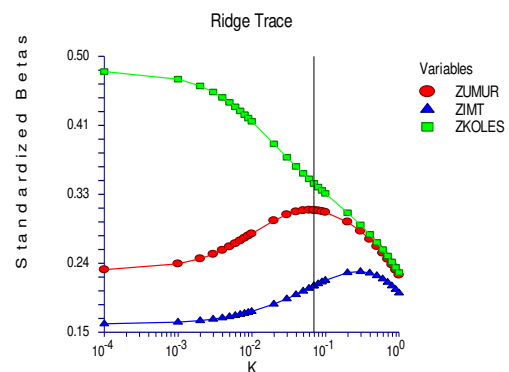
Sebelum dianalisis dengan regresi ridge, dilakukan pemusatan dan penskalaan pada data sehingga diperoleh data Y_i^* dan X_{ij}^* . Dalam

penentuan koefisien regresi, penetapan tetapan bias k merupakan hal yang sangat penting dalam penelitian ini. Dalam analisis regresi ridge akan ditampilkan nilai koefisien estimator $\hat{\beta}(k)$ parameter dengan berbagai tetapan nilai k.

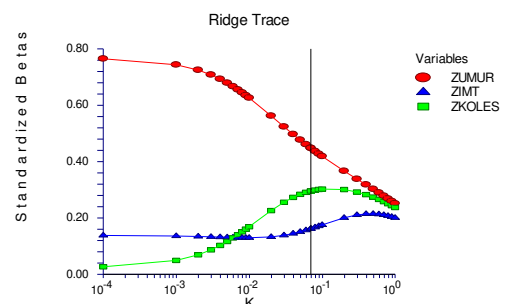
Pada $k=0$ maka hasil regresi yang terbentuk akan sama dengan persamaan regresi yang dihasilkan dengan analisis regresi linier ganda

yaitu $Y_{sistol} = 18,3978 + 0,5397 \text{ Umur} + 1,3834 \text{ IMT} + 0,3153 \text{ Kolesterol}$. Dan $Y_{diastol} = 28,9583 + 0,8582 \text{ Umur} + 0,5664 \text{ IMT} + 0,0083 \text{ Kolesterol}$.

Pada penelitian ini, pemilihan tetapan bias k dengan pendekatan nilai VIF dan ridge trace. Pada k 0,02 sampai 1, nilai VIF $\hat{\beta}(K)$ kurang dari 10. Maka k yang terpilih adalah k yang menampilkan koefisien $\hat{\beta}$ lebih stabil. Berikut adalah ridge trace dari tekanan darah sistol.



Gambar 1. Ridge Trace dari Tekanan Darah Sistol dengan Berbagai Nilai K



Gambar 2. Ridge Trace dari Tekanan Darah Diastol dengan Berbagai k

Dari ridge trace diatas untuk $K = 0,07$ koefisien $\hat{\beta}$ lebih stabil sehingga nilai K yang terpilih adalah 0,07.

Sehingga koefisien regresi ridge untuk tekanan darah sistol yang diperoleh adalah $Y_{sistol}^* = 0,3047Umur^* + 0,2089 IMT^* + 0,3389Kolesterol^*$. Dan koefisien regresi ridge untuk tekanan darah diastol yang diperoleh adalah $Y_{diastol}^* = 0,4479Umur^* + 0,1619 IMT^* + 0,2957 Kolesterol^*$. Proses pengembalian Y_{sistol}^* kebentuk Y_{sistol} dan $Y_{diastol}^*$ kebentuk $Y_{diastol}$ dengan rumus $\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \dots - \beta_k\bar{X}_k$ dengan $\bar{Y}_{sistol} = 170,57$; $\bar{Y}_{diastol} = 96,10$; $\bar{X}_{umur} = 57,38$; $\bar{X}_{IMT} = 27,71$; $\bar{X}_{kolesterol} = 262,82$; $S_{sistol} = 22,989$;

$S_{diastol} = 10,963$; $S_{umur} = 9,771$; $S_{IMT} = 2,670$ dan $S_{kolesterol} = 35,055$.

