PENGGUNAAN REGRESI AKAR LATEN UNTUK MEMPREDIKSI PENJUALAN MOBIL DI AMERIKA SERIKAT TAHUN 1961-1990

Edi Purwanto, Nar Herrhyanto, Maman Suherman

Departemen Pendidikan Matematika FPMIPA Universitas Pendidikan Indonesia

*Coresponding author: ediii.purwanto@gmail.com

ABSTRAK: Variabel tak bebas tidak hanya cukup dipengaruhi oleh satu variabel bebas dalam analisis regresi. Semakin banyak variabel bebas yang dilibatkan pada analisis regresi semakin baik pula untuk menaksir variabel tak bebasnya. Akan tetapi, hal ini juga menyebabkan peluang terjadinya multikolinearitas akan semakin besar. Regresi akar laten merupakan salah satu analisis regresi di mana terjadi hubungan di antara variabel-variabel bebasnya. Regresi akar laten akan menggunakan akar laten (nilai eigen) dan vektor laten (vektor eigen) yang diperoleh dari matriks yang entri-entrinya merupakan variabel bebas dan variabel tak bebas untuk membentuk persamaan regresi.

Kata kunci : *multikolinearitas, nilai eigen, vektor eigen, regresi akar laten.*

ABSTRACT: Dependent variableis notonlysufficientlyinfluencedbyanindependent variablein the regression analysis. The moreindependent variableswere includedin theregressionanalysisto estimatethe greater theindependentvariable. However, this is alsocausingthe possibility ofmulticollinearitywill be greater. latentrootregressionis one of theregressionanalysiswhere there isa relationshipbetweenthe independent variables. Latentrootregressionwilluse thelatentroots(eigenvalues) andlatentvectors(eigenvectors) areobtainedfrom thematrixwhose entriesareindependent variablesand thedependent variabletoestablish the regression equation.

Keywords: *multicollinearity, eigenvalues, eigenvectors, latentrootregression.*

PENDAHULUAN

Mobil adalah kendaraan darat yang digerakkan oleh tenaga mesin, beroda empat atau lebih biasanya menggunakan bahan bakar minyak (bensin atau solar) untuk menghidupkan mesinnya. Banyak manfaat yang diperoleh dari penggunaan mobil itu sendiri, salah satunya adalah dapat membantu seseorang dalam beraktifitas khususnya dalam bekerja. Hal ini menyebabkan setiap orang memiliki keinginan untuk memilikinya. Penduduk di Amerika Serikat dikenal produktif dalam pekerjaan, untuk mendukung dalam pekerjaannya mereka memiih menggunakan mobil pribadi sebagai alat transportasi daripada menggunakan kendaraan umum yang akan memakan waktu lebih lama.

Penjualan mobil di negara ini merupakan salah satu yang berkembang. Melihat minat penduduk untuk memiliki mobil semakin tinggi, maka pemerintah Amerika Serikat mulai mengembangkan pembuatan mobil. Pemerintah mencatat bahwa penjualan mobil setiap tahunya mengalami peningkatan. Hal ini diduga ada kaitannya dengan semakin meningkatnya pendapatan penduduk dari hasil kerja. Untuk melihat faktor- faktor apa saja yang berpengaruh dalam penjualan mobil bisa menggunakan analisis regresi. Analisis regresi adalah hubungan fungsional antara variabel-variabel yang pada umumnya dinyatakan dalam bentuk persamaan matematik (Sudjana, 1989: 310). Dalam analisis regresi dibedakan dua jenis variabel, yaitu variabel bebas atau variabel prediktor dan variabel tak bebas atau variabel respon.

