

PENERAPAN REGRESI AKAR LATEN DALAM MENANGANI MULTIKOLINEARITAS PADA MODEL REGRESI LINIER BERGANDA

DWI LARAS RIYANTINI¹, MADE SUSILAWATI², KARTIKA SARI³

^{1,2,3}Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali
e-mail: ¹dwie_rianti@yahoo.com, ²susilawati.made@gmail.com,
³sari_kaartika@yahoo.co.id

Abstract

Multicollinearity is a problem that often occurs in multiple linear regression. The existence of multicollinearity in the independent variables resulted in a regression model obtained is far from accurate. Latent root regression is an alternative in dealing with the presence of multicollinearity in multiple linear regression. In the latent root regression, multicollinearity was overcome by reducing the original variables into new variables through principal component analysis techniques. In this regression the estimation of parameters is modified least squares method. In this study, the data used are eleven groups of simulated data with varying number of independent variables. Based on the VIF value and the value of correlation, latent root regression is capable of handling multicollinearity completely. On the other hand, a regression model that was obtained by latent root regression has R_{adj}^2 value of 0.99, which indicates that the independent variables can explain the diversity of the response variables accurately.

Keywords: *Multiple Linear Regression, Multicollinearity, Latent Root Regression, Least Squares Method Modified*

1. Pendahuluan

Analisis regresi adalah suatu alat statistik yang dapat digunakan untuk melihat hubungan sebab akibat. Dalam analisis regresi terdapat peubah bebas dan peubah tak bebas. Peubah bebas dapat diukur, sedangkan peubah tak bebas atau yang juga disebut dengan peubah respon dijelaskan oleh satu atau lebih peubah bebas. Pada analisis regresi linier, peubah responnya memiliki skala pengukuran minimal interval. Berdasarkan banyak peubah bebas yang digunakan, analisis regresi linier dibagi menjadi dua yaitu analisis regresi linear sederhana dan analisis regresi linear berganda. Analisis regresi linier yang hanya melibatkan satu peubah bebas disebut analisis regresi linier sederhana, sedangkan analisis regresi linier dengan peubah respon dipengaruhi oleh lebih dari satu peubah bebas disebut analisis regresi

linier berganda (Myers & Milton, 1991). Dalam analisis regresi linier berganda, permasalahan yang sering muncul adalah adanya multikolinieritas.

Multikolinearitas ditandai dengan adanya korelasi di antara peubah-peubah bebas. Adanya multikolinearitas pada peubah-peubah bebas mengakibatkan model regresi yang diperoleh jauh dari akurat, diantaranya pengujian hipotesis parameter berdasarkan metode kuadrat terkecil (*ordinary least square*) memberikan hasil yang tidak valid yaitu peubah-peubah bebas yang seharusnya berpengaruh signifikan terhadap peubah respon dinyatakan sebaliknya secara statistik, tanda koefisien regresi dugaan yang dihasilkan bertentangan dengan kondisi aktual, penduga koefisien regresi bersifat tidak stabil sehingga mengakibatkan sulitnya menduga nilai-nilai peubah respon yang tentunya akan

¹Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

^{2,3}Staf Pengajar Jurusan Matematika FMIPA Universitas Udayana

mengakibatkan tidak akuratnya peramalan (Gujarati, 1995).

Terdapat beberapa metode untuk mengatasi adanya multikolinearitas dalam regresi linier berganda, salah satunya adalah dengan menggunakan regresi komponen utama (*principal component regression*). Pada regresi komponen utama, peubah-peubah bebas yang saling berkorelasi diubah ke dalam bentuk peubah-peubah baru yang tidak saling berkorelasi tanpa kehilangan banyak informasi dari peubah asal dan disebut dengan komponen utama. Teknik meregresikan komponen utama dengan peubah respon melalui metode kuadrat terkecil disebut regresi komponen utama (Gujarati, 1995). Pemilihan komponen utama pada regresi komponen utama adalah dengan memilih komponen utama yang memiliki akar ciri lebih besar dari 1 (Draper & H. Smith, 1992). Akan tetapi, proses ini memungkinkan komponen utama yang berguna untuk prediksi terhadap peubah respon akan terabaikan, karena pembentukan komponen utama yang tidak melibatkan informasi dari peubah respon (Vigneau, E., Qannari, E.M., 2002).

Perluasan regresi komponen utama diajukan oleh J.T. Webster et. al, dalam "Latent root regression analysis", *Technometrics*, 16, 1974. Webster dan rekan kerjanya menggandengkan matriks data yang berasal dari peubah respon yang telah dibakukan dan peubah bebas yang telah dibakukan. Perluasan ini dinamakan regresi akar laten (Draper & H. Smith, 1992). Perbedaan regresi akar laten dibandingkan regresi komponen utama adalah komponen utama yang terbentuk pada regresi akar laten diperoleh dengan menghitung hubungan antara peubah bebas dan peubah respon, sehingga komponen utama pada regresi akar laten lebih banyak mengandung informasi dibandingkan regresi komponen utama (Vigneau, E., Qannari, E.M., 2002).

