

PEMODELAN REGRESI ROBUST M-ESTIMATOR DALAM MENANGANI PENCILAN (STUDI KASUS PEMODELAN JUMLAH KEMATIAN IBU NIFAS DI JAWA TENGAH

Alan Prahutama¹, Agus Rusgiyono², Dwi Ispriyanti³, Tiani Wahyu Utami⁴

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro, Semarang

⁴Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Muhammadiyah Semarang

Email: alan.prahutama@gmail.com

ABSTRACT

Regression analysis is statistical method that used to model between predictor variables and response variables. In the regression model, the residual assumed normal distribution, non-autocorrelation, and homoscedasticity. When the assumptions doesn't fulfilled, then the measurement of goodness not well enough. One of the causes may be outlier of data. Handling the outlier can be used robust regression, which one of method is robust M-estimator. In this article, we purposed modelling the number of maternal postpartum in Central Java province with predictor variables are the percentage of pregnant who consumed Fe tablet (X_1), the percentage of household whom applied clean and health lifestyle(X_2), and the percentage of pregnant who First visited to midwife of doctor (X_3) (X_3). In the multiple regression only X_3 was significantly with R-square was 14.25209%, and Mean Square Error (MSE) was 20.4177. Moreover, in outlier detection, there were two outlier in the data, then modelled with Robust M-estimator. The measurement of goodness used R-square of regression robust M-estimator was 21.74% with MSE was 15.02766. Robust M-estimator regression resulted better model than multiple regression to model the number of maternal postpartum in Central Java Province.

Keywords: Multiple linier regression, Robust M-estimator, the number maternal postpartum.

PENDAHULUAN

Analisis regresi merupakan analisis didalam statistika untuk memodelkan hubungan antara variabel respon dengan variabel prediktor. Model regresi dapat digunakan untuk mengetahui signifikansi variabel prediktor terhadap variabel respon. Model regresi berganda dapat dinotasikan sebagai berikut [1]:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (1)$$

Dengan $i = 1, 2, \dots, n$, dengan n menyatakan banyaknya pengamatan dan $j = 1, 2, \dots, k$ dengan k menyatakan banyaknya variabel prediktor. Di dalam analisis regresi, estimasi parameter menggunakan metode *Ordinary Least Square* (OLS) atau kudrat terkecil

dilakukan dengan meminimumkan jumlah kudrat error.

$$J = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \sum_{j=0}^k \beta_j x_{ij})^2 \quad (2)$$

Hasil estimasi parameter model regresi dengan menggunakan OLS dapat ditulis sebagai berikut (dalam bentuk matriks):

$$\hat{\beta} = (X'X)^{-1}X'y \quad (3)$$

Pengujian signifikansi model regresi linier berganda terdiri dari dua yaitu pengujian secara serentak (menggunakan uji F) dan pengujian secara partial (uji t) [2]. Pada model regresi terdapat bebera dari model berdistribusi normal, tidak terjadi autokorelasi dan tidak terjadi heterokedastisitas, selain itu untuk penggunaan regresi linier berganda

diasumsikan juga tidak terjadi multikolinieritas antar variabel prediktor. Pengujian normalitas residual bisa menggunakan uji Kolmogorov Smirnof (KS test), sedangkan pengujian autokorelasi menggunakan Durbin Watson test. Untuk pengujian heterokedastisitas menggunakan uji Gleitser sedangkan pengujian multikolinieritas menggunakan uji VIF (*Variance Inflation Factors*) [3].

Asumsi tersebut harus dipenuhi didalam model regresi agar estimasi model tidak bias. Selain itu untuk melihat performansi dari model regresi dapat dilihat dari koefisien determinasi (R^2) dan *Mean Square Error* (MSE).

Seringkali data yang dianalisis bisa mengandung pencilan/outlier yang bisa menyebabkan asumsi dari model regresi tidak terpenuhi atau mempengaruhi performansi dari model regresi. Untuk menguji outlier pada suatu data dapat menggunakan diagram box plot, pengujian *DFITS* (*Difference in Fit Standardized*) [4]. Untuk menangani pencilan dapat dilakukan beberapa cara, salah satunya adalah menghapus data yang mengandung pencilan. Akan tetapi hal ini tentunya mengakibatkan berkurangnya informasi dari data tersebut, apalagi untuk data dengan pengamatan yang sedikit bisa menimbulkan bias dari model karena berkurangnya informasi pengamatan. Salah metode yang dapat digunakan untuk mengatasi data outlier tanpa menghilangkan data tersebut dapat digunakan regresi Robust. Regresi Robust terdiri dari 5 jenis yaitu regresi *M-estimator*, *S-estimator*, dan *MM-estimator* *Least Median Square (LMS)-estimator*, *Least Trimmed Square (LTS)-estimator* [5].

