

**PENERAPAN METODE *LEAST MEDIAN SQUARE-MINIMUM
COVARIANCE DETERMINANT (LMS-MCD)*
DALAM REGRESI KOMPONEN UTAMA**

**I PUTU EKA IRAWAN¹, I KOMANG GDE SUKARSA²,
NI MADE ASIH³**

^{1,2,3}Jurusan Matematika FMIPA Universitas Udayana, Bukit Jimbaran-Bali
e-mail: ¹ekaone17@yahoo.co.id, ²sukarsakomang@yahoo.co.id, ³asihmath77@gmail.com

Abstract

Principal Component Regression is a method to overcome multicollinearity techniques by combining principal component analysis with regression analysis. The calculation of classical principal component analysis is based on the regular covariance matrix. The covariance matrix is optimal if the data originated from a multivariate normal distribution, but is very sensitive to the presence of outliers. Alternatives are used to overcome this problem the method of Least Median Square-Minimum Covariance Determinant (LMS-MCD). The purpose of this research is to conduct a comparison between Principal Component Regression (RKU) and Method of Least Median Square - Minimum Covariance Determinant (LMS-MCD) in dealing with outliers. In this study, Method of Least Median Square - Minimum Covariance Determinant (LMS-MCD) has a bias and mean square error (MSE) is smaller than the parameter RKU. Based on the difference of parameter estimators, still have a test that has a difference of parameter estimators method LMS-MCD greater than RKU method.

Keywords: *Multicollinearity, Outlier, Principal Component Regression, LMS, MCD*

1. Pendahuluan

Multikolinearitas (*multicollinearity*) merupakan suatu kondisi pada regresi linear berganda dengan variable-variabel bebas saling berkorelasi. Regresi Komponen Utama merupakan metode yang mengatasi multikolinearitas dengan cara mengombinasikan teknik analisis komponen utama dengan analisis regresi. Pada analisis komponen utama klasik, perhitungannya didasarkan pada matriks kovarian biasa. Matriks kovarian ini akan optimal apabila data berasal dari suatu distribusi normal multivariat, tetapi sangat sensitif terhadap adanya pencilan (*oulier*) [2]. Alternatif yang digunakan untuk mengatasi masalah ini adalah Metode *Least Median Square-Minimum Covariance Determinant (LMS-MCD)*. Adapun tujuan dalam penelitian ini adalah melihat nilai bias dan MSE parameter dari metode RKU setelah matriks kovarian biasa diganti dengan matriks kovarian

¹ Mahasiswa Jurusan Matematika FMIPA Universitas Udayana

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

robust (LMS-MCD). Pada penelitian ini, setelah matriks kovarian biasa yang digunakan pada RKU diganti dengan matriks kovarian *robust* (LMS-MCD) memiliki nilai bias dan *mean square error* (MSE) parameter lebih kecil dibandingkan dengan RKU. Berdasarkan selisih penduga parameter dari 100 ulangan, masih ada ulangan yang memiliki selisih penduga parameter dari metode LMS-MCD yang lebih besar dari metode RKU.

Metode *Least Median Square* (LMS)

Salah satu metode regresi *robust* yang juga sering digunakan dalam analisis regresi adalah metode LMS (*Least Median Square*). Metode ini mempunyai keuntungan untuk mengurangi pengaruh dari sisaan (*residual*) terhadap keakuratan koefisien regresi. Penduga LMS diperoleh dengan mencari model regresi yang meminimumkan median kuadrat sisaan (e_i^2) dari himpunan-himpunan bagian pada data yang masing-masing sebanyak h , dengan $e_i^2 = (y_i - x_i^T b)^2$ [3]. Nilai h ditentukan dengan menggunakan rumus $h = \left\lfloor \frac{n}{2} \right\rfloor + \left\lfloor \frac{(p+1)}{2} \right\rfloor$, dengan n = ukuran sampel (banyaknya data), p = banyaknya parameter.

