Jurnal Sains, Teknologi dan Industri, Vol. 12, No. 1, Desember 2014, pp. 48 - 57
ISSN 1693-2390 print/ISSN 2407-0939 online

ANALISIS JUMLAH TENAGA KERJA TERHADAP JUMLAH PASIEN RSUD ARIFIN ACHMAD PEKANBARU MENGGUNAKAN METODE REGRESI GULUD

[1]Rahmadeni, [2]Defi Anggreni
[1,2]Jurusan Matematika Fakultas Sains dan Teknologi UIN Suska Riau
E-mail: r4dieni@gmail.com

ABSTRAK

Penelitian ini menjelaskan tentang analisis jumlah tenaga kerja terhadap jumlah pasien RSUD Pekanbaru pada Tahun 2012 yang menjelaskan masalah multikolenieritas. Penyelesaian masalah multikolinearitas menggunakan metode regresi gulud *(ridge regression)* dilakukan dengan mentransformasi masing-masing peubah X dan Y melalui prosedur pemusatan dan pengskalaan. Suatu acuan yang digunakan untuk memilih besarnya nilai Ridge Parameter θ, dengan melihat besarnya nilai VIF (*Varian Inflation Factor*). Nilai VIF lebih besar dari 10 mengindetifikasi adanya multikolenier. Untuk memperjelas penggunaan regresi gulud untuk mengatasi multikolinearitas dibahas contoh kasus multikolinearitas, yaitu hubungan antara jumlah tenaga kerja (Y) dan (X) jumlah pasien berdasarkan cara bayar. Dari pembahasan contoh studi kasus, diperoleh persamaan regresi gulud:

$$\hat{Y} = 6{,}8x_1 + 0{,}018x_2 + 0{,}015x_3 - 0.02x_4 + 0.02x_5 - 0{,}012x_6 + 0{,}009x_7 - 0{,}0036x_8$$

Katakunci: *metode kuadrat terkecil, multikolenieritas, regresi gulud, regresi linier berganda, dan ridge parameter*

ABSTRACT

This research describes the analysis of the amount of labor against toward number of patients in the Hospital area Arifin Achmad of Pekanbaru in year 2012 which is describe the problem multicollinearity. Problem solved multicollinearity of ridge regression method was done by transform each X and Y variables through of the centering and rescaling. A reference that is used to select higher than of the value of the parameter θ, with a ridge to see choosing of the value of the VIF (variant of the Inflation Factor) that if there is a correlation between the free variables, the value of the VIF would be higher than (VIF > 10). VIF values higher than 10 identification of multicollinearity. To clarify of using of ridge regression method to resolved the case discussed multicollinearity which relationship between the amount of labor (Y) and (X) number of patients based on how to pay. From discussion of example case studies, the regression equation obtained gulud:

$$\hat{Y} = 6{,}8x_1 + 0{,}018x_2 + 0{,}015x_3 - 0.02x_4 + 0.02x_5 - 0{,}012x_6 + 0{,}009x_7 - 0{,}0036x_8$$

Keywords: *least squares method, multikolenieritas, multiple linear regression, and the ridge parameter, and ridge regression*.

PENDAHULUAN

Analisis regresi merupakan suatu cara yang dapat digunakan untuk mengetahui hubungan sebuah variabel tak bebas (*regressand*) dengan sebuah atau lebih variabel bebas (*regressor*). Menurut Drapper and Smith (1992) analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel terhadap variabel lainnya. Bila dalam analisisnya hanya melibatkan sebuah variabel bebas, maka analisis yang digunakan adalah analisis regresi linier sederhana. Hubungan atau korelasi antara dua variabel melalui persamaan regresi sederhana untuk meramalkan nilai Y dengan X yang sudah diketahui nilainya tidak cukup, sebab selain X masih ada variabel lainnya. Apabila dalam persamaan analisis regresi melibatkan dua atau lebih variabel bebas, maka regresi ini disebut analisis regresi linier berganda *(multiple linier regression)*.

Analisis regresi linier berganda mempunyai lebih dari satu variabel bebas, sering menimbulkan masalah karena terjadinya hubungan kuat antara dua variabel bebasnya yang mengakibatkan terjadinya kolenieritas ganda (*multikolenierity*). Gejala ini menimbulkan masalah dalam pemodelan regresi. Kolerasi yang sangat tinggi akan menghasilkan penaksiran yang berbias, tidak stabil dan mungkin jauh dari nilai sasaran Gonst dan Mason (1997) sehingga galat yang dihasilkan menjadi besar dan variansi parameternya menjadi tak hingga. Metode kuadrat terkecil akan memberikan efek dari kolenieritas yaitu tingginya nilai koefisien determinasi tetapi tidak diikuti dengan hasil uji hipotesis yang signifikan.

Salah satu cara untuk mendapatkan koefisien regresi pada persamaan regresi linier berganda adalah melalui metode kuadrat terkecil. Metode ini menghasilkan penaksiran terbaik (tak bias dan variansi minimum) apabila tidak ada kolerasi antara variabel bebas. Tetapi jika hal itu terjadi, maka salah satu cara untuk mengatasi masalah tersebut adalah melalui metode regresi gulud (*Ridge Regression*). Dengan menggunakan metode ini dapat mengurangi dampak terjadinya multikolenieritas dengan menentukan pendugaan yang bias tetapi mempunyai varians yang lebih kecil dari varians penduga regresi linier berganda. Pada dasarnya metode ini merupakan modifikasi dari metode kuadrat terkecil dengan cara menambahkan tetapan bias k pada diagonal matriks $X'X$, dimana $0 < k < 1$. Dengan asumsi matriks korelasi dari variabel bebas dapat diinverskan dengan menggunakan metode regresi gulud sehingga nilai dugaan koefisien regresi mudah didapat.