1.1 Analisis Regresi Linier

Analisis regresi adalah suatu metode dalam statistik yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif, sehingga peubah respon (*dependent variable*)

bisa diramalkan dari peubah bebas (*independent variable*) (Neter, 1997). Selain untuk melihat hubungan antara peubah bebas dengan peubah respon, analisis regresi juga bertujuan untuk melihat kontribusi relatif dari masing-masing peubah bebas terhadap peubah respon.

Pola atau bentuk hubungan pada analisis regresi dapat dinyatakan dalam bentuk persamaan regresi. Model regresi linier yang melibatkan lebih dari satu peubah bebas dengan satu peubah respon disebut model regresi linier berganda. Analisis regresi linier berganda sangat berguna di dalam situasi percobaan yang memungkinkan peneliti mengontrol peubah-peubah bebasnya.

1.1.1 Model Ordo-Pertama

Misalkan terdapat n tripel data $(y_1, x_{11}, x_{12}), (y_2, x_{21}, x_{22}), \dots, (y_n, x_{n1}, x_{n2})$, (Neter, 1997) maka model regresinya dapat dinyatakan sebagai:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i \quad (1)$$

dengan:

y_i adalah respon dari amatan ke- i ,

β_0, β_1 , dan β_2 adalah koefisien regresi,

e_i adalah suku galat ke- i ,

$i = 1, 2, \dots, n$.

Jika $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$ maka persamaan

(1) dengan $i = 1, 2, \dots, n$ dapat ditulis sebagai:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (2)$$

Persamaan (2) dinamakan model ordo-pertama

dengan dua peubah bebas, yaitu $X_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}$

dan $X_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix}$. Model ini bersifat linier dalam

parameter dan juga linier dalam peubah-peubah bebasnya.

Apabila diasumsikan $E\{\varepsilon_i\} = 0$, maka fungsi respon bagi model (2) adalah (Neter, 1997):

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \quad (3)$$

Pada model regresi (3), parameter β_0 adalah intersep Y pada bidang regresi tersebut. Nilai parameter β_0 melambangkan rata-rata respon, apabila peubah bebas X_1 dan X_2 bernilai 0. Jika tidak demikian, β_0 tidak memiliki makna di dalam model regresi tersebut. Parameter β_1 menunjukkan perubahan rata-rata respon untuk setiap kenaikan X_1 satu satuan apabila X_2 dipertahankan konstan. Begitu pula, parameter β_2 menunjukkan perubahan rata-rata respon untuk setiap kenaikan X_2 satu satuan, apabila X_1 dipertahankan konstan. Parameter β_1 dan β_2 sering disebut *koefisien regresi parsial*.

Peubah bebas X_1 dan X_2 dikatakan memiliki *pengaruh aditif* atau *tidak berinteraksi*, apabila pengaruh X_1 terhadap rata-rata respon tidak bergantung pada taraf X_2 , dan sebagai akibatnya pengaruh X_2 terhadap respon juga tidak bergantung pada taraf X_1 (Neter, 1997).

Sebagai generalisasi dari model ordo-pertama dengan dua peubah bebas berikut ini dibahas model ordo-pertama dengan lebih dari dua peubah bebas. Oleh karena itu, apabila terdapat $p - 1$ peubah bebas $X_1 =$

$$\begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{bmatrix}, X_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{n2} \end{bmatrix}, \dots, X_{p-1} = \begin{bmatrix} x_{1,p-1} \\ x_{2,p-1} \\ \vdots \\ x_{n,p-1} \end{bmatrix}, \text{ maka}$$

modelnya [4] adalah:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_{p-1} x_{i,p-1} + e_i \quad (4)$$

dengan :

- p banyaknya parameter,
- $\beta_0, \beta_1, \dots, \beta_{p-1}$ adalah parameter,
- $x_{i1}, x_{i2}, \dots, x_{i,p-1}$ adalah peubah bebas yang diketahui nilainya,
- e_i adalah suku galat,
- $i = 1, 2, \dots, n$,
- n adalah banyak amatan.

$$\text{Jika } Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \varepsilon = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \text{ maka persamaan}$$

(4) dengan $i = 1, 2, \dots, n$ dapat ditulis sebagai:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon \quad (5)$$

Adapun fungsi respon (Neter, 1997) untuk model (5) adalah:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} \quad (6)$$