Jika hubungan antara variabel bebas dan variabel tak bebas digambarkan dalam diagram pencar mengikuti pola garis lurus, maka disebut regresi linear, sebaliknya jika hubungan antara variabel bebas dan variabel tak bebas digambarkan dalam diagram pencar mengikuti pola sebuah kurva, maka disebut regresi non-linear. Selanjutnya analisis yang membahas bentuk hubungan antara satu variabel bebas dengan satu variabel tak bebas yang mengikuti pola garis lurus disebut analisis regresi linear sederhana, sedangkan analisis yang membahas bentuk hubungan antara lebih dari satu variabel bebas dengan satu variabel tak bebas disebut analisis regresi linear berganda.

Dalam analisis regresi berganda yang harus diperhatikan adalah diantara variabel-variabel bebasnya tidak terjadi multikolinearitas. Semakin banyak variabel bebas yang dilibatkan pada analisis regresi linear berganda semakin baik pula untuk menaksir variabel tak bebasnya. Akan tetapi, hal ini juga menyebebkan peluang terjadinya multikolinearitas akan semakin besar. Multikolinearitas adalah adanya hubungan linear diantara variabel-variabel bebasnya. Adanya multikolinearitas pada penaksiran parameter regresi linear berganda akan menghasilkan galat yang besar. Ada beberapa regresi linear berganda, dimana terjadi hubungan diantara variabel-variabel bebasnya, seperti regresi *ridge*, regresi komponen utama. Selain itu ada lagi regresi yang lainnya, yaitu regresi akar laten (Sharma dan James, 1981: 154).

LANDASAN TEORI

Analisis Regresi Linear

Analisis regresi linear adalah metode statistika yang dapat digunakan untuk mempelajari hubungan antar sifat permasalahan yang sedang diselidiki. Model analisis regresi linear dapat memodelkan hubungan antara dua variabel atau lebih. Pada model ini terdapat variabel tak bebas yang mempunyai hubungan fungsional dengan satu atau lebih variabel bebas. Jika pada regresi terdapat satu variabel bebas dan satu variabel tak bebas, maka disebut regresi sederhana. Sedangkan jika terdapat hubungan antara satu variabel tak bebas dan lebih dari satu variabel bebas, maka disebut regresi berganda.

Persamaan untuk regresi linear sederhana adalah :

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2.1)$$

dimana:

Y : variabel tak bebas

β_0 : konstanta

β_1 : koefisien regresi

X : variabel bebas

ε : galat

persamaan di atas akan ditaksir oleh persamaan $\hat{Y} = b_0 + b_1 X$. Koefisien b_0 dan b_1 dapat diperoleh dengan menggunakan persamaan dibawah :

$$b_0 = \frac{(\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} x_i y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

$$b_1 = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}$$

Untuk analisis regresi linear berganda persamaannya adalah :

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + \varepsilon \qquad (2.2)$$

di mana :

Y : variabel tak bebas

β_0 : konstanta

$\beta_1, \ldots, \beta_n$: koefisien regresi

$X_1, \ldots, X_n$: variabel bebas

ε : galat

Pernyataan di atas akan ditaksir oleh persamaan :

$$\hat{Y} = b_0 + b_1 X_1 + \cdots + b_n X_n$$

Dalam melakukan analisis regresi linear berganda, sering dijumpai masalah multikolinearitas pada variabel bebasnya. Menurut Damodar N. Gujarati dalam bukunya yang berjudul *Basic Econometrics*, persamaan regresi linear berganda harus memenuhi beberapa asumsi, salah satunya adalah tidak terjadi multikolinearitas di antara variabel bebasnya.