Salah satu aplikasi pemakaian metode regresi gulud yang dapat dianalisis yaitu jumlah tenaga kerja pada RSUD Arifin Achmad Pekanbaru yang sangat berpengaruh terhadap kunjungan pasien ke rumah sakit karena pihak rumah sakit harus dapat memberikan pelayanan kesehatan dengan baik kepada masyarakat yang membutuhkan pelayanan kesehatan. Adapun tenaga kerja yang dimaksud disini adalah dokter atau paramedis dan tenaga kerja nonparamedis yang ditempatkan disetiap poli pada RSUD Arifin Achmad Pekanbaru. Dalam menaksirkan atau memprediksi tenaga kerja yang diperlukan untuk rumah sakit mengalami masalah atau gangguan karena disebabkan faktor-faktor yang mempengaruhi atau yang menjadi variabel bebasnya mengalami korelasi ganda atau hubungan diantara variabel–variable bebasnya mengalami masalah multikolinieritas. Oleh karena itu tujuan dalam penelitian ini adalah untuk menyelesaikan masalah multikolenieritas pada studi kasus data jumlah tenaga kerja terhadap jumlah pasien di RSUD Arifin Achmad Pekanbaru dengan menggunakan metode regresi gulud.

Tinjauan Pustaka

Definisi Tenaga Kerja

Tenaga kerja merupakan penduduk yang berada dalam usia kerja. Menurut UU No.13 tahun 2003 Bab 1 pasal 1 ayat 2 disebutkan bahwa tenaga kerja adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan barang atau jasa baik untuk memenuhi kebutuhan sendiri maupun untuk masyarakat.

Tenaga medis adalah tenaga ahli kedokteran dengan fungsi utamanya adalah memberikan pelayanan medis kepada pasien dengan mutu sebaik-baiknya dengan menggunakan tata cara dan teknik berdasarkan ilmu kedokteran dan etik yang berlaku serta dapat dipertanggungjawabkan Anireon (1984).

Jumlah atau tenaga kerja yang dimaksud pada studi kasus ini adalah dokter atau paramedis dan non paramedis yang bekerja disetiap poli di RSUD Arifin Achmad.

Analisis Regresi Linier Berganda

Analisis regresi merupakan salah satu analisis statistik yang sering digunakan untuk menganalisis hubungan antara dua variabel atau lebih. Menurut Drapper dan Smith (1992) analisis regresi merupakan metode analisis yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna tentang hubungan ketergantungan variabel terhadap variabel lainnya. Hubungan yang didapat pada umumnya dinyatakan dalam bentuk persamaan matematika yang menyatakan hubungan antara variabel bebas (*idependent variable*) x dan variabel tak *bebas (dependent variable)* y dalam bentuk persamaan sederhana.

$$y = \beta_0 + \beta_1 X_1 \qquad i = 1,2,\dots,n$$

Regresi linier berganda merupakan perluasan dari regresi linier sederhana.

Perluasan terlihat dari banyaknya variabel bebas pada model regresi tersebut. Bentuk umum regresi linier berganda dapat dinyatakan secara statistik sebagai berikut:

$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$

keterangan:

Y_i = variabel tak bebas
X_{ki} = variabel bebas
$\beta_1, \cdots \beta_k$ = parameter regresi
ε_i = variabel gangguan

Asumsi Regresi Linier Berganda

Dalam metode regresi linier berganda ada beberapa asumsi yang harus dipenuhi, asumsi tersebut adalah:

1. Nilai rata-rata kesalahan pengganggu nol, yaitu $E(\varepsilon_i) = 0$, untuk i= 1,2,...,n
2. Varian $(\varepsilon_i) = E(\varepsilon_i^2) = \sigma^2$
3. Tidak ada autokorelasi antara kesalahan pengganggu (galat/*error*), berarti kovarian $(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$
4. Variabel bebas $x_1, x_2, \ldots, x_k$, konstan dalam sampling yang terulang dan bebas terhadap kesalahan penganggu ε_i.
5. Tidak ada multikolenieritas diantara variabel bebas x
6. $\varepsilon_i \sim N(0; \sigma^2)$, artinya kesalahan pengganggu mengikuti distribusi normal dengan rata-rata 0 dan varian σ^2.

Metode Kuadrat Terkecil

Salah satu metode penduga parameter dalam model regresi adalah metode kuadrat terkecil. Metode ini memerlukan beberapa asumsi klasik yang harus dipenuhi oleh komponen ε, yaitu memenuhi asumsi kenormalan, kehomogenan ragam, dan tidak memiliki autokorelasi.

Metode kuadrat terkecil merupakan suatu metode yang digunakan untuk menaksir parameter regresi dengan cara meminimumkan jumlah kuadrat kekeliruan (*error*) dari model regresi yang terbentuk.

Jumlah kuadrat kekeliruan *(error)* untuk persamaan regresi linier sederhana, yaitu:

$$J = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \; ; i = 1,2,\cdots,n$$

Sehingga diperoleh pendugaan kuadrat terkecil dari β pada regresi linier berganda adalah, sebagai berikut:

$$\hat{\beta} = (X'X)^{-1} X'Y$$

Persamaan regresi linear berganda dengan dua variabel bebas mempunyai perhitungan nilai a, b_1, dan b_2 . Menggunakan metode kuadrat terkecil yaitu dengan meminimumkan nilai $G = \sum e^2$ atau $G = \sum (Y - a - b_1 X_1 - b_2 X_2)^2$.