1.2 Koefisien Determinasi Ganda Terkoreksi

Dalam regresi linear berganda, proporsi keragaman data yang dapat diterangkan dalam model regresi dilihat dari koefisien determinasi ganda yang dilambangkan dengan R_{adj}^2 . (Neter, 1997) Koefisien determinasi ganda terkoreksi didefinisikan sebagai berikut:

$$R_{adj}^2 = 1 - \frac{JKG/(n-p)}{JKT/(n-1)} \quad (7)$$

Interval nilai R_{adj}^2 adalah $0 \leq R_{adj}^2 \leq 1$. Jika nilai R_{adj}^2 semakin mendekati 1, maka semakin besar nilai keragaman data peubah respon yang dapat dijelaskan oleh peubah bebas.

1.3 Multikolinearitas

Istilah multikolinearitas pertama kali diperkenalkan oleh Ragnar Frisch pada tahun 1934, yang berarti adanya korelasi di antara peubah-peubah bebas dari model regresi.

Multikolinearitas dapat memberi dampak untuk model regresi, antara lain (Neter, 1997):

1. Multikolinearitas antara peubah-peubah bebas dalam model regresi linier mengakibatkan variansi penduga kuadrat terkecil menjadi besar sehingga menghasilkan galat baku yang lebih besar. Hal ini mengakibatkan selang kepercayaan untuk parameter model regresi menjadi lebih besar.
2. Satu atau lebih peubah bebas menjelaskan peubah respon benar-benar sama dengan yang dijelaskan oleh peubah bebas lain.
3. Pengujian hipotesis parameter berdasarkan metode kuadrat terkecil memberikan hasil yang tidak valid.

Pada analisis regresi, dikatakan terdapat multikolinearitas apabila terdapat beberapa kondisi sebagai berikut:

1. Nilai korelasi antar peubah bebas (r_{XY}) melebihi 0,5 (Gujarati, 1995) Misalkan $(x_1, y_1), \dots, (x_n, y_n)$, pasangan data yang

diperoleh dari dua peubah acak $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

dan $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$. Nilai korelasi tersebut

diperoleh melalui rumus [7] sebagai berikut:

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{[\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2]^{\frac{1}{2}}} \quad (8)$$

Dalam hal ini X dan Y dianggap setara, tidak dipersoalkan apakah X dan Y yang menjadi peubah bebas atau peubah respon.

2. Nilai VIF lebih dari 4 (O'Brien, 2007) *Variance Inflation Factor* (VIF) atau faktor inflasi ragam dapat menginterpretasikan akibat dari korelasi antar variabel bebas ke- i pada varians penduga koefisien regresi. Adapun perhitungan VIF sebagai berikut (Neter, 1997):

$$VIF(i) = \frac{1}{1 - R_i^2} \quad (9)$$

Nilai $1 - R_i^2$ menunjukkan nilai toleransi yang mewakili varians dari peubah bebas ke- i yang tidak dihubungkan dengan peubah bebas lain pada model, sehingga nilai toleransi berbanding terbalik dengan nilai VIF. Nilai R_i^2 menunjukkan nilai korelasi antar peubah, kenaikan korelasi antar peubah akan mengakibatkan kenaikan nilai VIF yang menunjukkan terjadinya multikolinearitas. Jika $R_i^2 = 0$ atau $VIF = 1$, mengindikasikan bahwa peubah bebas ke- i orthogonal dengan peubah bebas lainnya.

1.4 Regresi Komponen Utama

Regresi komponen utama merupakan salah satu metode yang dapat digunakan untuk menangani multikolinearitas. Tahap pertama pada regresi komponen utama adalah menghitung komponen utama yang merupakan

kombinasi linier dari peubah bebas. Langkah selanjutnya, beberapa komponen utama yang terbentuk diregresikan dengan peubah respon melalui analisis regresi (Myers & Milton, 1991). Kriteria pemilihan komponen utama yang akan digunakan yaitu dengan memilih komponen utama yang bersesuaian dengan akar ciri lebih besar dari 1 (Draper, N.R. and H. Smith, 1992)