2.2 Multikolineritas

Multikolinieritas adalah terjadinya hubungan linier antara variabel bebas dalam suatu model regresi linier berganda (Gujarati, 2004). Salah satu dampak adanya multikolinearitas adalah penaksir parameter mempunyai varians yang besar sehingga sulit mendapatkan taksiran yang tepat. Ada beberapa cara untuk mendeteksi multikolinearitas, yaitu dengan melihat nilai *variance inflation factor* (VIF) pada model regresi dan membandingkan nilai koefisien determinasi individual (r^2)dengan nilai determinasisecara serentak(R^2). Jika nilai VIF lebih besar dari 5, maka variabel tersebut mempunyai persoalan multikolinearitas dengan variabel bebas lainnya. Selanjutnya untuk mendeteksi adanya multikolinieritas dengan membandingkan nilai koefisien determinasi individual (r^2) adalah dengan cara meregresikan setiap variabel bebas dengan variabel bebas lainnya, dengan tujuan untuk mengetahui nilai koefisien r^2 untuk setiap variabel yang diregresikan. Selanjutnya nilai r^2 tersebut dibandingkan dengan nilai koefisien determinasi R^2. Kriteria pengujian yaitu jika $r^2 > R^2$ maka terjadi multikolinearitas dan jika $r^2 < R^2$ maka tidak terjadi multikolinearitas.

REGRESI AKAR LATEN

Bentuk umum regresi akar laten adalah :

$$y^* = \omega \quad + \varepsilon \qquad (3.1)$$

dimana :

y^* : matriks berukuran $n \times 1$ yang berisi variabel tak bebas yang telah dipusatkan dan diskalakan

ω : matriks berukuran $n \times p$ yang berisi dari p variabel bebas yang telah dipusatkan dan diskalakan

η : matriks berukuran $p \times 1$ yang berisi parameter tak diketahui

ε : matriks berukuran $n \times 1$ yang berisi nilai galat

Diasumsikan variabel tak bebas *y* yang dipusatkan dan diskalakan kemudian disimbolkan dengan y^* :

$$y^* = \frac{y_i - \bar{y}}{s_y} \qquad (3.2)$$

di mana :

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \; d \quad S_y = \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

(Draper dan Smith, 1992)

Selanjutnya juga diasumsikan variabel bebas *x* yang dipusatkan dan diskalakan, kemudian disimbolkan dengan ω;

$\omega_p = \frac{x_i - \bar{x}}{S_x\sqrt{p}}$ (3.3) dimana :

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \; d \quad S_x = \sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(Draper dan Smith, 1992)

Untuk memperoleh penaksir parameter menggunakan regresi akar laten, diperlukan nilai eigen dan vektor eigen yang diperoleh dari perkalian matriks A^TA. Dimana entri-entri dari matriks matriks A adalah nilai-nilai dari variabel tak bebas dan variabel bebas yang sebelumnya telah dipusatkan dan diskalakan. Nilai eigen yang paling besar digunakan terlebih dahulu untuk menaksir koefisiennya. Setelah diperoleh nilai eigen, maka penaksir akar laten yang pertama dapat diperoleh dengan rumus :

$$\widehat{y_{\mathrm{U}}^{*(1)}} = -\gamma_0^{-1} \sum_{k=1}^{p} \gamma_{\mathrm{U}k}\omega_k \qquad (3.4)$$

Selanjutnya dengan menggunakan metode kuadrat terkecil, variabel tak bebas y^* diregresikan pada $\widehat{y_{\mathrm{U}}^{*(1)}}$ akan menghasilkan model penaksir bagi y^* yaitu $\widehat{y^*}$

$$\widehat{y^*} = q_1 \widehat{y_{\mathrm{U}}^{*(1)}} \qquad (3.5)$$

Langkah selanjutnya adalah melakukan penaksiran akar laten kedua. Penaksir ini diperoleh dengan terlebih dahulu menghitung nilai residu antara nilai variabel yang telah dipusatkan dan diskalakan dengan taksiran regresi akar laten. Nilai residu untuk variabel bebas dapat diperoleh dengan menggunakan persamaan berikut

$$r_{w_j}^{(1)} = \widehat{w_j} - w_j (j = 1,2, \dots, p) \qquad (3.6)$$

Dimana $\widehat{w_j}$ adalah nilai taksiran dari regresi setiap $w_j (j = 1,2, \dots, p)$ pada $\widehat{y_{\mathrm{U}}^{*(1)}}$ melalui persamaan

$$\widehat{w_j} = b_j^{(1)} \widehat{y_{\mathrm{U}}^{*(1)}} (j = 1,2, \dots, p) \qquad (3.7)$$