Dengan menyamakan fungsi-fungsi turunan pertama parsial dari jumlah G terhadap setiap nilai a, b_1, dan b_2, sehingga diperoleh :

$$b_1 = \frac{\sum x_1 y \sum x_2^2 - \sum x_2 y \sum x_1 x_2}{(\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2)}$$

$$b_2 = \frac{\sum x_2 y \sum x_1^2 - \sum x_1 y \sum x_1 x_2}{\sum x_2^2 \sum x_1^2 - (\sum x_1 x_2)^2}$$

Multikolinieritas

Multikolinieritas adalah suatu kondisi dimana terjadi korelasi yang kuat diantara variabel-variabel bebas (X) yang diikutsertakan dalam pembentukan model regresi linier. Jelas bahwa multikolinieritas adalah suatu kondisi yang menyalahi asumsi regresi linier. Tentu saja, multikolinieritas tidak mungkin terjadi apabila variabel bebas (X) yang diikut sertakan hanya satu.

Dalam bentuk matriks, multikolinearitas adalah suatu kondisi buruk atau *ill condition* dari matriks $X'X$ yaitu suatu kondisi yang tidak memenuhi asumsi klasik. Jika multikolinearitas terjadi antara dua variabel atau lebih dalam suatu persamaan regresi, maka nilai perkiraan koefisien dari variabel yang bersangkutan menjadi tak berhingga, sehingga tidak mungkin lagi menduganya. Hal ini disebabkan $X'X$ menjadi singular atau $X'X$ mendekati nol.

Ada beberapa cara untuk mengetahui ada tidaknya multikolinieritas, yaitu:

a. Nilai Korelasi (Korelasi antar Peubah Bebas)

Prosedur ini merupakan pendeteksian yang paling sederhana dan paling mudah. Nilai korelasi yang tinggi antara peubah satu dengan yang lainnya memperlihatkan adanya hubungan linier pada peubah-peubah tersebut.

b. Nilai Kondisi

Ada beberapa metode untuk menghitung nilai kondisi (ϕ) yang menunjukkan tingkat multikolenieritas Vonod dan Ulah (1981) menyarankan bahwa nilai kondisi diberikan oleh:

$$\phi = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \tag{2.14}$$

Montgomery dan Peck (1992) mendefinisikan nilai kondisi merupakan perbandingan dari λ_{max} dan λ_{min} yang didapat dari matriks korelasi dan memberikan kategori multikolenieritas berdasarkan nilai kondisi yang diperoleh, adapun Persamaan perbandingan dari λ_{max} dan λ_{min} sebagai berikut :

$$\phi = \frac{\lambda_{max}}{\lambda_{min}} \tag{2.15}$$

dengan:

λ_{max} adalah nilai eigen yang terbesar (maksimum)

λ_{min} adalah nilai eigen yang terkecil (minimum)

jika:

$\phi < 100$ maka disebut multikolenieritas rendah

$100 \leq \phi < 1000$ maka disebut multikolenieritas cukup kuat

$\phi \geq 1000$ maka disebut multikolenieritas kuat

Nilai kondisi yang terlalu besar mengindifikasikan multikolenieritas yang serius. Nilai kondisi yang terlalu besar menunjukkan ketidakstabilan koefisien regresi terhadap perubahan dalam data variabel bebas.

Pagel dan Lunnebor (1985) menyatakan bahwa nilai kondisi adalah:

$$\phi = \sum_{i=1}^{p} \frac{1}{\lambda_i} \tag{2.16}$$

c. VIF (*Varians Inflantion Factors*)

VIF adalah elemen-elemen diagonal utama dari invers matriks korelasi. VIF digunakan sebagai kriteria untuk medeteksi multikolenieritas pada regresi linier berganda yang melibatkan lebih dari dua variabel bebas. Nilai VIF lebih besar dari 10 mengidentifikasikan adanya masalah mutikolenieritas yang serius. VIF untuk koefisien regresi ke-j didefinisikan sebagi berikut:

$$\mathrm{VIF_j} = \frac{1}{1-R_j^2} \tag{2.17}$$

dengan :

R_j^2 = koefisien deterninasi antar X_j dengan variabel bebas lainnya ; $j = 1,2,\cdots 3, n$

Uji *F*

Uji F dilakukan untuk melihat pengaruh variabel-variabel bebas secara keseluruhan terhadap variabel terikat. Pengujian ini dilakukan dengan membandingkan nilai F_{hitung} dengan F_{tabel} . Untuk uji statistik koefisien berganda, uji statistiknya menggunakan uji F ini dengan memakai rumus:

$$F_{hitung} = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

keterangan:

$R =$ koefisien korelasi berganda

$k =$ jumlah variabel bebas

$n =$ jumlah anggota sampel

Prosedur uji statistiknya adalah sebagai berikut:

a. Menentukan hipotesis

H_0 : $\beta_i = 0$ (tidak ada pengaruh secara signifikan antara variabel bebas secara simultan atau bersama-sama terhadap variabel terikat atau Y)

H_1 : $\beta_i \neq 0$ (ada pengaruh secara signifikan antara variabel bebas secara simultan atau bersama-sama terhadap variabel terikat atau Y)

b. Mentukan taraf nyata (α) dan nilai F tabel

Nilai taraf nyata yang digunakan 0.05 dan nilai F tabel memiliki $v_1 = k$ dan $v_2 = n - k - 1$.

c. Menentukan kriteria pengijuan dan memberikan kesimpulan

jika $F_{hitung} > F_{tabel}$ berarti H_0 di tolak, dan jika $F_{hitung} > F_{tabel}$ berarti H_0 diterima. Kemudian menghitung nilai F tabel dan membuat kesimpulan ketika F_{hitung} bandingkan dengan F_{tabel}.