1.5 Regresi Akar Laten (*Latent Root Regression*)

Metode regresi akar laten merupakan perluasan dari regresi komponen utama. Perbedaan kedua metode ini terletak pada nilai akar laten yang dihasilkan dari matriks korelasi yang dihasilkan. Pada regresi akar laten, matriks korelasi diperoleh dari penggabungan peubah respon yang telah dibakukan dan peubah bebas yang telah dibakukan, yang dapat ditulis sebagai berikut (Draper, N.R. and H. Smith, 1992):

$$\mathbf{Z}^* = [\mathbf{Z}_y, \mathbf{Z}] \quad (10)$$

dengan \mathbf{Z}_y dan \mathbf{Z} secara berturut-turut merupakan matriks \mathbf{Y} dan \mathbf{X} yang telah dipusatkan dan diskalakan (dibakukan). Pembakuan data pada peubah respon diperoleh melalui rumus:

$$\mathbf{Z}_y = \frac{(\mathbf{Y} - \mathbf{1}\bar{y})}{\sqrt{S_{YY}}} \quad \text{dengan } \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad S_{YY} = \frac{(\mathbf{Y} - \mathbf{1}\bar{y})^T (\mathbf{Y} - \mathbf{1}\bar{y})}{n-1} \quad (11)$$

sedangkan, Pembakuan data pada peubah bebas diperoleh melalui rumus:

$$\mathbf{Z} = \frac{(\mathbf{X} - \mathbf{1}\bar{x})}{\sqrt{S_{XX}}} \quad \text{dengan}$$

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,p-1} \\ x_{21} & x_{22} & \dots & x_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{n,p-1} \end{bmatrix}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n},$$

$$\mathbf{1} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}_{n \times p-1},$$

$$S_{XX} = \frac{(X - \mathbf{1}\bar{x})^T (X - \mathbf{1}\bar{x})}{n-1} \quad (12)$$

Untuk matriks \mathbf{Y} dan \mathbf{X} seperti pada persamaan (10), setelah matriks \mathbf{Y} dan \mathbf{X} dibakukan, maka:

$$\mathbf{Z}^* = \begin{bmatrix} Z_{1y} & Z_{11} & \dots & Z_{1,p-1} \\ Z_{2y} & Z_{21} & \dots & Z_{2,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{ny} & Z_{n1} & \dots & Z_{n,p-1} \end{bmatrix}$$

Langkah berikutnya adalah melakukan analisis komponen utama berdasarkan matriks \mathbf{Z}^* . Seperti halnya dalam analisis komponen utama, akar laten dan vektor latennya kemudian dihitung dari matriks korelasi gandingan $\mathbf{Z}^{*T}\mathbf{Z}^*$

Misalkan $\Gamma_j^T = (\gamma_{0j}, \gamma_{1j}, \gamma_{2j}, \dots, \gamma_{rj})$ merupakan vektor laten dari matriks $\mathbf{Z}^{*T}\mathbf{Z}^*$ dan $\Gamma_j^0 = (\gamma_{1j}, \gamma_{2j}, \dots, \gamma_{rj})$ merupakan vektor yang terbentuk dari elemen yang sama dengan Γ_j^T kecuali elemen pertama yang telah dibuang, maka komponen utama (Sharma, S., James, W.L., 1986) dari \mathbf{Z}^* adalah:

$$C_j = \mathbf{Z}^* \Gamma_j \quad (13)$$

yang dapat dituliskan sebagai:

$$C_j = \gamma_{0j} \mathbf{Z}_y + \mathbf{Z} \Gamma_j^0 \quad (14)$$

Pada regresi akar laten, unsur pertama koefisien Y (γ_{0j}) setiap vektor laten digunakan untuk meramalkan peubah responnya oleh vektor laten tersebut. Untuk menentukan komponen utama yang akan digunakan, yaitu dengan membuang komponen utama yang bersesuaian dengan nilai akar laten $\lambda_j \leq 0.05$ atau elemen pertama vektor laten $|\gamma_{0j}| < 0.10$ (Vigneau, E., Qannari, E.M., 2002). Adanya akar laten yang kecil menandakan adanya kemungkinan ketergantungan atau ketidakbebasan linear di antara peubah-peubah bebas. Semakin kecil akar laten, semakin kuat ke tidak bebas linearan tersebut. Akar laten yang bernilai 0 menandakan adanya singularitas, dan nilai 0 pada elemen pertama dari suatu vektor laten menunjukkan bahwa vektor laten tersebut tidak memiliki kontribusi

variansi dalam Y (Sharma, S., James, W.L., 1986). Oleh karena itu, Webster menyarankan akar laten $\lambda_j \leq 0.05$ atau unsur pertama vektor laten padanannya $|\gamma_{0j}| < 0.10$, disarankan untuk dibuang.