Metode *Minimum Covariance Determinant* (MCD)

Pada Metode ini dicari matriks varians-kovarians dari setiap himpunan bagian data yang banyak elemennya h' , dengan $h' = \frac{n+p+1}{2}$. Selanjutnya diantara matriks varian-kovarians tersebut dipilih yang memiliki determinan terkecil, dengan p menyatakan banyak variabel dan n menyatakan banyak pengamatan [1].

2. Metode Penelitian

Data yang digunakan adalah data simulasi yang diperoleh dengan membangkitkan data yang berdistribusi normal dengan menggunakan bantuan program R i386 2.15.2. Penyelesaian Analisis Komponen Utama dan *Least Median Square – Minimum Covariance Determinant* (LMS-MCD) pada data yang dibangkitkan menggunakan software MINITAB 15 dan software R i386 2.15.2. Pada simulasi data ini, banyaknya sampel yang digunakan adalah 100 amatan (dinotasikan $n = 100$) . Sedangkan presentase pencilan yang digunakan adalah 10%, 15%, dan 20%.

Adapun tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Membangkitkan nilai sisaan (e), $X_1 \sim (50,5)$, $X_2 \sim (100,10)$, $X_3 \sim (30,5)$, $X_4 = X_1 + X_3 + e$, Menentukan nilai Y dengan cara $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$. Dimana $\beta_0 = 2$ dan $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$
2. Melakukan regresi komponen utama pada X_1, X_2, X_3, X_4 terhadap Y , sehingga diperoleh nilai penduga yang fit ($\hat{\beta}^{fit}$) dengan menggunakan R i386 2.15.2.
3. Membangkitkan pencilan yang berdistribusi $N(40,0.05)$ sebanyak 10% dari banyak amatan ($n = 100$) dengan menggunakan program R i386 2.15.2. dan

memasukan pencilan pada data simulasi dengan cara menjumlahkan pencilan pada masing-masing variabel bebas yaitu X_1, X_2, X_3

4. Menghitung nilai Y_p dari peubah bebas yang sudah terkontaminasi pencilan. Dengan $Y_p = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$, $\beta_0 = 2$ dan $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$.
5. Melakukan analisis Regresi Komponen Utama pada data yang sudah terkontaminasi pencilan dengan menggunakan software R i386 2.15.2 sehingga diperoleh $\hat{\beta}^{(RKU)p}$
6. Melakukan analisis data menggunakan R i386 2.15.2 dengan menggunakan metode *Least Median Square – Minimum Covariance Determinant* (LMS-MCD) sehingga diperoleh $\hat{\beta}^{(LMS-MCD)p}$.
7. Ulangi langkah 3 sampai 6 untuk presentase 15%, 20%.
8. Menganalisa dan membandingkan hasil perhitungan Regresi Komponen Utama dengan *Least Median Square-Minimum Covariance Determinant* (LMS-MCD) berdasarkan bias parameter dan *Mean Square Error* (MSE).

3. Hasil dan Pembahasan

Teknik perhitungan analisis komponen utama menggunakan matriks kovarian biasa. Matriks kovarian ini akan optimal apabila data berasal dari suatu distribusi normal multivariat. Akan tetapi, matriks kovarian ini sensitif terhadap adanya pencilan. Apabila pada data terdapat pencilan, maka matriks kovarian akan kehilangan efisiensinya, sehingga diperlukan suatu matriks kovarian yang *robust* terhadap pencilan. Penaksir *robust* yang digunakan untuk kovarian adalah *Minimum Covariance Determinant* (MCD). Setelah matriks kovarian biasa pada metode RKU diganti dengan matriks kovarian *robust*, maka diperoleh komponen-komponen utama yang *robust*. Selanjutnya komponen-komponen utama yang *robust* ini diregresikan terhadap peubah tidak bebas dengan menggunakan metode *Least Median Square*.

Berikut hasil perhitungan bias parameter dari metode RKU sebelum matriks kovarian biasa diganti dengan matriks kovarian *robust* dan sesudah matriks kovarian biasa diganti dengan matriks kovarian *robust* (LMS-MCD) yang disajikan dalam tabel 1.

