

METODE REGRESI RIDGE UNTUK MENGATASI KASUS MULTIKOLINEAR

Margaretha Ohyver

Jurusan Matematika, Fakultas Sains dan Teknologi, Binus University
Jl. Kh.Syahdan No.9, Palmerah, Jakarta 11480
ethaohyver@binus.ac.id, mohyver@binus.edu

ABSTRACT

Multicollinear is a case that occurs in multi-linear regression analysis. Using multicollinear, it will be difficult to separate the influence of each independent variable towards the response variables. It also occurs in a farm production like cabbage. To solve this problem, Ridge regression method is used. This research aims to obtain a Ridge regression model to solve the multicollinear case. By using this method, the alleged regression coefficient is obtained by variance inflation factor less than ten for six free variables.

Keywords: multicollinear, Ridge regression.

ABSTRAK

Multikolinear merupakan salah satu kasus yang terjadi dalam analisis regresi linear ganda. Dengan adanya multikolinear, akan sulit memisahkan pengaruh masing-masing variabel bebas terhadap variabel respon. Kasus ini pun terjadi pada hasil produksi usaha tani kol bulat. Untuk mengatasi kasus ini, digunakan metode regresi Ridge. Tujuan penelitian ini adalah memperoleh model regresi Ridge yang dapat mengatasi kasus multikolinear. Berdasarkan metode ini diperoleh koefisien regresi dugaan dengan variance inflation factor yang kurang dari sepuluh untuk keenam variabel bebas.

Kata kunci: multikolinear, regresi Ridge.

PENDAHULUAN

Hubungan antara dua variabel (variabel bebas X dan variabel tak bebas Y) dalam suatu sistem yang kompleks tidak cukup dinyatakan dalam suatu persamaan regresi sederhana. Dalam situasi yang demikian, suatu variabel tak bebas atau variabel respon dapat dipengaruhi oleh lebih dari satu variabel bebas. Apabila persamaan regresi memuat lebih dari satu variabel bebas, model regresinya disebut model regresi ganda. Seperti halnya metode statistika lainnya, model regresi ganda mempunyai beberapa asumsi, di antaranya galat ε_i saling bebas dan berdistribusi normal $N(0, \sigma^2)$ serta tidak terjadi multikolinear (Supranto, 1986).

Asumsi yang terakhir terkadang dilanggar dalam data yang diambil dari keadaan tak terkontrol. Variabel-variabel bebas dalam situasi tersebut cenderung berkorelasi, bahkan berkorelasi tinggi. Korelasi antar variabel bebas inilah yang disebut dengan multikolinear (*multicollinearity*). Adanya kasus ini dapat menyebabkan sulitnya memisahkan pengaruh masing-masing variabel bebas terhadap variabel responnya (Rietveld & Sunaryanto, 1994). Multikolinear juga dapat menyebabkan kesalahan tanda (positif atau negatif) dari dugaan koefisien regresi kuadrat terkecil (Ryan, 1997). Akibat adanya pengaruh yang ditimbulkan oleh multikolinear tersebut diperlukan suatu metode untuk mengatasinya. Dan salah satu metode yang dapat digunakan adalah metode regresi Ridge.

Penelitian ini dilakukan dengan tujuan untuk memperoleh persamaan regresi Ridge yang dapat mengatasi kasus multikolinear. Dengan adanya penelitian ini diharapkan dapat memberikan pengetahuan mengenai cara mengatasi adanya multikolinear serta penerapan regresi Ridge pada data, yang dalam hal ini adalah data Hasil Produksi Usaha Tani Kol Bulat.

METODE

Metode regresi Ridge (*Ridge regression*) dapat digunakan untuk mengatasi korelasi yang tinggi antara beberapa variabel bebas (Hoerl dan Kennard, 1970). Regresi Ridge merupakan metode pendugaan koefisien regresi yang diperoleh melalui penambahan konstanta bias c pada diagonal $\mathbf{X}'\mathbf{X}$. Meskipun metode ini menghasilkan penduga koefisien regresi yang berbias, penduga ini bisa mendekati nilai parameter yang sebenarnya. Hal ini dapat diketahui dari perbandingan *mean square error* (MSE) antara penduga Ridge dengan penduga kuadrat terkecil (*least square*), dimana MSE penduga Ridge lebih kecil daripada MSE penduga kuadrat terkecil.