Selanjutnya dihitung vektor koefisien kuadrat terkecil termodifikasinya (Webster, et al. 1974) dengan rumus:

$$\beta^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_p^* \end{bmatrix} = c \sum_j^* \gamma_{0j} \lambda_j^{-1} \begin{bmatrix} \gamma_{1j} \\ \gamma_{2j} \\ \vdots \\ \gamma_{np} \end{bmatrix}; \quad (15)$$

$$c = -\{\sum_j^* \gamma_{0j} \lambda_j^{-1}\}^{-1} \{\sum_{i=1}^n (Y_i - \bar{Y})^2\}^{1/2} \quad (16)$$

dengan:

λ_j adalah akar laten ke- j dari matriks $\mathbf{Z}^{*T}\mathbf{Z}$

γ_j adalah elemen vektor laten ke- j

γ_{0j} adalah elemen pertama dari vektor laten ke- j

$j = 0, 1, 2, \dots, p$

Selanjutnya, pendugaan koefisien regresi pada peubah awal diperoleh dengan membagi penduga koefisien regresi pada peubah yang telah dibakukan dengan S_j , (Draper, N.R. and H. Smith, 1992) sehingga diperoleh:

$$\beta_j = \frac{\beta_j^*}{S_j} \text{ dengan } S_j = \sqrt{\sum (x_j - \bar{x}_j)^2},$$

$$j = 1, 2, \dots, p-1 \quad (17)$$

Sedangkan, perhitungan koefisien regresi β_0 (Draper, N.R. and H. Smith, 1992) diperoleh berdasarkan rumus:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}_1 - \beta_2 \bar{x}_2 - \beta_3 \bar{x}_3 - \beta_4 \bar{x}_4 \quad (18)$$

Setelah persamaan kuadrat terkecil termodifikasinya diperoleh, Webster dan rekan-rekannya menyarankan untuk melakukan eliminasi langkah mundur untuk mengeluarkan peubah peramal dari persamaan itu (Webster, et al. 1974).

2. Metode Penelitian

Jenis data yang digunakan dalam penelitian ini adalah data sekunder berupa simulasi yang terdiri dari sebelas kelompok data dengan banyak peubah bebas bervariasi. Program yang digunakan dalam penelitian ini

adalah program Microsoft Excel dan Minitab 15.

Adapun tahap analisis data menggunakan regresi akar laten dengan langkah-langkah sebagai berikut:

- Melakukan pembakuan data pada peubah respon dan peubah bebas secara berturut-turut melalui persamaan (11) dan (12) dengan bantuan program Microsoft Excel.
- Memasangkan matriks data yang berasal dari peubah bebas dan peubah respon yang telah dibakukan.

$$\mathbf{Z}^* = [\mathbf{Z}_y, \mathbf{Z}]$$

- Menghitung akar laten λ_j dan vektor laten padanannya Γ_j dari matriks korelasi $\mathbf{Z}^{*T}\mathbf{Z}^*$ dengan bantuan Program Minitab 15.
- Melakukan pembentukan komponen utama melalui analisis komponen utama berdasarkan akar laten λ_j dan vektor laten padanannya Γ_j yang telah terbentuk pada program Minitab15.
- Memilih komponen utama yang digunakan dengan membuang komponen utama yang mempunyai nilai akar laten $\lambda_j \leq 0.05$ dan elemen pertama vektor laten $|\gamma_{0j}| < 0.10$ (Webster, et al., 1974).
- Berdasarkan langkah (e), komponen utama yang telah ditentukan diregresikan dengan peubah respon.
- Menghitung nilai VIF dan nilai korelasi antar peubah untuk mendeteksi apakah masalah multikolinearitas sudah teratasi.
- Melakukan pendugaan koefisien regresi pada data yang dibakukan melalui persamaan (15) dan (16).
- Melakukan pendugaan koefisien regresi pada peubah awal melalui persamaan (17) dan (18).

3. Hasil dan Pembahasan

Hasil analisis regresi linier berganda dengan menggunakan metode kuadrat terkecil pada sebelas kelompok data yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Model Regresi Linier Berganda

Mdl	Model Regresi Linier Berganda
I	$Y = 2,00 + 0,00 X_1 + 1,00X_2 + 0,00X_3 + 2,00X_4$
II	$Y = 10,2 + 1,15 X_1 + 1,02X_2 + 1,27X_3 + 0,737X_4 + 0,925X_5$
III	$Y = 2,00 + 1,00 X_1 + 1,00X_2 + 1,00X_3 + 1,0X_4 + 2,00X_5$
IV	$Y = 6,13 + 1,04 X_1 + 1,01X_2 + 1,06X_3 + 0,945X_4 + 0,953X_5 + 0,975X_6$
V	$Y = 2,00 + 1,00 X_1 + 1,00X_2 + 1,00X_3 + 1,00X_4 + 1,00X_5 + 2,00X_6$
VI	$Y = -2,53 + 0,823 X_1 + 0,973X_2 + 0,984X_3 + 1,06X_4 + 1,10X_5 + 0,991X_6$
VII	$Y = 2,00 + 1,00 X_1 + 1,00X_2 - 1,00X_3 + 1,00X_4 + 1,00X_5 + 2,00X_6$
VIII	$Y = -6,85 + 1,42 X_1 + 1,01X_2 + 1,15X_3 + 1,07X_4 + 1,03X_5 + 1,02X_6 + 1,00X_7 + 0,792X_8$
IX	$Y = -5,85 + 1,38 X_1 + 1,00X_2 + 1,28X_3 + 0,861X_4 + 1,21X_5 + 0,886X_6 + 1,01X_7 + 0,907X_8$
X	$Y = 3,91 - 0,108 X_1 + 0,982X_2 + 0,607X_3 + 1,29X_4 + 1,27X_5 + 0,955X_6 + 1,01X_7 + 1,13X_8$
XI	$Y = 3,89 + 0,614 X_1 + 0,985X_2 + 0,667X_3 + 1,16X_4 + 1,21X_5 + 1,11X_6 + 1,02X_7 + 0,913X_8$