Berdasarkan uji F dari F_{hitung} yang lebih kecil dari F_{tabel} berarti semua variabel bebas X berpengaruh secara signifikan terhadap nilai taksiran Y.

Ridge Regression (Regresi Gulud)

Menurut RE Walpole dan R.H Mayers pada Tahun 1985, dengan adanya multikolenieritas dapat menyebabkan pendugaan koefisien regresi sangat tidak stabil dan sensitif terhadap perubahan data. Selain itu dapat menyebabkan perbedaan koefisien untuk data sampel yang berbeda cenderung besar. Oleh sebab itu diperlukan suatu metode penaksiran alternatif yang memberi hasil penaksiran yang baik yang menghasilkan penduga koefisien regresi bias tetapi cenderung mempunyai ketepatan yang lebih baik.

Prosedur regresi gulud diusulkan pertama kali oleh A.E Hoerl pada Tahun 1962 dan dibahas secara mendalam dalam dua tulisan Hoerl dan Kennard. Prosedur tersebut ditujukan untuk mengatasi suatu multikolenieritas dan kolom matriks dari X tidak bebas linier yang menyebabkan matriks $X'X$ hampir singular.

Pada metode regresi gulud, penduga koefisien regresi yang dihasilkan adalah penduga bias. Penaksiran metode alternatif tidak sebaik metode kuadrat terkecil karena jumlah kuadrat residual tidak terlalu kecil dan koefisien korelasi ganda tidak terlalu besar tetapi lebih potensial untuk ketepatan yang lebih baik.

Regresi gulud merupakan salah satu metode yang dapat digunakan untuk mengatasi masalah multikolinearitas melalui modifikasi terhadap metode kuadrat terkecil Neter, Waserman dan Kutner (1990) dalam Herwindiati (1997). Modifikasi tersebut ditempuh dengan cara menambah tetapan bias *k* yang relatif kecil pada diagonal matriks $X'X$, sehingga koefisien penduga gulud dipengaruhi oleh besarnya tetapan bias *k*. Dengan demikian paramater dugaan akan menjadi:

$$\widehat{\beta}\,(k)\left(X^{*'}X^{*}+kI\right)^{-1}X^{*\prime}Y \qquad ,k \geq 0$$

dengan :

$\widehat{\beta}$: Vektor koefisien regresi gulud

$X'X$: Matriks korelasi peubah *X*

k : Tetapan bias

I : Matriks identitas

$X'Y$: Vektor korelasi antara Y dan peubah *X*

Pemilihan besarnya tetapan bias *k* merupakan masalah yang perlu diperhatikan. Tetapan bias *k* yang diinginkan adalah tetapan bias yang menghasilkan bias relatif kecil dan menghasilkan koefisien penduga yang relatifstabil. Ada beberapa acuan yang digunakan untuk memilih besarnya *k*, diantaranya dengan melihat besarnya VIF dan melihat pola kecendrungan jejak gulud. Jejak gulud berupa plot dari penduga regresi gulud secara bersama dengan berbagai kemungkinan nilai tetapan bias *k* Gibbons dan McDonald (1984) dalam Herwindiati 1997. Nilai *k* yang dipilih yaitu *k* yang memberikan nilai penduga regresi gulud $\hat{\beta}\,(k)$ yang relatif stabil.

Hoerl dan Kennard (1970) dalam Gusriani (2004) menentukan nilai k dengan menggunakan jejak gulud yang merupakan suatu plot data antara $\hat{\beta}(k)$ dengan beberapa nilai k dalam selang antara 0 dan 1 hingga tercapai kestabilan pada parameter dugaannya. Akan tetapi pemilihan k dengan jejak gulud menjadi prosedur yang subjektif karena memerlukan keputusan peneliti untuk menentukan nilai k yang akan dipilih, Montgomery dan Peck (1992). Hoerl, Kennard, dan Balwin (1975) dalam Gusriani (2004) menyarankan pemilihan nilai k dengan menggunakan rumus HKB :

$$\widehat{K}(HKB) = P\hat{\sigma}^2/\,\hat{\beta}\,'\beta$$

dengan:

P adalah : banyaknya parameter diluar β_0

$\hat{\sigma}^2$ dan $\hat{\beta}$: diperoleh dari metode kuadrat terkecil

Pada penelitian selanjutnya Montgomery dan Peck (1992) mengajukan prosedur iterasi dengan menggunakan nilai *k* pada persamaan diatas sebagai nilai awal untuk menghitung nilai k dan selanjutnya $\hat{\beta}$ dan $\hat{\sigma}$ yang digunakan diperoleh dari metode regresi gulud dengan demikian prosedur ini akan berhenti jika:

$$\frac{[K(a_{i-1})-K(a_i)]}{K(a_i)} > \delta \text{ dan } T = \frac{trace(X'X)^{-1}}{p} \qquad (2.19)$$

dengan $\delta = 20\,T^{-13}$

Pemusatan dan Pengskalaan (*Centering and Scaling*)

Pemusatan dan pengskalaan data merupakan bagian dari membakukan (*standardized*) variabel. Pemusatan merupakan perbedaan antara masing-masing pengamatan dan rata-rata dari semua pengamatan untuk variabel. Sedangkan pengskalaan meliputi gambaran pengamatan pada kesatuan (unit) standar deviasi dari pengamatan untuk variabel Kutner (2005). Berikut ini merupakan pembakuan variabel terikat *Y* dan variabel bebas $X_1, \cdots, X_k$:

$$\frac{Y_i - \overline{Y}}{S_Y}$$

$$\frac{X_{ij}-\bar{Y}_j}{S_{x_j}}\;, j = 1,2,\cdots,k$$

dengan :

