

PERBANDINGAN REGRESI KOMPONEN UTAMA DAN ROBPCA DALAM MENGATASI MULTIKOLINEARITAS DAN PENCILAN PADA REGRESI LINEAR BERGANDA

NI WAYAN YULIANI¹, I KOMANG GDE SUKARSA²,
I GUSTI AYU MADE SRINADI³

^{1,2,3}Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali
e-mail: ¹yuliani_ani@rocketmail.com, ²sukarsakomang@yahoo.co.id,
³srinadiigustiyumade@yahoo.co.id

Abstract

Multiple linear regression analysis with a lot of independent variable always makes many problems because there is a relationship between two or more independent variables. The independent variables which correlated each other are called multicollinearity. Principal component analysis which based on variance covariance matrix is very sensitive toward the existence of outlier in the observing data. Therefore in order to overcome the problem of outlier it is needed a method of robust estimator toward outlier. ROBPCA is a robust method for PCA toward the existence of outlier in the data. In order to obtain the robust principal component is needed a combination of Projection Pursuit (PP) with Minimum Covariant Determinant (MCD). The results showed that the ROBPCA method has a bias parameter and Mean Square Error (MSE) parameter lower than Principal Component Regression method. This case shows that the ROBPCA method better cope with the multicollinearity observational data influenced by outlier.

Keywords: *Multiple Linear Regression, Principal Component Regression, ROBPCA (Robust Principal Component Analysis), multicollinearity, Outlier*

1. Pendahuluan

Analisis regresi linear berganda adalah salah satu metode statistika yang digunakan untuk mengetahui pengaruh sebuah variabel tidak bebas (*dependent variable*) dengan dua atau lebih variabel bebas (*independent variable*)[2]. Adapun tujuan dari analisis regresi linier berganda adalah mengetahui seberapa besar pengaruh beberapa variabel bebas terhadap variabel tidak bebas dan juga dapat meramalkan nilai variabel tidak bebas apabila seluruh variabel bebas sudah diketahui nilainya.

Pada analisis regresi linier berganda dengan banyak variabel bebas, sering timbul masalah karena adanya hubungan antara dua atau lebih variabel bebas. Variabel bebas yang saling berkorelasi disebut multikolineari. Permasalahan yang terjadi pada analisis regresi berganda dapat mengakibatkan hasil analisis yang

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

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

kurang akurat. Multikolinearitas merupakan salah satu masalah yang terjadi pada analisis regresi linear berganda. Masalah lain yang dapat memengaruhi hasil analisis data adalah pencilan (*outlier*). Pada penelitian ini akan dibahas dua permasalahan statistik tersebut.

Pada kasus multikolinearitas, korelasi antar variabel akan menyebabkan jumlah kuadrat galat yang semakin besar sehingga menghasilkan keputusan yang tidak *significant*. Kasus multikolinearitas juga sangat berpengaruh pada bentuk matriks. Pada pendugaan parameter $\hat{\beta} = (X'X)^{-1}X'Y$, apabila terjadi multikolinearitas maka matriks $X'X$ *singular*, sehingga persamaan untuk pendugaan estimasi parameter tidak lagi mempunyai penyelesaian yang tunggal. Hal ini akan berdampak pada dugaan koefisien variabel tidak tunggal, melainkan tidak terhingga banyaknya sehingga tidak mungkin untuk menduganya [3].

Metode regresi komponen utama (*Principal Component Regression*) merupakan salah satu teknik dalam mengatasi multikolinearitas dengan cara mereduksi variabel-variabel yang ada menjadi beberapa variabel baru yang saling bebas dan merupakan kombinasi linier dari variabel asal (Montgomery [1]).

Dalam menentukan komponen utama pada metode Regresi Komponen Utama yakni melalui tahapan *Principal Component Analysis* (PCA). Analisis komponen utama yang berdasarkan matriks varian kovarian sangat sensitif terhadap adanya pencilan pada data pengamatan, sehingga untuk mengatasi masalah pencilan diperlukan suatu metode penduga yang tegar terhadap pencilan. *ROBPCA* (*Robust Principal Component Analysis*) adalah suatu metode yang kuat (*robust*) untuk PCA terhadap keberadaan pencilan pada data, untuk mendapatkan komponen utama yang *robust* diperlukan penggabungan konsep *Projection Pursuit* (PP) dengan penduga *robust Minimum Covariance Determinant* (MCD)[4].