Berdasarkan Tabel 1, model regresi linier I yang terbentuk adalah:

$$Y = 2,00 - 0,000000X_1 + 1,00X_2 + 0,000000X_3 + 2,00X_4$$

Model tersebut menginterpretasikan bahwa apabila semua peubah bebas diasumsikan konstan, maka peubah respon akan bernilai 2,00. Peubah respon tidak mengalami perubahan setiap kenaikan X_1 satu satuan selama X_2, X_3, X_4 dipertahankan konstan. Peubah respon akan meningkat sebesar 1,00 satuan setiap kenaikan X_2 satu satuan selama X_1, X_3, X_4 dipertahankan konstan. Interpretasi peubah bebas X_3 dan X_4 dapat dilakukan dengan cara yang sama. Model regresi lainnya dapat diinterpretasi dengan cara yang sama.

Untuk mendeteksi adanya multikolinearitas pada peubah bebas dapat dilihat berdasarkan nilai korelasi dan nilai VIF. Untuk model regresi I, nilai korelasi dan nilai VIF dapat dilihat pada Tabel 2.

Tabel 2 Nilai Korelasi dan Nilai VIF pada Model Regresi Linier I

NK	X_1	X_2	X_3	X_4	VIF
X_1	1				14,9
X_2	0,170	1			1,2
X_3	0,951	0,251	1		23,3
X_4	-0,961	-0,262	-0,977	1	30,7

Pada Tabel 2, terlihat bahwa X_1 dan X_3 memiliki nilai korelasi sebesar 0,951, X_1 dengan X_4 memiliki nilai korelasi sebesar -0,961, dan X_3 dengan X_4 memiliki nilai korelasi sebesar -0,977. Hal ini mengindikasikan adanya multikolinearitas di antara peubah bebas X_1 , X_3 dan X_4 . Selain berdasarkan nilai korelasi, indikasi adanya multikolinearitas antar peubah bebas X_1 , X_3 dan X_4 dipertegas dengan adanya nilai VIF yang lebih besar dari 4, sehingga dapat disimpulkan bahwa terdapat multikolinearitas pada ketiga peubah bebas tersebut. Di lain pihak, nilai korelasi pada peubah bebas X_2 kurang dari 0,5 dan nilai VIF kurang dari 4 menandakan bahwa peubah bebas X_2 tidak mengalami masalah multikolinearitas. Dengan cara yang sama, diperoleh bahwa terdapat beberapa peubah bebas yang mengalami multikolinearitas pada model regresi yang lain.

Adanya multikolinearitas pada peubah-peubah bebas mengakibatkan model regresi yang diperoleh jauh dari akurat, sehingga diperlukan alternatif dalam menangani multikolinearitas yang dalam penelitian ini dilakukan melalui regresi akar laten.

Regresi Akar Laten dalam Menangani Multikolinearitas

Langkah pertama dalam regresi akar laten adalah membakukan data dengan cara data dipusatkan (*centering*) dan diskalakan (*scaling*). Hal ini dilakukan untuk memudahkan perhitungan dan juga meminimumkan kesalahan pembulatan dalam perhitungan. Pada penelitian ini, pembakuan data dilakukan pada peubah respon dan peubah bebas. Data yang telah merupakan elemen-elemen pada matriks Z^* .