Jika $\hat{\beta}^*$ adalah penduga dari vektor β , jumlah kuadrat residual dapat ditulis sebagai berikut (Hoerl & Kennard, 1970):

$$\begin{aligned}\phi &= (\mathbf{Y} - \mathbf{X}\hat{\beta}^*)'(\mathbf{Y} - \mathbf{X}\hat{\beta}^*) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta}^*)'(\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}\hat{\beta} - \mathbf{X}\hat{\beta}^*) \\ &= [\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}(\hat{\beta} - \hat{\beta}^*)]'[\mathbf{Y} - \mathbf{X}\hat{\beta} + \mathbf{X}(\hat{\beta} - \hat{\beta}^*)] \\ &= (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) + (\hat{\beta} - \hat{\beta}^*)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \hat{\beta}^*)\end{aligned}\quad (1)$$

dimana $\hat{\beta}$ adalah penduga kuadrat terkecil dari β . Untuk ϕ tetap, maka dipilih nilai $\hat{\beta}^*$ dan dibuat meminimumkan $\hat{\beta}^{*'}\hat{\beta}^*$ dengan kendala $(\hat{\beta} - \hat{\beta}^*)'\mathbf{X}'\mathbf{X}(\hat{\beta} - \hat{\beta}^*) = \phi_0$, sehingga fungsi lagrangennya (Hoerl & Kennard, 1970) adalah:

$$F = \hat{\beta}^{*'}\hat{\beta}^* + \frac{1}{c}[(\hat{\beta}^* - \hat{\beta})'\mathbf{X}'\mathbf{X}(\hat{\beta}^* - \hat{\beta}) - \phi_0]\quad (2)$$

dimana $\frac{1}{c}$ adalah pengganda lagrange. Kemudian didiferensialkan terhadap $\hat{\beta}^*$.

$$\begin{aligned}
 \frac{\partial F}{\partial \hat{\beta}^*} &= 2\hat{\beta}^* + \frac{1}{c}[2(\mathbf{X}'\mathbf{X})\hat{\beta}^* - 2(\mathbf{X}'\mathbf{X})\hat{\beta}] \\
 &= 2\hat{\beta}^* + \frac{2}{c}(\mathbf{X}'\mathbf{X})\hat{\beta}^* - \frac{2}{c}(\mathbf{X}'\mathbf{X})\hat{\beta} \\
 &= \hat{\beta}^* + \frac{1}{c}(\mathbf{X}'\mathbf{X})\hat{\beta}^* - \frac{1}{c}(\mathbf{X}'\mathbf{X})\hat{\beta} \\
 \frac{\partial F}{\partial \hat{\beta}^*} &= 0 \\
 \hat{\beta}^* [1 + \frac{1}{c}(\mathbf{X}'\mathbf{X})] &= \frac{1}{c}(\mathbf{X}'\mathbf{X})\hat{\beta} \\
 \hat{\beta}^* [c\mathbf{I} + (\mathbf{X}'\mathbf{X})] &= (\mathbf{X}'\mathbf{X})\hat{\beta} \\
 \hat{\beta}^* &= [c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{X})\hat{\beta} \\
 \hat{\beta}^* &= [c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{Y}. \tag{3}
 \end{aligned}$$

Jadi penduga Ridge adalah $\hat{\beta}^* = [c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{Y}$.