$\overline{Y}$: rata-rata dari *Y*

$\overline{X}_j$: rata-rata dari pengamatan X_j

S_Y : standar deviasi dari Y

S_{X_j} : standar deviasi dari X_j

Transformasi korelasi merupakan fungsi sederhana dari pembakuan variabel Sehingga melalui transformasi diperoleh persamaan sebagai berikut:

$$Y_i^* = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i - \bar{Y}}{S_Y}\right)$$

$$X_{ij}^* = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ij} - \bar{X}_j}{S_{X_j}}\right),\ j = 1,2,\ldots,k$$

Berdasarkan transformasi variabel Y_i^* dan X_{ij}^* yang didefinisikan dengan transformasi korelasi pada model Persamaan di atas diperoleh model regresi sebagai berikut:

$$y_i = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \ldots + \beta_k^* X_{ik}^*$$

Terdapat hubunga antara parameter regresi yang baku dengan parameter regresi, yaitu diantara parameter $\beta_1^*, \beta_2^*, \cdots, \beta_k^*$ pada model regresi baku dengan parameter $\beta_1, \beta_2, \cdots, \beta_k$ pada model regresi linear berganda yang biasa terdapat suatu hubungan linear. Hubungan antara kedua parameter dari dua model yang berbeda tersebut dijabarkan seperti di bawah ini Kutner (2005):

$$\beta_j = \left(\frac{S_Y}{S_{X_j}}\right)\beta_j^*,\ j = 1,2,\ldots,k$$

$$\beta_0 = \bar{Y} - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \ldots - \beta_k\bar{X}_k$$

$$= \bar{Y} - \sum_{j=1}^{k}\beta_j\bar{X}_j$$

Matriks Korelasi

Persamaan atau model yang didapat dari prosedur pemusatan dan penskalaan pada dapat dituliskan dalam bentuk matriks seperti berikut:

$$\begin{bmatrix} y_1 * \\ y_2 * \\ \vdots \\ y_k * \end{bmatrix} = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_k^* \end{bmatrix}\begin{bmatrix} X_{11}^* & X_{12}^* & \ldots & X_{1k}^* \\ X_{21}^* & X_{22}^* & \ldots & X_{2k}^* \\ \vdots & \vdots & & \vdots \\ X_{n1}^* & X_{n2}^* & \ldots & X_{nk}^* \end{bmatrix} + \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{bmatrix}$$

selanjutnya dari persamaan bentuk matriks diatas didapat matriks $X^{*'}X^*$ dan $X^{*'}Y^*$, yaitu:

$$X^{*'}X^* = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^{*2} & \sum_{i=1}^{n} X_{i1}^* X_{i2}^* & \ldots & \sum_{i=1}^{n} {}^*_{i1} X_{ik}^* \\ \sum_{i=1}^{n} X_{i2}^* X_{i1}^* & \sum_{i=1}^{n} X_{i2}^{*2} & \ldots & \sum_{i=1}^{n} X_{i2}^* X_{i2}^* \\ \vdots & \vdots & & \\ \sum_{i=1}^{n} X_{ik}^* X_{i1}^* & \sum_{i=1}^{n} X_{ik}^* X_{i2}^* & \ldots & \sum_{i=1}^{n} X_{ik}^{*2} \end{bmatrix}$$

$$X^{*'}y^* = \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^* y_i^* \\ \sum_{i=1}^{n} X_{i2}^* y_i^* \\ \vdots \\ \sum_{i=1}^{n} X_{ik}^* y_i^* \end{bmatrix} \quad (2.24)$$

pada persamaan di atas dapat diubah dalambentuk matriks korelasi. Matriks pertama adalah matriks korelasi dari variabel X dan dinotasikan dengan r_{xx}. dan matriks yang kedua adalah vektor yang berisikan koefisien korelasi sederhana diantara variabel Y dan setiap variabel X yang dinotasikan dengan r_{xy}, yaitu:

$$r_{xy} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{bmatrix}$$

$$r_{yx} = \begin{bmatrix} r_{y1} \\ r_{y2} \\ \vdots \\ r_{yk} \end{bmatrix}$$

Berdasarkan Persamaan tersebut diatas diperoleh persamaan seperti berikut ini:

$$X^{*'}X^* = r_{xy}$$

Untuk matriks kedua,yaitu adalah vektor yang berisikan koefisien korelasi sederhana diantara variabel terikat Y dan setiap variabel bebas X, yang dinotasikan dengan r_{yx}. Matriks korelasinya didefinisikan sebagai berikut:

$$X^{*'}Y^* = r_{yx}$$

Metode Penelitian

Sumber Data

Data yang digunakan untuk contoh pemakaian ini adalah data dari RSUD Arifin Achmad Pekanbaru Tahun 2012. (2.27)

Jurnal Sains, Teknologi dan Industri, Vol. 12, No. 1, Desember 2014, pp. 48 - 57
ISSN 1693-2390 print/ISSN 2407-0939 online

Peubah-peubah yang digunakan adalah sebagai berikut:

Y = Jumlah Tenaga Kerja
X_1 = Jumlah Pasien dengan Cara Bayar Melalui Umum
X_2 = Jumlah Pasien dengan Cara Bayar Melalui Askes Wajib
X_3 = Jumlah Pasien dengan Cara Bayar Melalui Askes Sukarela
X_4 = Jumlah Pasien dengan Cara Bayar Melalui Jamkesmas
X_5 = Jumlah Pasien dengan Cara Bayar Melalui Jamkesda
X_6 = Jumlah Pasien dengan Cara Bayar Melalui Jamsostek
X_7 = Jumlah Pasien dengan Cara Bayar Melalui Jampersal
X_8 = Jumlah Pasien dengan Cara Bayar Melalui Perusahaan