Akar laten λ_j dan vektor laten Γ_j dengan $j = 1, \dots, p - 1$ yang bersesuaian dengan λ_j

dibentuk dari matriks korelasi $Z^{*T}Z^*$. Untuk model regresi linier I diperoleh nilai-nilai akar laten yaitu:

$$\lambda_0 = 2,8029$$

$$\lambda_1 = 1,3790$$

$$\lambda_2 = 0,8015$$

$$\lambda_3 = 0,0167$$

$$\lambda_4 = 0,0000$$

Dari akar laten $\lambda_j, j = 0,1,2,3,4$, diperoleh vektor-vektor laten Γ_j yang bersesuaian dengan λ_j yaitu:

$$\Gamma_0 = \begin{bmatrix} -0,587 \\ -0,355 \\ -0,357 \\ -0,325 \\ -0,544 \end{bmatrix}$$

$$\Gamma_1 = \begin{bmatrix} -0,072 \\ -0,478 \\ -0,448 \\ 0,705 \\ 0,263 \end{bmatrix}$$

$$\Gamma_2 = \begin{bmatrix} 0,180 \\ -0,639 \\ 0,676 \\ 0,125 \\ -0,294 \end{bmatrix}$$

$$\Gamma_3 = \begin{bmatrix} 0,262 \\ -0,487 \\ -0,618 \\ -0,618 \\ 0,526 \end{bmatrix}$$

$$\Gamma_4 = \begin{bmatrix} -0,741 \\ -0,000 \\ 0,425 \\ 0,000 \\ 0,521 \end{bmatrix}$$

Tidak ada kriteria yang pasti dalam penentuan akar laten dan vektor laten yang digunakan untuk pembentukan komponen utama. Webster menyarankan untuk membuang akar laten $\lambda_j \leq 0,05$ atau unsur pertama vektor laten padanannya $|\gamma_{0j}| < 0,10$ [10]. Sedangkan, Sharma membuang akar laten $\lambda_j \leq 0,1$ atau unsur pertama vektor laten padanannya $|\gamma_{0j}| < 0,3$ dalam penelitiannya [8], dan Reichert membuang akar laten $\lambda_j \leq 0,3$ atau unsur pertama vektor laten padanannya $|\gamma_{0j}| < 0,10$ (Reichert, A.K.,

James, S.M., 1986). Dalam penelitian ini, penulis menggunakan kriteria pemilihan yang disarankan oleh Webster karena dengan menggunakan kriteria tersebut, model regresi yang diperoleh akan lebih akurat dengan data yang digunakan dalam penelitian ini. Oleh karena itu, dipilih akar laten $\lambda_j \leq 0.05$ atau elemen pertama vektor laten $|\gamma_{0j}| < 0.10$ [10]. Diperhatikan bahwa:

- a. $\lambda_0 > 0,05$ dan $|\gamma_{00}| = 0,587 > 0,10$. Oleh karena itu, vektor yang bersesuaian tetap dipertahankan.
- b. $\lambda_1 > 0,05$ dan $|\gamma_{01}| = 0,072 > 0,10$. Oleh karena itu, vektor ini tetap dipertahankan meskipun γ_{01} bernilai kecil.
- c. Karena $\lambda_2 > 0,05$ dan $|\gamma_{02}| = 0,180 > 0,10$ maka vektor ini tetap dipertahankan.
- d. $\lambda_3 < 0,05$ menandakan kemungkinan adanya ke tidak bebas linieran di antara peubah-peubah bebas. Akan tetapi, nilai $|\gamma_{03}| = 0,262 > 0,10$ menandakan keterampilan yang tinggi sehingga vektor ini tetap dipertahankan.
- e. $\lambda_4 = 0$ menandakan adanya singularitas, dan menandakan keadaan tidak bebas linier di antara peubah-peubah bebas yang menyebabkan pendugaan koefisien regresi menjadi tidak stabil, sehingga vektor ini dibuang walaupun nilai $|\gamma_{04}| = 0,741 > 0,10$ menandakan keterampilan yang tinggi.

Selanjutnya, dilakukan pembentukan komponen utama berdasarkan koefisien matriks (vektor laten). Berikut merupakan proses pembentukan dari lima komponen yang akan digunakan:

$$KU\ 1\ (C_0) = -0,587Z_y - 0,355Z_1 - 0,357Z_2 - 0,325Z_3 - 0,544Z_4$$

$$KU\ 2\ (C_1) = -0,072Z_y - 0,478Z_1 - 0,448Z_2 - 0,705Z_3 - 0,263Z_4$$

$$KU\ 3\ (C_2) = 0,180Z_y - 0,639Z_1 - 0,676Z_2 - 0,125Z_3 - 0,294Z_4$$

$$KU\ 4\ (C_3) = 0,262Z_y - 0,487Z_1 - 0,187Z_2 - 0,618Z_3 + 0,526Z_4$$

Komponen utama yang terbentuk merupakan kombinasi linier dari peubah asal yang saling

tegak lurus dan tidak berkorelasi. Berdasarkan hasil analisis regresi akar laten, adapun model regresi I yang terbentuk adalah:

$$Y = 366 - 9,94C_0 - 1,22C_1 + 3,06C_2 + 4,44C_3$$

Hasil perhitungan dengan menggunakan regresi akar laten pada model regresi I diperoleh nilai VIF masing-masing peubah bebas sebesar 1,0 dan nilai korelasi yang bernilai kurang dari 0,5 antar peubah bebas yang menandakan bahwa masalah multikolinearitas dapat diatasi secara tuntas. Nilai korelasi dan nilai VIF melalui regresi akar laten dapat dilihat pada tabel 3.

