

EFEKTIFITAS METODE *JACKKNIFE* DALAM MENGATASI MULTIKOLINEARITAS DAN PENYIMPANGAN ASUMSI NORMALITAS PADA ANALISIS REGRESI BERGANDA

Muhlasah Novitasari Mara¹, Neva Satyahadewi², Ryan Iskandar³

Jurusan Matematika, FMIPA, Universitas Tanjungpura

¹novee_mara@yahoo.co.id, ²neva.satya@gmail.com, ³ryan_virgo90@yahoo.co.id

Abstrak

Regresi merupakan metode dalam statistik dimana variabel dependen dimodelkan sebagai kombinasi linear dari satu atau lebih variabel independen. Terdapat beberapa asumsi klasik pada analisis regresi berganda diantaranya adalah normalitas. Dampak dari tidak terpenuhinya asumsi normalitas adalah baik uji F maupun uji t serta estimasi nilai variabel dependen menjadi tidak valid. Selain asumsi normalitas, dalam pemakaian analisis regresi ganda masih terdapat satu permasalahan yang perlu mendapat perhatian, yaitu bilamana interkorelasi di antara variabel independen yang ada cukup tinggi (multikolinearitas). Bila variabel independen saling berkorelasi tinggi maka varian estimatornya juga akan meningkat dan dapat menghasilkan kuadrat koefisien korelasi ganda yang signifikan, sekalipun dalam persamaan regresinya masing-masing variabel independen sebenarnya memiliki korelasi yang tidak terlalu tinggi terhadap variabel dependen. Efek lain dari multikolinearitas adalah mengakibatkan penduga parameter regresi menjadi tidak efisien karena mempunyai bias dan varians yang besar. Pada paper ini akan dilakukan simulasi metode *Jackknife* untuk mengetahui efektifitas metode *Jackknife* dalam menduga parameter regresi dengan kasus multikolinearitas sekaligus penyimpangan asumsi normalitas pada analisis regresi berganda. *Jackknife* merupakan teknik nonparametrik dan *resampling* yang bertujuan untuk menaksir parameter regresi. Prinsip metode *Jackknife* ialah menghilangkan satu buah data dan mengulanginya sebanyak jumlah sampel yang ada.

Kata kunci: *Multikolinearitas, Normalitas, Jackknife*

A. PENDAHULUAN

Istilah regresi diperkenalkan pertama kali oleh Francis Galton (1886). Galton menemukan ada kecenderungan bagi orang tua yang tinggi mempunyai anak-anak yang tinggi dan bagi orang tua yang pendek untuk mempunyai anak-anak yang pendek (Gujarati, 1999). Regresi adalah hubungan variabel dependen yang dipengaruhi oleh satu atau lebih dari variabel independen. Regresi dapat digunakan untuk melakukan peramalan nilai suatu variabel berdasarkan variabel lain.

Terdapat empat asumsi klasik pada analisis regresi berganda yakni normalitas, multikolinearitas, homogenitas dan autokorelasi. Pada kenyataannya sering terjadi penyimpangan asumsi tersebut. Bahkan seringkali terjadi penyimpangan dua asumsi secara bersamaan seperti terjadi multikolinearitas dan penyimpangan asumsi normalitas secara bersamaan. Multikolinearitas adalah adanya hubungan linear atau korelasi antar variabel bebas. Efek dari multikolinearitas ini dapat mengakibatkan penduga parameter regresi yang dihasilkan dari analisis regresi linear berganda menjadi tidak efisien karena dapat menyebabkan regresi berganda mempunyai bias dan varians yang besar. Multikolinearitas juga akan menyebabkan hasil-hasil

dugaan menjadi peka terhadap perubahan-perubahan kecil. Selain itu multikolinearitas juga dapat menyebabkan terjadinya perbedaan kesimpulan antara uji statistik F dan uji statistik t (Gujarati, 1999). Sedangkan normalitas adalah kondisi dimana error berdistribusi normal. Dampak dari tidak terpenuhinya asumsi normalitas adalah baik uji statistik F maupun uji statistik t serta estimasi nilai variabel dependen menjadi tidak valid.

Ada beberapa cara untuk mengatasi masalah multikolinearitas dan penyimpangan asumsi normalitas. Salah satu cara mengatasi multikolinearitas adalah dengan metode *Jackknife*. Penyimpangan asumsi normalitas dapat diatasi dengan menghilangkan nilai outlier dari data, melakukan transformasi atau menggunakan alat analisis nonparametrik seperti metode *Jackknife*. Pada paper ini akan dibahas efektifitas metode *Jackknife* dalam menduga parameter regresi dengan kasus multikolinearitas sekaligus penyimpangan asumsi normalitas pada analisis regresi berganda.