Metode Analisis

Analisis data dilakukan dalam bebrapa tahap, yaitu :

1. Mendeteksi keberadaan multikolenieritas. Adapun langkah-langkahnya sebagai berikut:
   a. Mendeteksi nilai Korelasi antar Peubah Bebas
   b. Mendeteksi nilai VIF (*Varians Inflantion Factors*)
   c. Menghitung Nilai Kondisi (ϕ)
2. Melakukan uji regresi linier dengan uji F
3. Penyelesaian masalah multikolenieritas menggunakan regresi gulud, dengan langkah-langkah sebagai berikut:
   a. Menentukan penaksiran koefisien regresi.
      1. Menghitung nilai penaksiran parameter β, selanjutnya menghitung simpangan baku.
      2. Menghitung $\hat{y}$ dan menganalisis tabel ANAVA.
   b. Menaksirkan koefisien regresi gulud. Dengan langkah-langkah sebagai berikut:
      1. Melakukan tranformasi terhadap matriks X dan vektor Y.
      2. Menghitung matriks $X'X$ atau r_{xx} yaitu matriks korelasi dari variabel bebas, serta menghitung $X'Y$ yang merupakan korelasi dari variabel bebas terhadap variabel tak bebas y
      3. Menghitung nilai penaksiran parameter β^* dengan berbagai kemungkinan tetapan bias k
   c. Menentukan nilai k yang bersesuaian dengan menggunakan prosedur iterasi Hoerl, Kennard, dan Balwin (1975) dalam Gusriani (2004) menyarankan pemilihan nilai k dengan menggunakan rumus HKB. Dan selanjutnya menentukan model regresi gulud.

HASIL DAN PEMBAHASAN

Mendeteksi Moltikolenieritas

Untuk pendeteksian multikolenieritas ada beberapa cara yang dapat digunakan antara lain:

1. Nilai Korelasi antar Peubah Bebas

Cara pertama adalah dengan melihat keeratan hubungan antar dua variabel bebas atau lebih yang dikenal dengan istilah korelasi. diperoleh koefisien korelasi parsial antar peubah bebasnya sebagai berikut :

Tabel 1. Matriks Korelasi dari Variabel *X*

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1	0.994	0.997	0.99	0.993	0.998	0.994	0.993
X_2	0.994	1	0.991	0.974	0.997	0.988	0.979	0.979
X_3	0.997	0.991	1	0.99	0.989	0.999	0.993	0.994
X_4	0.99	0.974	0.99	1	0.972	0.993	0.995	0.996
X_5	0.933	0.997	0.989	0.972	1	0.986	0.997	0.979
X_6	0.998	0.988	0.999	0.993	0.986	1	0.996	0.996
X_7	0.994	0.979	0.993	0.995	0.977	0.996	1	0.997
X_8	0.993	0.979	0.994	0.996	0.979	0.996	0.997	1

Berdasarkan Tabel 1 terlihat bahwa korelasi yang sangat tinggi mendekati nilai 1 antar variabel bebasnya. Hal ini menunjukkan adanya multikolenieritas.

Mendeteksi nilai VIF (*Varians Inflantion Factors*)

Mendeteksi apakah suatu model memiliki gejala multikolenieritas, akan digunakan cara VIF dan uji korelasi. Dengan cara ini akan dilihat apakah nilai VIF masing-masing variabel lebih besar 10 atau tidak. Bila nilai VIF lebih besar 10, maka diidentifikasikan model persamaan regresi tersebut mengalami multikoleniritas.

Tabel 2. Nilai VIF antar Variabel Bebas

Model	Pendugaan Parameter	VIF
Konstanta	4,071	
Umum	−0,129	1.136E3
Askes Wajib	0,040	294.870
Askes sukarela	0,048	797.145
Jamkesmas	−0,022	179.558
Jamkesda	0,071	340.964
Jamsostek	−0,080	1.584E3
Jampersal	0,105	395.610
Perusahaan	−0,039	362.543

Berdasarkan Tabel 2 dapat dilihat bahwa seluruh variabel bebas memiliki nilai VIF lebih besar dari 10, maka dapat disimpulkan model regresi ini memiliki masalah multikoleniritas.

Mendeteksi nilai kondisi $(\boldsymbol{\phi})$

Pendeteksian multikolenieritas dilakukan salah satunya dengan menghitung nilai kondisi (ϕ) yang menunjukkan tingkat multikolenieritas. Montgomery dan Peck (1992) mendefinisikan nilai kondisi dapat dilihat pada tabel berikut :

Tabel 3. Nilai Eigen dari $X'X$

No.	Model	Nilai Eigen
1	Umum	0,354
2	Askes Wajib	0,02
3	Askes Sukarela	0,0023
4	Jamkesmas	0,0017
5	Jamkesda	0,00158
6	Jamsostek	0,00073
7	Jampersal	0,00029
8	Perusahaan	0,00014757

Berdasarkan Tabel 3 dapat terlihat bahwa kedelapan nilai eigennya mendekati nol ini berarti terdapat multikoleniritas. Selain itu mutikolenieiritas dapat diukur dalam bentuk rasio atau perbandingan dari nilai terbesar dan terkecil nilai eigen. Nilai (ϕ) yang besar mengindikasikan multikolenieritas yang serius.

$$\phi = \frac{\lambda_{max}}{\lambda_{min}} = \frac{0,354}{0,00014757} = 2398,862$$

Berdasarkan semua proses pemeriksaan di atas berdasarkan Table 3, menurut Montgomery dan Peck (1992) ternyata nilai kondisinya mencapai 2398,862 lebihkecil dari 1000 yang menunjukkan adanya masalah multikolenieiritas yang kuat dalam data yang dianalisis. Untuk mengatasi masalah multikoleniritas tersebut, maka dilakukan analisis regresi gulud.