Tabel 3. Nilai Korelasi Antar dan Nilai VIF pada regresi akar laten

NK	C_1	C_2	C_3			VIF
C_1	1					1,0
C_2	-0,00	1				1,0
C_3	0,00	-0,00	1			1,0
C_4	0,00	0,00	0,00		1	1,0

dengan nilai koefisien determinasi ganda terkoreksi (R_{adj}^2) sebesar 1,00. Setelah itu, untuk memperoleh penduga koefisien regresi untuk regresi akar laten pada peubah awal digunakan persamaan (13) dan (14). Sehingga, untuk model regresi I, penduga koefisien pada data awal adalah sebagai berikut

$$Y = 19,095 + 7,054 X_1 + 1,095 X_2 + 7,489 X_3 - 5,515 X_4$$

Model tersebut menginterpretasikan jika pada saat semua peubah bebas diasumsikan konstan, maka peubah respon akan bernilai 19,095. Peubah respon akan meningkat sebesar 7,054 setiap kenaikan X_1 satu satuan selama X_2, X_3, X_4 dipertahankan konstan. Peubah respon akan berkurang sebesar 1,095 setiap kenaikan X_2 satu satuan selama X_1, X_3, X_4 dipertahankan konstan. Peubah respon akan meningkat 7,489 setiap kenaikan X_3 satu satuan selama X_1, X_2, X_4 dipertahankan konstan, dan peubah respon akan berkurang sebesar 5,515 setiap kenaikan X_4 satu satuan selama X_1, X_2, X_3 dipertahankan konstan.

Selanjutnya, untuk melihat seberapa akurat model yang diperoleh, dihitung R_{adj}^2 dengan menggunakan persamaan (3) dari masing-masing model. Nilai R_{adj}^2 hasil RAL pada masing-masing model dapat dilihat pada Tabel 4.

Tabel 4. Nilai R_{adj}^2 Model Hasil RAL

Model	R_{adj}^2
I	1,000
II	1,000
III	1,000
IV	1,000
V	1,000
VI	1,000
VII	1,000
VIII	1,000
IX	1,000
X	1,000
XI	1,000

Berdasarkan Tabel 4, nilai R_{adj}^2 sebesar 1,000 merupakan hasil pembulatan karena data yang digunakan dalam penelitian ini merupakan bilangan desimal, yang kemudian dalam prosesnya mengalami pembulatan berkali-kali.

4. Kesimpulan

Berdasarkan pembahasan dapat disimpulkan bahwa regresi akar laten dapat mengatasi multikolinearitas dengan tuntas dan menghasilkan persamaan regresi yang akurat.

Daftar Pustaka

- Draper, N.R. and H. Smith. 1992. *Analisis Regresi Terapan*, Edisi Kedua. Diterjemahkan oleh Bambang Sumantri. Jurusan Statistika FMIPA IPB. Bogor
- Gujarati N, Damorar. 1995. *Ekonometrika Dasar*. Erlangga. Jakarta.
- Myers, R.H. & Milton, J.S. 1991. *A First Course In The Theory Of Linier Statistical Models*. PWS-KENT Publishing Company, Boston
- Neter, J. 1997. *Model Linier Terapan*. Diterjemahkan oleh Bambang Sumantri. IPB, Bandung.

- O'Brien, R M. 2007. *A Caution Regarding Rules of Thumb for Variance Inflation Factor*. Departement of Sociology of Oregon, Eugene, USA.
- Reichert, A.K., James, S.M., 1986. Using Latent Root Regression to Identify Nonpredictive Collinearity in Statistical Appraisal Models. *AREUEA Journal*. 14, 136-152
- Sembiring, R.K. 2003. *Analisis Regresi*. Edisi Kedua. Bandung : Penerbit ITB.
- Sharma, S., James, W.L., 1986. *Latent Root Regression: An Alternate Procedure for Estimating Parameters in the Presence of Multicollinearity*. *JMR, Journal of Marketing Research*. 18, 154-161.
- Vigneau, E., Qannari, E.M., 2002. *A New Algorithm for Latent Root Regression Analysis*. *Computational Statistics & Data Analysis*. 41, 231-242.
- Webster, J.T., R. F. Gunts, and R. L. Mason.(1974). *Latent Root Regression Analysis*. *Technometrics* 16, 513-522.