B. PEMBAHASAN

Jackknife adalah metode resampling yang diperkenalkan oleh Quenouille (1949) untuk estimasi bias dan Tukey (1958) memperkenalkan *Jackknife* untuk menduga standar deviasi. Prinsip metode *Jackknife* dalam estimasi parameter regresi ialah menghilangkan satu buah data dan mengulanginya sebanyak jumlah sampel data yang ada. Model regresi linear berganda dapat dinyatakan dalam notasi matriks sebagai berikut:

$$Y = X\beta + \varepsilon \tag{1}$$

dengan Y adalah matriks berukuran $N \times 1$ berisi sampel data observasi variabel terikat, X adalah matriks berukuran $N \times k$ berisi data observasi variabel bebas, β adalah matriks dari parameter regresi berukuran $k \times 1$, ε adalah matriks variabel galat acak berukuran $N \times 1$, sehingga persamaan (1) dapat dituliskan kembali sebagai berikut

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{k-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Data *Jackknife* ke-1 diperoleh dengan menghilangkan baris ke 1 dari matriks Y^{*1} dan X^{*1} sehingga diperoleh

$$Y^{*1} = X^{*1}\beta^{*1} + \varepsilon^{*1}$$

$$\begin{bmatrix} y_1^{*1} \\ y_2^{*1} \\ \vdots \\ y_{n-1}^{*1} \end{bmatrix} = \begin{bmatrix} 1 & x_{11}^{*1} & x_{12}^{*1} & \dots & x_{1k}^{*1} \\ 1 & x_{21}^{*1} & x_{22}^{*1} & \dots & x_{2k}^{*1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{(n-1)1}^{*1} & x_{(n-1)2}^{*1} & \dots & x_{(n-1)k}^{*1} \end{bmatrix} \begin{bmatrix} \beta_0^{*1} \\ \beta_1^{*1} \\ \vdots \\ \beta_{k-1}^{*1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1^{*1} \\ \varepsilon_2^{*1} \\ \vdots \\ \varepsilon_{n-1}^{*1} \end{bmatrix}$$

Estimasi parameter β^{*1} pada persamaan (2) dicari menggunakan metode kuadrat terkecil sehingga diperoleh estimator $\hat{\beta}^{*1}$ berikut

$$\hat{\beta}^{*1} = (X^{*1T}X^{*1})^{-1}X^{*1TY^{*1}} \tag{4}$$

Langkah mengambil menghilangkan baris ke i terus dilakukan hingga diperoleh parameter *Jackknife* $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*n}$. Estimasi parameter *Jackknife* β^{**} merupakan rata-rata dari estimator $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \dots, \hat{\beta}^{*n}$

$$\hat{\beta}^{**} = \sum_{i=1}^n \hat{\beta}^{*i} / n$$

Dengan demikian persamaan regresi linear berganda pada persamaan (1) dapat diduga dengan persamaan regresi linear berganda *Jackknife* berikut (Sahinler dan Topuz, 2007):

$$\hat{Y} = X\hat{\beta}^{**} + \varepsilon$$

Pada penelitian ini dilakukan simulasi dengan perulangan sebanyak 100 kali untuk mengkaji efektifitas metode *Jackknife* dalam mengatasi multikolinearitas sekaligus

penyimpangan normalitas. Variabel bebas X_1, X_2, X_3 dibangkitkan sebanyak 20 data dari variabel acak berdistribusi normal. Sedangkan Galat ε dibangkitkan sebanyak 20 data dari variabel acak berdistribusi normal, uniform, weibull dan gamma. Variabel-variabel bebas yang telah dibangkitkan tersebut kemudian dibuat saling berkorelasi menggunakan persamaan (5) sehingga diperoleh variabel bebas baru X_{2br} yang berkorelasi dengan X_1 dan X_{3br} yang berkorelasi dengan X_{2br} (Kusnandar, 2001):

$$X_{ibr} = \rho \times X_{i-1} + \sqrt{(1 - \rho^2)} \times X_i \quad i = 2,3 \quad (5)$$

Koefisien korelasi (ρ) yang digunakan dalam penelitian ini adalah 0,9. Variabel X_1, X_{2br} , dan X_{3br} kemudian digunakan sebagai variabel bebas dalam persamaan (6) berikut