Penduga *robust* MCD merupakan nilai matriks rata-rata dan matriks kovarian dari sebagian pengamatan yang meminimumkan determinan matriks kovarian. Penduga ini didapat dengan cara mencari h pengamatan yang memberikan nilai minimum dari matrik kovarian (Sunaryo, [4]).

Nilai matriks rata-rata t_l dan matriks kovarians C_l dirumuskan sebagai:

$$t_l = \frac{1}{h} (H_b)^T \cdot V^* \quad (1.1)$$

$$C_l = \frac{1}{h} (H_b - V^*(t_l)^T)^T (H_b - V^*(t_l)^T) \quad (1.2)$$

2. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data simulasi yang diperoleh dengan membangkitkan data yang berdistribusi normal, pemeriksaan multikolinearitas, pemeriksaan pencilan, serta penyelesaian Regresi Komponen Utama dan *Robust Principle Component Analysis*. Software yang digunakan adalah MINITAB 15 dan R i386 2.15.2.

Langkah – langkah yang dilakukan dalam penelitian ini yakni:

1. Membangkitkan data yang mengandung multikolinearitas dengan empat variabel bebas dan satu variabel tidak bebas, amatan yang dibangkitkan sebanyak 100 amatan ($n=100$). Nilai sisaan (ε) yang dibangkitkan berdistribusi normal dengan rata-rata nol dan ragam satu. Nilai sisaan yang dibangkitkan berukuran 100 amatan.
2. Melakukan analisis regresi linear berganda setelah membentuk variabel tidak bebas dari beberapa variabel bebas. Pada langkah ini model regresi yang didapat tidak sesuai dengan yang diharapkan sehingga diperlukan suatu analisis untuk mendapatkan nilai penduga yang mendekati nilai yang diharapkan.
3. Menganalisis data bangkitan dengan metode regresi Komponen utama. Hasil analisis ini akan menjadi nilai penduga yang fit untuk mencari bias parameter.
4. Membangkitkan pencilan yang berdistribusi normal ($N(40;0,05)$) dengan persentase pencilan 10%, 15%, dan 20%. Selanjutnya menambahkan pencilan pada masing-masing peubah bebas dan melakukan analisis komponen utama.
5. Melakukan analisis dengan *Robust Principle Component Analisis* (ROBPCA) pada data pengamatan yang dipengaruhi oleh pencilan.
6. Langkah terakhir yakni membandingkan dan menganalisis kedua metode tersebut. Yang digunakan sebagai pembanding adalah nilai bias parameter dan nilai *Mean Square Error*.

3. Hasil dan Pembahasan

3.1 Perbandingan Regresi Komponen Utama (RKU) dan *Robust Principle Component Analysis* (ROBPCA)

Bias parameter merupakan nilai harapan dari selisih antara nilai estimasi dan nilai yang sebenarnya. Nilai yang digunakan sebagai acuan adalah nilai penduga dari metode Regresi Komponen Utama yang tidak dipengaruhi oleh pencilan, sedangkan nilai yang digunakan sebagai nilai estimasi adalah nilai penduga dari metode Regresi Komponen Utama dan *Robust Principle Component Analysis* yang dipengaruhi oleh pencilan dengan persentase 10%, 15%, dan 20%. Bias parameter untuk 100 kali ulangan diperoleh dengan rumus berikut ini:

$$BIAS(\beta_i) = \left(\frac{\sum_{j=1}^{100} \hat{\beta}_j^{fit}}{100} \right) - \left(\frac{\sum_{j=1}^{100} \hat{\beta}_j^p}{100} \right), \quad i = 1, \dots, 4 \quad (3.1)$$

Mean Square Error (MSE) suatu estimator merupakan nilai harapan dari bias kuadrat. *Mean Square Error* parameter untuk 100 kali ulangan diperoleh dengan menggunakan rumus berikut ini:

$$MSE(\beta_i) = \frac{\sum_{j=1}^{100} (\hat{\beta}_j^{fit} - \hat{\beta}_j^p)^2}{100}, \quad i = 1, \dots, 4 \quad (3.2)$$

Hasil perhitungan bias parameter dan *Mean Square Error* untuk 100 kali ulangan disajikan berturut-turut pada Tabel 1 dan Tabel 2.