Selain itu, penduga Ridge dapat juga ditulis dalam bentuk:

$$\begin{aligned}
 \hat{\beta}^* &= [c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{Y} \\
 &= [c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
 &= [c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} \\
 &= \mathbf{Z}\hat{\beta} \tag{4}
 \end{aligned}$$

Nilai harapan dari penduga Ridge adalah:

$$\begin{aligned}
 E(\hat{\beta}^*) &= E([c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{X}\hat{\beta}) \\
 &= E([c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}[(\mathbf{X}'\mathbf{X}) + c\mathbf{I} - c\mathbf{I}]\hat{\beta}) \\
 &= E([c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\mathbf{X}'\mathbf{X}\hat{\beta} + c[c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\hat{\beta} - k[k\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\hat{\beta}) \\
 &= E([c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}[\mathbf{X}'\mathbf{X}\hat{\beta} + c\mathbf{I}\hat{\beta}] - c[c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\hat{\beta}) \\
 &= E([c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}[\mathbf{X}'\mathbf{X} + c\mathbf{I}]\hat{\beta} - c[c\mathbf{I} + (\mathbf{X}'\mathbf{X})]^{-1}\hat{\beta}) \\
 &= E(\hat{\beta} - c(\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1}\hat{\beta}) \\
 &= \beta - c(\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1}\beta \tag{5}
 \end{aligned}$$

Karena $E(\hat{\beta}^*) = \beta - c(\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1}\beta$, penduga Ridge merupakan penduga yang bias dengan besarnya bias adalah $c(\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-1}\beta$.

Untuk memperoleh ragam dari penduga Ridge adalah sebagai berikut. Misalkan $\lambda_1, \lambda_2, \dots, \lambda_{k+1}$ adalah nilai-nilai eigen dari $(\mathbf{X}'\mathbf{X})$ dan $\hat{\beta}_{(c),0}^*, \hat{\beta}_{(c),1}^*, \hat{\beta}_{(c),2}^*, \dots, \hat{\beta}_{(c),k}^*$ adalah penduga koefisien regresi Ridge. Diketahui bahwa $\hat{\beta}^* = \mathbf{Z}\hat{\beta}$ sehingga

$$\text{var} [\hat{\beta}^*] = \text{var} [\mathbf{Z}\hat{\beta}]$$

$$\begin{aligned}
&= \mathbf{Z} \text{var}[\hat{\boldsymbol{\beta}}] \mathbf{Z}' \\
&= \mathbf{Z} \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{Z}' \\
&= \sigma^2 \mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{Z}' .
\end{aligned}$$

Karena yang akan dicari adalah $\sum \text{var} [\hat{\boldsymbol{\beta}}^*]$ maka

$$\begin{aligned}
\sum \text{var} [\hat{\boldsymbol{\beta}}^*] &= \sigma^2 \text{tr} [\mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{Z}'] \\
&= \sigma^2 \text{tr}[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{Z}] \\
&= \sigma^2 \sum \frac{\lambda_i}{(\lambda_i + c)^2} .
\end{aligned} \tag{6}$$

Sehingga ragam untuk penduga Ridge adalah $\sigma^2 \sum_{i=1}^{k+1} \frac{\lambda_i}{(\lambda_i + c)^2}$.

MSE diperoleh dengan menjumlahkan ragam dan besarnya bias dari penduga regresi Ridge.

$$\begin{aligned}
MSE(\boldsymbol{\beta}^*) &= \text{var}(\boldsymbol{\beta}^*) + \text{bias}(\boldsymbol{\beta}^*) \\
&= \sigma^2 \sum_{i=1}^{k+1} \frac{\lambda_i}{(\lambda_i + c)^2} + c^2 \boldsymbol{\beta}' (\mathbf{X}'\mathbf{X} + c\mathbf{I})^{-2} \boldsymbol{\beta} .
\end{aligned} \tag{7}$$

Adapun transformasi yang digunakan adalah transformasi korelasi, yaitu:

$$\begin{aligned}
y'_i &= \frac{1}{\sqrt{n-1}} \left(\frac{y_i - \bar{y}}{S_Y} \right) \\
x'_{ij} &= \frac{1}{\sqrt{n-1}} \left(\frac{x_{ij} - \bar{x}_j}{S_{X_j}} \right) \quad (j = 1, \dots, k)
\end{aligned} \tag{8}$$

dengan \bar{y} adalah rata-rata nilai y_i , \bar{x}_j adalah rata-rata nilai x_{ij} , S_Y adalah simpangan baku dugaan variabel Y , S_{X_j} adalah simpangan baku dugaan variabel X_j .