Sehingga persamaan regresi ridge yang diperoleh adalah $Y_{sistol} = 21,2828 + 0,7169 Umur + 1,7986 IMT + 0,2222 Kolesterol$ dan $Y_{diastol} = 24,5290 + 0,5025 Umur + 0,6648 IMT + 0,0946 Kolesterol$.

Uji Keberartian Regresi

Kemudian akan diuji keberartian dari model yang diperoleh dengan menggunakan tabel ANOVA.

$H_0: \beta_0 = \beta_1 = \dots = \beta_j = 0$ (regresi tidak

berarti)

$H_1: \beta_j \neq 0$ (regresi berarti)

Tabel 2. Tabel ANOVA Ridge Tekanan Darah Sistol

Varian	DK	JK	RK	F HITUNG	P
Sisa	1	3,9744	3,9744	84,1692	0,000
Regresi	3	0,7068	0,2356		
Error	101	0,2331	2,9024		
Total	105	1,0001	9,615		

Dari tabel diatas tampak bahwa nilai $P=0,000$ lebih kecil dari α (0.1). Artinya H_0 ditolak atau dapat dinyatakan bahwa regresi yang menghubungkan umur, obesitas(IMT)

dan kolesterol terhadap tekanan darah sistol berarti. Dari tabel ANOVA tersebut juga dapat di ambil kesimpulan bahwa terdapat hubungan yang linier antara variable independen dengan variable dependen.

Tabel 3 Tabel ANOVA Ridge Tekanan Darah Diastol

Varian	DK	JK	RK	F HITUNG	P
Sisa	1	1,9157	1,9157	146,16	0,000
Regresi	3	0,8036	0,2678		
Error	101	0,1968	0,9446		
Total	105	0,9999	9,6150		

Dari tabel diatas tampak bahwa nilai $P=0,000$ lebih kecil dari α (0.1). Artinya H_0 ditolak atau dapat dinyatakan regresi yang

menghubungkan umur, obesitas (IMT), dan kolesterol berarti. Dari tabel ANOVA tersebut juga dapat di ambil kesimpulan bahwa terdapat hubungan

yang linier antara variable independen dengan variable dependen.

KESIMPULAN DAN SARAN

Kesimpulan

1. Data mengalami multikolinearitas terlihat pada hasil uji multikolinearitas yaitu nilai VIF > 10
2. Untuk mengatasi masalah multikolinearitas yang terdapat pada data digunakan analisis regresi ridge sehingga model regresi yang di dapat adalah $Y_{sistol} = 21,2828 + 0,7169 \text{ Umur} + 1,7986 \text{ IMT} + 0,2222 \text{ Kolesterol}$. dan $Y_{diastol} = 24,5290 + 0,5025 \text{ Umur} + 0,6648 \text{ IMT} + 0,0946 \text{ Kolesterol}$

Saran

1. Multikolinearitas dalam data dapat menyebabkan model regresi kurang baik dalam peramalan sehingga disarankan kepada pembaca untuk terlebih dahulu mengatasi masalah multikolinearitas salah satunya dengan uji regresi ridge.

DAFTAR PUSTAKA

Riset Kesehatan Dasar. 2013. **Laporan Hasil Riset Kesehatan Dasar (Riskesdas) 2013**. Jakarta: Badan Libankes, Depkes RI, 2013

RSUDS. 2015. **Profil Rumah Sakit Umum Daerah Sidikalang Tahun 2015**. Rumah Sakit Umum Daerah Sidikalang

Ryan, T.P. 1997. **Modern Regression Method**. New York: Wiley

Sembiring, R.K. 1995. **Analisis Regresi Terapan**. Bandung: ITB Bandung

SIRS. 2011. **Data dan Inforamasi Kesehatan Penyakit Tidak Menular**. Jakarta: Kemenkes RI.

Soemartini. (2008). **Principal Component Analysis(PCA) sebagai Salah Satu Metode untuk Mengatasi Masalah Multikolinieritas**. Skripsi S1. Jatinangor :Jurusan Statistika, FMIPA Universitas Padjadjaran.

Supranto. 1995. **Ekonometrik**. Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia.

WHO. 2013. **A Global Brief on Hypertension Silent Killer, Global Public Health Crisis**. Geneva: World Health Organization