Tabel 1. Perbandingan Regresi Komponen Utama Dan ROBPCA Berdasarkan Nilai Bias Parameter

Pencilan	Komponen	Nilai Bias Parameter	
		Regresi Komponen Utama	ROBPCA
10%	Intercept	19,6368	1,3057
	PC1	1,4974	0,3775
	PC2	0,4915	0,1309
	PC3	0,0803	0,0023
	PC4	0,6893	0,1160
15%	Intercept	29,8135	6,3881
	PC1	1,5809	0,3581
	PC2	0,5197	0,0714
	PC3	0,2581	0,0610
	PC4	0,9431	0,3668
20%	Intercept	39,8165	13,0083
	PC1	1,1576	0,1045
	PC2	0,4013	0,1743
	PC3	1,4120	0,0457
	PC4	1,1039	0,4341

Sumber : Data diolah (2013)

Estimasi yang baik adalah estimasi yang menghasilkan nilai bias yang rendah atau kecil. Semakin besar nilai bias, maka semakin jauh penyimpangan dari nilai yang sebenarnya. Pada Tabel 1 menunjukkan bahwa nilai penduga dari metode ROBPCA selalu lebih kecil dibandingkan dengan Regresi Komponen Utama. Oleh Karena itu nilai penduga dari metode ROBPCA lebih baik dibandingkan dengan metode Regresi Komponen Utama. Hal ini karena metode ROBPCA dapat mengatasi data pengamatan yang dipengaruhi oleh pencilan.

Tabel 2. Perbandingan Regresi Komponen Utama Dan ROBPCA Berdasarkan Nilai *Mean Square Error* Parameter

Pencilan	Komponen	Nilai Bias Parameter	
		Regresi Komponen Utama	ROBPCA
10%	Intercept	392,0726	6,6843
	PC1	3,9182	1,5095
	PC2	2,8191	2,4226
	PC3	0,1875	0,0206
	PC4	2,7543	1,5256
15%	Intercept	891,3347	48,4942
	PC1	4,0961	1,4267
	PC2	3,1972	3,0118
	PC3	0,4456	0,0494
	PC4	3,0174	1,6895
20%	Intercept	1587,84599	173,4866
	PC1	3,012465527	1,8322
	PC2	2,682504765	2,2665
	PC3	2,164030949	0,0714
	PC4	3,060462252	1,8983

Sumber : Data diolah (2013)

Semakin kecil nilai *Mean Square Error* suatu estimator, maka hasil estimasinya akan semakin baik. Pada Tabel 2 menunjukkan bahwa nilai penduga dari metode ROBPCA selalu lebih kecil dibandingkan dengan Regresi Komponen Utama. Sehingga nilai penduga dari metode ROBPCA lebih baik dibandingkan dengan metode Regresi Komponen Utama. Hal ini karena metode ROBPCA dapat mengatasi data pengamatan yang dipengaruhi oleh pencilan.

4. Kesimpulan

Pada penelitian ini nilai penduga dari metode ROBPCA memiliki nilai bias parameter dan nilai *Mean Square Error* (MSE) yang lebih kecil dibandingkan dengan nilai penduga dari metode Regresi Komponen Utama, sehingga metode ROBPCA memiliki nilai estimasi yang lebih baik dibandingkan dengan metode Regresi Komponen Utama.

Daftar Pustaka

- [1] Montgomery, D.C. dan Peck, E.A. (1991) *Introduction to Linear Regression Analysis*, 2nd edition, A Wiley-Interscience, New York.
- [2] Neter, J. (1997) *Model Linear Terapan*, Bandung: Diterjemahkan oleh Bambang Sumantri, IPB.
- [3] Notiragayu.2008.*Pembandingan Beberapa Metode Analisis Regresi Komponen Utama Robust*.Prosiding Seminar Hasil Penelitian dan Pengabdian kepada Masyarakat, Universitas Lampung.
- [4] Sunaryo, S. (2011) *Mengatasi Masalah Multikolinearitas dan Outlier dengan Pendekatan ROBPCA (Studi Kasus: Angka Kematian Bayi di Jawa Timur)*, Jurnal Matematika, Saint dan Teknologi, Jurusan Statistika, ITS, vol. 12, Nomor 1, Maret, pp. 1-10.