Data

Data yang akan digunakan dalam penelitian ini adalah data sekunder mengenai hasil produksi usaha tani kol bulat yang digunakan dalam skripsi S-1 Haerunissa (2004) dari Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Haluoleo. Ada 6 (enam) variabel bebas yang digunakan, yaitu: X_1 adalah benih (*ml*); X_2 adalah pupuk urea (*kg*); X_3 adalah pupuk TSP (*kg*); X_4 adalah pupuk KCL (*ml*); X_5 adalah pestisida (*ml*); X_6 adalah curahan hari kerja (*HKP*); Y adalah hasil produksi (*kg*).

Penelitian ini dilakukan dengan langkah-langkah sebagai berikut: (1) menganalisis data dengan metode regresi ganda, serta menentukan nilai VIF yang lebih dari sepuluh mengindikasikan adanya multikolinear; (2) menganalisis data dengan metode regresi Ridge.

HASIL DAN PEMBAHASAN

Data yang ada dianalisis dengan analisis regresi ganda. Nilai penduga kuadrat terkecil dan hasil ANOVA dapat dilihat pada Tabel 1 dan Tabel 2. Pada tabel 1 terlihat bahwa ada nilai VIF yang

lebih dari sepuluh. Hal ini berarti terjadi kasus multikolinear. Karena adanya kasus ini, maka diperlukan metode untuk mengatasinya. Untuk penelitian kali ini akan digunakan metode regresi Ridge.

Tabel 1
Penduga Parameter dan VIF

Peubah	Penduga	VIF
Konstan	-1286,72	
X ₁	20,53	27, 87
X ₂	2,92	3, 23
X ₃	13,65	4, 12
X ₄	-3,35	2, 71
X ₅	0,36	11, 16
X ₆	31,94	19, 79

Tabel 2
ANOVA untuk Regresi Linear Ganda

Sumber Keragaman	Jumlah Kuadrat	Derajat Bebas	Kuadrat Tengah	F hitung
Regresi	1,37x10 ⁷	6	2279279	122,497
Error	427957	23	18606,82	
Total	1,41x10 ⁷	29	486332,1	

Berdasarkan hasil tersebut, diperoleh persamaan regresi sebagai berikut:

$$\hat{y} = -1286.72 + 20.53X_1 + 2.92X_2 + 13.65$$

Langkah yang dilakukan untuk mengatasi kasus multikolinear ini adalah menganalisisnya dengan menggunakan metode regresi Ridge. Dan hasilnya dapat dilihat pada Tabel 3. Koefisien-koefisien regresi untuk berbagai nilai c dapat dilihat secara lengkap pada Tabel 3 di atas. Untuk memilih koefisien regresi yang mana yang akan digunakan dilakukan dengan melihat nilai VIF serta jejak Ridge (Tabel 4).