Nilai residu untuk variabel tak bebas dapat diperoleh dengan menggunakan persamaan berikut

$$r_{y^*}^{(1)} = \widehat{y^*} - y^* \qquad (3.8)$$

Dimana $\widehat{y^*}$ adalah nilai taksiran dari regresi setiap y^*pada $\widehat{y_{\mathrm{U}}^{*(1)}}$ melalui persamaan

$$\widehat{y^*} = q^{(1)} \widehat{y_{\mathrm{U}}^{*(1)}} \qquad (3.9)$$

Selanjutnya nilai residu tersebut dipusatkan dan diskalakan dengan cara yang sama seperti halnya pada variabel-variabel asal. Untuk memperoleh penaksir akar laten kedua bagi analisis regresi akar laten diperlukan matriks A_R, yaitu matriks hasil penggabungan variabel $r_y^{*(1)}$ dan variabel-variabel $r_w^{*(1)}$. Berdasarkan matriks tersebut dapat diperoleh nilai eigen dan vektor eigen dengan terlebih dahulu menghitung nilai dari perkalian matriks $A_R^T A_R$. Jika diperoleh nilai eigen yang mendekati satu, maka proses pendugaan akar laten selesai. Kemudian variabel yang telah diperoleh dikembalikan kedalam bentuk yang memuat variabel asal.

STUDI KASUS

Berikut ini diberikan data tentang penjualan mobil yang dipengaruhi oleh indeks harga mobil baru (X_1), indeks harga konsumen (X_2), pendapatan pribadi (X_3), tingkat bunga (X_4), dan banyaknya tenaga kerja (X_5).

Tabel 4.1 Data Penjualan Mobil di Amerika Serikat Tahun 1961 s.d. 1990

Y	X_1	X_2	X_3	X_4	X_5	Y	X_1	X_2	X_3	X_4	X_5
10,23	112,00	121,30	776,80	4,89	79,37	11,45	224,40	323,40	3022,10	6,31	109,60
10,87	111,00	125,30	839,60	4,55	82,15	10,22	213,50	310,20	3109,20	3,43	112,30
11,35	111,10	133,10	949,80	7,38	85,06	11,92	225,70	315,60	3150,50	5,34	135,60
8,78	117,50	147,70	1038,40	8,61	86,79	9,13	201,30	299,80	3175,90	5,30	140,30
8,54	127,60	161,20	1142,80	6,16	85,85	9,00	219,90	305,20	3289,90	6,73	147,80
9,99	135,70	170,50	1252,60	5,22	88,75	10,23	222,70	317,90	3318,60	7,20	152,60
11,05	142,90	181,50	1379,30	5,50	92,02	11,24	225,10	328,50	3357,50	7,89	155,60
11,16	153,80	195,30	1551,20	7,78	96,05	12,99	229,90	336,50	3425,50	8,20	175,90
10,56	166,00	217,70	1729,30	10,25	98,82	11,67	240,70	339,50	3479,90	9,45	160,40
8,98	179,30	247,00	1918,00	11,28	99,30	10,12	246,30	345,60	3561,10	12,34	177,30
8,54	190,20	272,30	2127,60	13,73	100,40	11,28	261,90	359,90	3599,90	10,23	186,30
7,98	197,60	286,60	2261,40	11,20	99,53	8,78	269,10	362,80	3629,80	16,29	183,40
9,18	202,60	297,40	2428,10	8,69	100,83	10,66	271,20	365,80	3658,80	16,30	192,50
10,39	208,50	307,60	2670,60	9,65	105,01	9,35	275,40	372,50	3728,90	16,98	193,50
11,04	215,20	318,50	2841,10	7,75	107,15	12,89	270,90	380,20	3800,00	17,56	201,60

Sumber : *Business Statistics, 1986,* A Supplement to the *Current Survey of Business*, U.S. Department of Commerce.