Menentukan Penaksiran Koefisien Regresi

Hasil analisis regresi dengan menggunakan metode kuadrat terkecil terhadap data pada tabel nilai penaksiran parameter (Tabel 4).

Tabel 4 Penaksiran Parameter Regresi Kuadrat Terkecil

Variabel	Pendugaan Parameter	Simpangan Baku
Konstanta	4,071	2,5424
Umum	−0,129	0,079
Askes Wajib	0,040	0,038
Askes Sukarela	0,048	0,065
Jamkesmas	−0,022	0,041
Jamkesda	0,071	0,039
Jamsostek	−0,080	0,088
Jampersal	0,105	0,046
Perusahaan	−0,039	0,047

Berdasarkan Tabel 4 di atas dapat diperoleh persamaan regresi linier berganda seperti pada persamaan berikut ini, yaitu:

$$\hat{Y} = 4.071 - 0.129\,X_1 + 0.040\,X_2 + 0.046\,X_3 - 0.022\,X_4 + 0.071\,X_5 - 0.08\,X_6 + 0.105\,X_7 - 0.039\,X_8$$

Meguji kecocokan model regresi linier ganda secara bersama-sama atau *simultan* dapat dilakukan melalui uij ANAVA, seperti yang terlihat pada Tabel 5 di bawah ini:

Pengujian keberartian model regresi linier ganda yang dilakukan secara *simultan* atau secara bersama-sama, dengan hipotesis:

H_0 : $\beta_i = 0$, untuk $i = 1,2,3$ (variabel bebas X secara individu tidak berpengaruh secara signifikan terhadap nilai taksiran Y)

H_1 : $\beta_i \neq 0$, untuk $i = 1,2,3$ (variabel bebas X secara individu berpengaruh secara signifikan terhadap nilai taksiran Y)

Tabel 5 ANAVA Untuk Data Awal

Sumber Keragaman	Jumlah Kuadrat	Derajat Kebebasan	Kuadrat Tengah	F_{hitung}	$F_{(8;6;0,005)}$	R^2-adj
Regresi	923,015	8	115,77	4,258	4,21	0,85
Galat	162,585	6	27,097			
Total	1085,600	14				

dengan kriteria uji:

Tolak H_0 jika $F_{hitung} \geq F_{(p;n-p-1;\alpha)}$ atau bisa juga dilihat dari nilai p, tolak $□_0$ jika nilai □ ≤ □. $□_{4.258}$> $□_{4.21}$yang berarti tolak $□_0$.

Berdasarkan tabel di atas terlihat bahwa nilai $□_{0.047}$ kurang dari $□_{0.05}$. Ini berarti semua variabel □ secara bersamaan berpengaruh terhadap nilai taksiran □, dari Tabel 5 bahwa $□^2 =$ 0,850 yang berarti mendekati satu, tidak diikuti dengan hasil uji hipotesis yang signifikan dari koefisien . Hal ini menunjukkan adanya kolinieritas.

Menentukan Penaksiran Koefisien Regresi Gulud

Untuk mengatasi terjadinya masalah multikoleniritas, maka dilakukan analisis regresi gulud. Adapun langkah-langkah dalam menaksirkan koefisien regresi gulud, yaitu:

1. Melakukan Transformasi terhadap Matriks X dan vektor Y

Dalam analisis regresi gulud digunakan data yang sudah ditransformasi terhadap matriks X dan vektor Y dapat dilakukan melalui metode pemusatan dan pengskalaan. Sebelum pemodelan regresi gulud dibentuk, data variabel bebas terlebih dahulu dipusatkan atau diskalakan untuk meminimumkan kesalahan pembulatan dan menganggap regresi sudah dipenuhi kenormalannya. Berdasarkan data maka diperoleh hasil dari transformasi variabel $□^*_□$ dan $□^*_{□□}$ melalui pemusatan dan pengskalaan.Dalam proses penaksiran regresi gulud, tetapan bias □ merupakan hal yang sangat penting dalam penelitian ini. Nilai dari koefisien penduga parameter $\hat{\beta}(□)$ dengan berbagai nilai □. Nilai □ yang terpilih dengan menggunakan rumus HKB adalah *0,004*. Sehingga persamaan regresi gulud yang diperoleh jika □ yang diambil sebesar *0,004*, yaitu:

$$\hat{□}^* = -0,7578\ □_1 * +1,6461\ □_2 * -0,0663\ □_3 * -1,4901\ □_4 * +2,3989\ □_5 * -1,1689\ □_6 * +0,878\ □_7 * -1,057\ □_8 *$$

Uji Keberartian Regresi Gulud

Model diperoleh kemudian akan diuji keberartian dari model tersebut, untuk melakukan pengujian regresi linier dilakukan sebagai berikut :

$□_0$: $□_0 = □_1 \cdots = □_□ = 0$ (regresi tidak berarti)

$□_1$: $□_□ \neq 0$(regresi berarti)

Pengujian keberartian model regresi gulud yang dilakukan secara *parsial* atau individu, dengan hipotesis:

$□_0$: $□_□ = 0$, untuk $□ = 1,2,3$ (variabel bebas □ secara individu tidak berpengaruh secara signifikan terhadap nilai taksiran □)