Tabel 3
Nilai koefisien regresi dugaan untuk berbagai c

c	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
0,000000	0,3319	0,0583	0,0737	-0,0140	0,0640	0,5099
0,001000	0,3321	0,0580	0,0738	-0,0136	0,0676	0,5056
0,002000	0,3322	0,0578	0,0739	-0,0132	0,0710	0,5015
0,003000	0,3323	0,0575	0,0741	-0,0127	0,0744	0,4976
0,004000	0,3322	0,0573	0,0742	-0,0123	0,0776	0,4939
0,005000	0,3321	0,0571	0,0744	-0,0119	0,0808	0,4903
0,006000	0,3319	0,0569	0,0746	-0,0115	0,0838	0,4868
0,007000	0,3317	0,0567	0,0747	-0,0111	0,0868	0,4835
0,008000	0,3314	0,0565	0,0749	-0,0107	0,0896	0,4803
0,009000	0,3311	0,0564	0,0751	-0,0103	0,0924	0,4773
0,010000	0,3308	0,0562	0,0753	-0,0099	0,0951	0,4743
0,020000	0,3264	0,0553	0,0771	-0,0063	0,1182	0,4493
0,030000	0,3214	0,0549	0,0788	-0,0028	0,1358	0,4302
0,040000	0,3165	0,0550	0,0804	0,0004	0,1495	0,4150
0,050000	0,3119	0,0552	0,0819	0,0034	0,1605	0,4023
0,060000	0,3076	0,0557	0,0832	0,0063	0,1693	0,3916
0,070000	0,3036	0,0562	0,0844	0,0090	0,1765	0,3823

0,080000	0,2999	0,0568	0,0855	0,0116	0,1825	0,3741
0,090000	0,2964	0,0575	0,0866	0,0141	0,1874	0,3668
0,100000	0,2932	0,0582	0,0876	0,0165	0,1916	0,3602
0,200000	0,2690	0,0656	0,0950	0,0354	0,2097	0,3162
0,300000	0,2526	0,0719	0,0998	0,0484	0,2115	0,2900
0,400000	0,2401	0,0769	0,1030	0,0577	0,2088	0,2712
0,500000	0,2298	0,0807	0,1052	0,0646	0,2047	0,2566
0,600000	0,2211	0,0835	0,1066	0,0697	0,2000	0,2445
0,700000	0,2136	0,0857	0,1075	0,0736	0,1954	0,2343
0,800000	0,2068	0,0873	0,1079	0,0765	0,1908	0,2254
0,900000	0,2008	0,0884	0,1081	0,0787	0,1864	0,2175
1,000000	0,1953	0,0892	0,1079	0,0804	0,1822	0,2105

Tabel 4
 Nilai VIF untuk Berbagai Nilai c

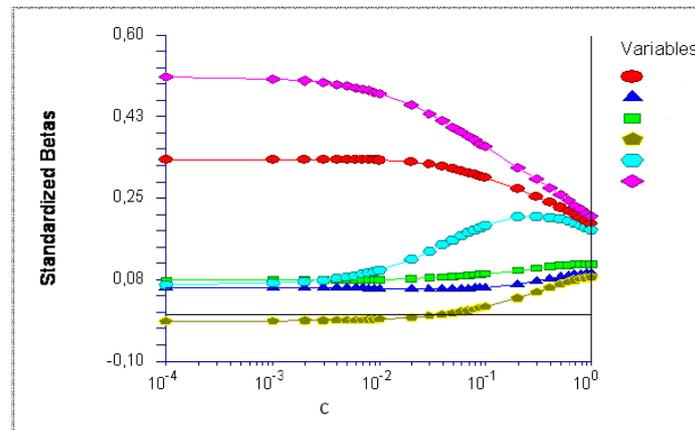
c	VIF (X_1)	VIF (X_2)	VIF (X_3)	VIF (X_4)	VIF (X_5)	VIF (X_6)
0,000000	27,8630	3,2260	4,1196	2,7077	11,1621	19,7908
0,001000	25,7442	3,1846	4,0412	2,6887	10,8057	18,4463
0,002000	23,8645	3,1454	3,9685	2,6701	10,4670	17,2468
0,003000	22,1890	3,1082	3,9008	2,6518	10,1448	16,1714
0,004000	20,6887	3,0729	3,8373	2,6338	9,8381	15,2029
0,005000	19,3400	3,0391	3,7777	2,6161	9,5457	14,3273
0,006000	18,1228	3,0066	3,7213	2,5988	9,2668	13,5325
0,007000	17,0205	2,9755	3,6679	2,5816	9,0005	12,8085
0,008000	16,0189	2,9454	3,6171	2,5647	8,7461	12,1469
0,009000	15,1059	2,9164	3,5686	2,5481	8,5027	11,5403
0,010000	14,2714	2,8883	3,5223	2,5317	8,2699	10,9827
0,020000	8,7707	2,6453	3,1417	2,3778	6,4027	7,2061
0,030000	5,9889	2,4474	2,8531	2,2396	5,1218	5,1930
0,040000	4,3801	2,2783	2,6171	2,1142	4,2032	3,9736
0,050000	3,3622	2,1301	2,4167	1,9997	3,5211	3,1703
0,060000	2,6753	1,9985	2,2427	1,8947	3,0001	2,6087
0,070000	2,1887	1,8805	2,0895	1,7982	2,5927	2,1983
0,080000	1,8306	1,7739	1,9534	1,7091	2,2678	1,8879
0,090000	1,5590	1,6771	1,8314	1,6268	2,0043	1,6466
0,100000	1,3477	1,5888	1,7215	1,5505	1,7874	1,4545
0,200000	0,5083	1,0058	1,0272	1,0187	0,7818	0,6330
0,300000	0,2916	0,7048	0,6926	0,7255	0,4656	0,3907
0,400000	0,2000	0,5264	0,5035	0,5459	0,3208	0,2787
0,500000	0,1511	0,4109	0,3854	0,4276	0,2403	0,2148
0,600000	0,1211	0,3314	0,3063	0,3452	0,1899	0,1736
0,700000	0,1010	0,2740	0,2506	0,2855	0,1558	0,1450
0,800000	0,0867	0,2312	0,2097	0,2408	0,1314	0,1240
0,900000	0,0759	0,1983	0,1787	0,2063	0,1131	0,1080
1,000000	0,0676	0,1723	0,1546	0,1791	0,0990	0,0954