Keterangan :

y : Banyak mobil yang terjual (dalam ribuan unit)
x_1 : Indeks harga mobil baru
x_2 : Indeks harga konsumen
x_3 : Pendapatan pribadi (ribu dollar)
x_4 : Tingkat bunga (%)
x_5 : Banyak tenaga kerja sipil (ribuan orang)

Lakukan pemuatan dan penskalaan pada masing-masing variabel. Gunakan Persamaan (3.1) untuk variabel tak bebas dan Persamaan (3.2) untuk variabel bebas. Kemudian tentukan nilai eigen dan vektor eigen yang diperoleh dari perkalian matriks A^TA.

Tabel 4.2 Nilai Eigen dan Vektor Eigen dari Matriks $A_\omega^T A_\omega$

Nilai Eigen		Vektor Eigen					
j	$\lambda_j^{*(1)}$	$\gamma_{0j}^{*(1)}$	$\gamma_{1j}^{*(1)}$	$\gamma_{2j}^{*(1)}$	$\gamma_{3j}^{*(1)}$	$\gamma_{4j}^{*(1)}$	$\gamma_{5j}^{*(1)}$
0	1,9241	0,8463	0,2557	0,2485	0,2668	0,0864	0,2793
1	0,7274	0,5154	-0,4087	-0,4017	-0,3811	-0,3704	-0,3514
2	0,1162	0,1136	-0,1472	-0,2568	-0,3422	0,8815	0,00733
3	0,03	-0,0716	-0,2174	-0,4154	0,0006	-0,2188	0,8527
4	0,00161	0,0088	-0,7826	0,1892	0,5694	0,1499	-0,0686
5	0,005	-0,0077	-0,2931	0,7089	-0,5850	-0,0891	0,2476

Perhatikan nilai $\lambda_j^{*(1)}$= 1,9241 > 1 sehingga penaksir akar laten pertama $\left(\widehat{y_0^{*(1)}}\right)$ akan dihitung menggunakan Persamaan (3.4), yaitu :

$$\widehat{y_0^{*(1)}} = -\gamma_0^{-1} \sum_{k=1}^{p} \gamma_{0k} \omega_k = \begin{bmatrix} -0,174 \\ -0,171 \\ \vdots \\ 0,166 \end{bmatrix}$$

Selanjutnya variabel tak bebas y^*diregresikan pada $y_0^{*(1)}$ sesuai dengan Persamaan (3.5) untuk mendapatkan q_1, yaitu penaksir kuadrat terkecil untuk parameter regresi y^* pada $y_0^{*(1)}$ sehingga akan diperoleh persamaan $\widehat{y^*} = 0,38637\widehat{y_0^{*(1)}}$.

Langkah selanjutnya adalah melakukan penaksiran akar laten kedua. penaksir ini diperoleh dengan terlebih dahulu membentuk nilai residu yang diperoleh dari regresi variabel ω_j ($j = 1,2,3,4,5$) dan y^* pada $y_0^{*(1)}$. Kemudian nilai-nilai residu tersebut dipusatkan dan diskalakan dengan cara yang sama seperti halnya pada variabel-variabel sebelumnya. Tentukan kembali nilai eigen dan vektor eigen, jika nilai eigen sudah mendekati satu, maka penaksiran akar laten sudah selesai.

Tabel 4.3 Nilai Eigen Dan Vektor Eigen Dari Matriks $A_R^{(1)^T} A_R^{(1)}$

Nilai Eigen		Vektor Eigen					
j	$\lambda_j^{*(1)}$	$\gamma_{0j}^{*(1)}$	$\gamma_{1j}^{*(1)}$	$\gamma_{2j}^{*(1)}$	$\gamma_{3j}^{*(1)}$	$\gamma_{4j}^{*(1)}$	$\gamma_{5j}^{*(1)}$
0	1,0005	-0,9607	0,1124	0,0965	-0,551	0,1509	-0,1709
1	0,6311	0,1016	0,4267	0,5210	0,4385	-0,364	-0,4595
2	0,2495	0,2478	0,4194	0,1576	-0,5313	0,5806	-0,3439