Tabel 1. Hasil Perhitungan Bias Parameter untuk 100 kali Ulangan

Presentase Pencilan	Metode	Intercept	PC1	PC2	PC3	PC4
10%	RKU	20,0055	1,2313	0,5453	0,0308	0,4356
	LMS-MCD	1,7792	0,0119	0,1443	0,0199	0,3892
15%	RKU	30,0025	1,5189	0,6074	0,3046	0,5959
	LMS-MCD	6,7617	0,0883	0,0272	0,0308	0,4804
20%	RKU	40,0035	1,0302	0,2414	1,4092	1,1498
	LMS-MCD	13,1999	0,0324	0,0826	0,0673	0,685

Hasil perhitungan MSE parameter dapat dilihat pada tabel 2.

Tabel 2. Hasil Perhitungan *Mean Square Error* Parameter untuk 100 kali Ulangan

Presentase Pencilan	Metode	Intercept	PC1	PC2	PC3	PC4
10%	RKU	400,222	3,7153	3,2863	0,1961	2,4759
	LMS-MCD	6,3224	1,1183	2,7129	0,0523	1,7137
15%	RKU	900,152	3,6949	3,15	0,4917	3,1602
	LMS-MCD	48,8361	1,761	2,8019	0,0324	2,0185
20%	RKU	1600,28	2,9369	2,916	2,0797	2,7384
	LMS-MCD	176,411	1,8399	2,5413	0,0437	2,1958

Berdasarkan Tabel 1 dan 2, terlihat bahwa pada presentase pencilan 10%, 15%, dan 20% nilai bias parameter dan *mean square error* (MSE) parameter dari metode RKU sebelum matriks kovarian biasa diganti dengan matriks kovarian *robust* selalu lebih besar dibandingkan dengan nilai bias parameter dan *mean square error* (MSE) parameter dari metode RKU sesudah matriks kovarian biasa diganti dengan matriks kovarian *robust* (LMS-MCD).

Selain itu, perbandingan untuk kedua metode juga dilakukan dengan menghitung selisih dari penduga parameter yang dihasilkan oleh kedua metode dengan penduga parameter dari metode RKU pada data tanpa pencilan. Setelah dilakukan pengulangan 100 kali dan dihitung selisih penduga parameter dari masing-masing ulangan pada kedua metode, diperoleh bahwa secara umum metode LMS-MCD lebih bagus daripada metode RKU.

4. Kesimpulan

Berdasarkan hasil yang diperoleh dapat disimpulkan, untuk presentase pencilan 10%, 15%, dan 20% pada data simulasi yang dibuat peubah-peubah bebasnya bermultikolinearitas menunjukkan bahwa metode RKU sesudah matriks kovarian biasa diganti dengan matriks kovarian *robust* (LMS-MCD) menunjukkan hasil yang lebih baik dibandingkan dengan metode RKU sebelum matriks kovarian biasa diganti dengan matriks kovarian *robust*. Untuk penelitian

selanjutnya disarankan menambah jumlah ulangan dalam simulasi dan membandingkan peubah bebas yang dibangkitkan dari sebaran normal yang mengalami multikolinearitas dengan peubah bebas yang dibangkitkan dari sebaran data normal multivariat.

Daftar Pustaka

- [1] Aderlina, D. 2011. Metode Minimum Covariance Determinant pada Analisis Regresi Linear Berganda dengan Kasus Pencilan. *Skripsi*. Jurusan Matematika F MIPA Universitas Padjadjaran.
- [2] Notiragayu. 2008. Perbandingan Beberapa Metode Analisis Regresi Komponen Utama Robust. *Prosiding Seminar Hasil Penelitian dan Pengabdian kepada Masyarakat*, Universitas Lampung.
- [3] Rousseeuw, P.J. 1984. Least Median of Square Regression. *Journal of the American Statistical Association*.