Pada Tabel 4 terlihat bahwa mulai dari $c = 0,00000$ sampai $c = 1,00000$ nilai VIF akan semakin kecil. Nilai VIF yang akan diambil adalah nilai VIF yang relatif mendekati satu. Sehingga koefisien regresi yang akan digunakan adalah koefisien pada nilai $c = 0,10000$. Pada nilai c ini juga koefisien regresi mulai stabil. Pada Gambar 1 disajikan jejak Ridge untuk berbagai nilai c .

Sehingga persamaan regresi untuk data kol bulat adalah:

$$\hat{y}^* = 0,2932 Z_1 + 0,0582 Z_2 + 0,0876 Z_3 + 0,0165 Z_4 + 0,1916 Z_5 + 0,3602 Z_6.$$

Jika dilihat dari hasil tersebut, diketahui bahwa benih, pupuk urea, pupuk TSP, pupuk KCL, pestisida, dan HKP, memberikan pengaruh positif terhadap hasil produksi kol bulat.



Gambar 1. Jejak Ridge.

PENUTUP

Berdasarkan penelitian yang dilakukan dapat disimpulkan bahwa persamaan regresi untuk data hasil produksi usaha tani kol bulat adalah:

Koefisien-koefisien regresi untuk persamaan ini diperoleh pada nilai bias dengan nilai VIF yang kurang dari sepuluh.

DAFTAR PUSTAKA

- Haerunissa. (2004), *Penggunaan Metode Regresi Komponen Utama pada Kasus Data yang Memiliki Kasus Kolinear Ganda*. Skripsi S1. Kendari: Jurusan Matematika, FMIPA Universitas Haluoleo.
- Hoerl, A. E., and Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *A Journal of Statistics for the Physical Chemical and Engineering Sciences*, 12 (1), 55-67.
- Rietveld, P., dan Sunaryanto, L. T. (1994). *87 Kasus Pokok dalam Regresi Berganda* (edisi pertama). Yogyakarta: Andi Offset.
- Ryan, T. P. (1997). *Modern Regression Method*. New York: Wiley.
- Supranto, J. (1986). *Pengantar Probabilita dan Statistik Induktif* (edisi pertama). Jakarta: Erlangga.