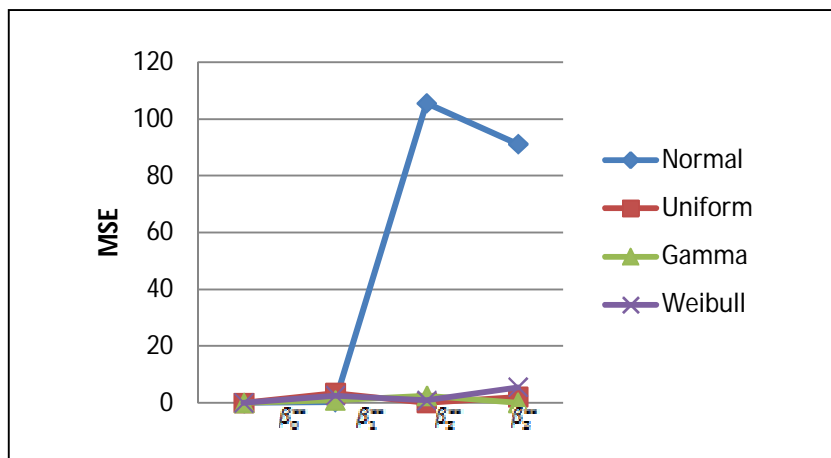
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \quad (6)$$

Nilai-nilai parameter (β) yang digunakan pada persamaan (6) dalam adalah $\beta_0 = 0$, dan $\beta_1 = \beta_2 = \beta_3 = 1$. Luaran nilai bias, standard deviasi dan MSE estimator *Jackknife* dari proses simulasi dapat dilihat pada tabel 1 berikut

Tabel 1. Nilai bias dan standard deviasi estimator *Jackknife*

Distribusi	Nilai Bias				Standar Deviasi				MSE			
	β_0^{**}	β_1^{**}	β_2^{**}	β_3^{**}	β_0^{**}	β_1^{**}	β_2^{**}	β_3^{**}	β_0^{**}	β_1^{**}	β_2^{**}	β_3^{**}
Normal	0.005	0.492	-10.257	9.541	0.017	0.322	0.513	0.296	0.000	0.345	105.471	91.116
Uniform	0.001	-1.861	-0.036	1.456	0.004	0.088	0.135	0.081	0.000	3.471	0.020	2.127
Gamma	-0.003	-0.928	1.566	-0.148	0.010	0.180	0.229	0.173	0.000	0.893	2.504	0.052
Weibull	0.004	-1.597	-0.805	2.349	0.016	0.269	0.450	0.224	0.000	2.621	0.850	5.567

Hasil simulasi yang disajikan pada tabel 1 menunjukkan bahwa pada kondisi terjadi multikolinearitas sekaligus penyimpangan asumsi normalitas estimator *Jackknife* memiliki nilai MSE yang kecil. Namun pada kondisi data terjadi multikolinearitas dan asumsi normalitas terpenuhi estimator *Jackknife* justru besar. Secara visual hal ini dapat dilihat dengan mudah dari gambar 1 berikut



Gambar 1. MSE Estimator *Jackknife*

C. KESIMPULAN

Metode *Jackknife* menghasilkan estimator dengan MSE kecil pada kondisi terjadi multikolinearitas sekaligus penyimpangan asumsi normalitas dalam hal ini *error* berdistribusi Uniform, Gamma dan Weibull. Namun ketika asumsi normalitas terpenuhi, MSE dari estimator *Jackknife* justru besar. Dengan demikian dapat disimpulkan bahwa metode *Jackknife* lebih efektif untuk estimasi parameter regresi pada kondisi terjadi multikolinearitas sekaligus

penyimpangan asumsi normalitas daripada kondisi data terjadi multikolinieritas dengan asumsi normalitas terpenuhi.

D. DAFTAR PUSTAKA

Efron, B.. 1982. *The Jackknife, The Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM. Philadelphia.

Gujarati, D.. 1999. *Ekonometrika Dasar*. Erlangga. Jakarta.

Kusnandar, D.. 2001. *The Identification and Interpretation of Genetic Variation in Forestry Plantation*. The University of Western Australia. Faculty of Agriculture. Australia (Thesis).

Kusnandar, D.. 2004. *Metode Statistik dan Aplikasinya dengan Minitab dan Excel*. Madyan Press. Yogyakarta.

Kutner, M.H., Neter, J., dan Wasserman, W..1997. *Model Linear Terapan*. Sumatri, Bambang (alih bahasa). Jurusan Statistik FMIPA IPB. Bogor.

Sahinler, S., dan Topuz, D.. 2007. Bootstrap and Jackknife Resampling Algorithm For Estimation of Regression Parameters. *Journal of Applied Quantitative Methods*. Vol. 2. No.2.

Sprent, P..1989. *Applied Nonparametric Statistical Methods*. Chapman & Hall. New York.