Nilai Eigen		Vektor Eigen					
j	$\lambda_j^{*(1)}$	$\gamma_{0j}^{*(1)}$	$\gamma_{1j}^{*(1)}$	$\gamma_{2j}^{*(1)}$	$\gamma_{3j}^{*(1)}$	$\gamma_{4j}^{*(1)}$	$\gamma_{5j}^{*(1)}$
3	0,0496	0,0597	-0,7456	0,4232	0,1664	0,3583	-0,3243
4	0,0101	-0,0408	-0,1271	0,6033	-0,5922	-0,4093	0,3159
5	-0,00003	-0,0008	0,2390	0,3888	0,3791	0,4599	0,6605

Dari Tabel 4.3, diperoleh nilai $\lambda_0^{*(2)} = 1{,}0005 \approx 1$ itu artinya penaksiran akar laten kedua $y_i^{(2)}$ tidak mempunyai kemampuan untuk menaksir y^*. Oleh sebab itu maka model taksiran bagi y^* hanya melibatkan penaksir akar laten pertama $\left(y_i^{(1)}\right)$. Sehingga model taksiran regresi akar laten untuk y^* (data penjualan mobil yang diskalakan) dengan variabel bebas indeks harga mobil baru (x_1), indeks harga konsumen (x_2), pendapatan pribadi (x_3), tingkat bunga (x_4) dan banyak tenaga kerja sipil (x_5) adalah $\widehat{y^*} = 0{,}38637\widehat{y_i^{*(1)}}$. Model penaksir yang diproleh dikembalikan kebentuk semula menjadi

$$y = 0{,}00129x_1 + 0{,}00079x_2 + 0{,}00658x_3 + 0{,}00583x_4 + 0{,}001825x_5 + 9{,}3833$$

KESIMPULAN

Untuk memperoleh persamaan regresi dengan menggunakan regresi akar laten tahapannya adalah sebagai berikut :

1. Lakukan pemusatan dan penskalaan terhadap variabel tak bebas dengan $y^* = \frac{y_i - \bar{y}}{s_y}$ dan variabel bebas dengan $\omega_p = \frac{x_i - \bar{x}}{s_x\sqrt{p}}$.
2. Bentuk matriks yang entri-entrinya berisi variabel tak bebas dan variabel bebas yang telah dipusatkan dan diskalakan.
3. Tentukan nilai eigen dan vektor eigen yang diperoleh dari perkalian transpose matriks dengan matriks asalnya.
4. Nilai eigen yang paling besar digunakan untuk menaksir akar laten yang pertama, sesuai dengan $\widehat{y_i^{*(1)}} = -\gamma_0^{-1}\sum_{k=1}^{p}\gamma_{ik}\omega_k$.
5. Hitung nilai residu untuk setiap variabel tak bebas dan variabel bebas.
6. Lakukan pemusatan dan penskalaan terhadap nilai residu untuk memperoleh penaksir akar laten yang berikutnya. Jika nilai eigen yang diperoleh mendekati satu maka proses pendugaan akar laten selesai.
7. Variabel tak bebas dan variabel bebas dikembalikan kebentuk semula.

DAFTAR PUSTAKA

Draper, N.R. dan Smith, H. (1992). *Applied Regression Analysis (Second Edition)*. New York: John Wiley & Sons.

Gujarati, D.N. (2004). *Basic Econometrics (Fourth Edition)*. New York: The McGraw-Hill Companied.

Herrhyanto, Nar. (2003). *Statistika Matematis Lanjutan*. Bandung: CV. Pustaka Setia.

Sharma, S dan James, W.L. (1981). *Latent Root Regression: An Alternate Procedure for Estimating Parameters in the Presence of Multicollonearity*. Journal of Marketing Research. 18, 154-161.

Sudjana. (1989). *Metode Statistika (Edisi ke-5)*. Bandung: Tarsito.