$□_1$: $□_□ \neq 0$, untuk $□ = 1,2,3$ (variabel bebas □ secara individu berpengaruh secara signifikan terhadap nilai taksiran □

Tabel 6 ANAVA Regresi Gulud

Sumber Keragaman	Jumlah Kuadrat	Derajat Kebebasan	Kuadrat Tengah	F_{hitung}	$F_{(8;6;0,005)}$	R^2-adj
Regresi	0,9498	8	0,11872	14,184	4,15	0,9498
Galat	0,0502	6	0,00837			
Total	1	14	0,12709			

dengan kriteria uji:

Tolak $□_0$ jika $□_{h□□□□□} \geq □_{(□;□-□-1;□)}$. $□_{14.1846}$> $□_{4.21}$ yang berarti tolak $□_0$. sehingga tolak $□_0$ maka dapat dinyatakan bahwa regresi berarti

Nilai $□^2 - □□□$ diatas dapat diartikan bahwa keragaman peubah terikat dapat dijelaskan oleh peubah bebas sebesar 85%. Hal ini mengalami peningkatan jika dibandingkan dengan R^2-adj yang diperoleh sebelumnya dengan Nilai $□^2 - □□□$ pada regresi gulud 94,98%.

Proses pemegembalian $□^*$ ke □ bentuk semula dengan $\bar{□} = 6,6$, $\bar{□}_1 = 740,733$, $\bar{□}_2 = 754,33$, $\bar{□}_3 = 731,533$, $\bar{□}_4 = 560,67$, $\bar{□}_5 = 806,267$, $\bar{□}_6 = 776,53$, $\bar{□}_7 = 745,67$, $\bar{□}_8 = 686,4$, $□_{□_1} = 3574,341$, $□_{□_2} = 3688,493$, $□_{□_3} = 3591,29$, $□_{□_4} = 2706,907$, $□_{□_5} = 3934,702$, $□_{□_6} = 3769,535$, $□_{□_7} = 3579,618$, $□_{□_8} = 3134,03$.

Sehingga setelah dikembalikan ke variabel-variabel asal diperoleh model persamaan regresinya menjadi:

$$\hat{\square} = 6{,}803 - 0{,}00859\,\square_1 + 0{,}018096\,\square_2 - 0{,}016993\,\square_3 - 0{,}02227\,\square_4 + 0{,}02467\,\square_5 - 0{,}01254\,\square_6 + 0{,}0099\,\square_7 - 0{,}001364\,\square_8$$

KESIMPULAN DAN SARAN

Kesimpulan

1. Adanya multikolenieritas dalam persamaan regresi linear berganda, ini terlihat dari besarnya nilai korelasi antar variabel bebas yang mendekati 1 dan seluruh nilai VIF lebih besar dari 10.
2. Dengan menggunakan metode regresi gulud, yaitu dengan menambahkan tetapan bias $\square = 0.004$, maka diperoleh persamaan regresi gulud:
$$\hat{\square} = 6{,}8\square_1 + 0{,}018\square_2 + 0{,}015\square_3 - 0.02\square_4 + 0.02\square_5 - 0{,}012\square_6 + 0{,}009\square_7 - 0{,}0036\square_8$$
3. Nilai korelasi determinasi estimator mendekati 1, yaitu $\square^2 = 0.949$. Hal ini menunjukkan bahwa estimator yang diperoleh sudah dapat digunakan.

Saran

Regresi gulud (*ridge regression*) belum tentu dapat digunakan untuk menyelesaikan semua model yang mengandung multikolinieritas, tetapi sudah cukup bukti bahwa regresi gulud merupakan salah satu metode yang baik. Ini dikarenakan melalui model ini diusahakan memperoleh variansi yang kecil dengan menentukan nilai $\square$ sehingga diperoleh keadaan yang lebih baik. Bagi pembaca yang tertarik melanjutkan penelitian ini bisa menyelesaikan masalah multikolinier dengan menggunakan metode lainnya.

DAFTAR PUSTAKA

Denereny, M., dan N.I. Rashwan . ”Solving Multicollinierity Problem Using Ridge Regression Models,” *Department of Statistics and Mathematics*, vol.6 halaman 585-600. 2011.

Draper, NR. 1998. *Applied Regression Analysis*. A Willey Interscience Publication, Canada

Gujarati, darmodar. *Ekonometrika Dasar*. Erlangga:Jakarta. 1997.

Hasan, Iqbal. *Analisis Data Penelitian dengan Statistik*. Bumi Aksara: Jakarta. 2004.

Myers, RH. 1990. *Clasissical and Modern Regression with Application 2nd edition*. Pws-Kent. Publishing Company, Boston.

Nurhasanah. “*Perbandingan Regresi Komponen Utama Terkoreksi dengan Regresi Ridge dalam Mengatasi Multikolenieritas*”. Pascasarjana. Jurusan Matematika FMIPA IPB. 2006.

Prenandita, Agriska. “*Penggunaan Metode Ridge Trace dan Variance Inflation Factors(VIF) Pada Regresi Ridge”*. Skripsi. Jurusan Matematika FMIPA UNY. 2011.

Sembiring. *Analisis Regresi*. Penerbit ITB: Bandung. 1995.

Sudarmanto, R.Gunawan. *Analisis Regresi Linier Berganda dengan SPSS*. Graha Ilmu: Yogyakarta. 2005.

Supranto, J. *Satistik Teori dan Aplikasi*. Erlangga:Jakarta. 2001.

Montgomery, D. C. and E. A. Peck (1992), *Introduction to Linear Regression Analysis*, 2nd edition, John Wiley & Sons, New York.