Setiap metode tersebut mempunyai kelebihan dan kekurangannya masing-masing. Regresi Robust M-estimator mudah digunakan baik secara teoritis ataupun perhitungan komputasinya. Konsep dari regresi Robust M-estimator

adalah mengganti estimasi pada jumlah kuadrat error (persamaan 2) dengan menggunakan persamaan berikut:

$$\min_{\beta} \sum_{i=1}^n \rho(e_i) = \min_{\beta} \sum_{i=1}^n \rho\left(y_i - \sum_{j=0}^k x_{ij}\beta_j\right) \quad (4)$$

dengan ρ merupakan fungsi simetrik dengan kondisi minimum (bernilai 0) [2].

Menurut Stuart (2011) [6], persamaan (4) tidak *equivariant*, sehingga dapat diatasi dengan menyederhanakan error-nya. Dengan estimasi s . Sehingga regresi Robust M-estimator meminimumkan fungsi objektif sebagai berikut:

$$\begin{aligned} \min_{\beta} \sum_{i=1}^n \rho(u_i) &= \min_{\beta} \sum_{i=1}^n \rho\left(\frac{e_i}{s}\right) \\ &= \min_{\beta} \sum_{i=1}^n \rho\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right) \end{aligned} \quad (5)$$

dimana s adalah skala estimasi *robust*. Estimasi s yang digunakan adalah:

$$s = \frac{MAD}{0,6745} = \frac{\text{median}|e_i - \text{median}(e_i)|}{0,6745} \quad \text{dan} \quad u_i = \frac{e_i}{s}$$

Untuk mendapatkan estimator dari model regresi robust M-estimator, dilakukan dengan mencari turunan partial persamaan 5 terhadap β_j sama dengan 0, diperoleh sebagai berikut:

$$\sum_{i=1}^n x_{ij} \psi\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right) = 0 \quad (6)$$

Dengan pembobotan Andrew sebagai berikut:

$$w_i = w(u_i) = \frac{\psi\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right)}{\left(\frac{y_i - \sum_{j=0}^k x_{ij}\beta_j}{s}\right)} \quad (7)$$

Sehingga estimasi model regresi robust M-estimator adalah sebagai berikut:

$$\sum_{i=1}^n x_{ij} w_i \left(y_i - \sum_{j=0}^k x_{ij}\beta_j\right) = 0 \quad (8)$$

Dengan cara proses *Iteratively Reweighted Least Square* (IRLS) didapatkan estimasi model regresi robust M-estimator sebagai berikut [7]:

$$\hat{\beta} = (X'WX)^{-1}X'Wy \quad (9)$$

Pada penelitian ini akan dilakukan pemodelan regresi pada jumlah kematian ibu nifas berdasarkan faktor-faktor yang mempengaruhi. Selain itu dimodelkan menggunakan regresi robust M-estimator.

METODOLOGI PENELITIAN

Pada penelitian ini dilakukan pemodelan regresi terhadap jumlah kematian ibu Nifas di provinsi Jawa Tengah tahun 2018. Kematian ibu nifas merupakan kematian yang terjadi pasca/ setelah melahirkan. Sumber data dari website BPS (Badan Pusat Statistika) provinsi Jawa Tengah. Variabel prediktor yang digunakan antara lain prosentase ibu hamil yang mengkonsumsi tablet Fe (X_1), prosentase rumah tangga yang menerapkan Perilaku Hidup Bersih dan Sehat (PKHBS) (X_2), dan prosentase ibu hamil yang melakukan kunjungan pertama (K1) (X_3) (Dinkes Jateng, 2018) [8]. Pada artikel ini dilakukan pemodelan regresi linier berganda, kemudian pengujian signifikansi model baik serentak maupun partial. Selanjutnya adalah pengujian asumsi model regresi linier berganda dan pengujian outlier. Kemudian dimodelkan menggunakan regresi robust M-estimator.

HASIL DAN PEMBAHASAN

Adapun hasil dan pembahasan pada paper ini meliputi pemodelan regresi linier berganda, pengujian asumsi model regresi dan pendeteksian pencilan dan pemodelan regresi robust M-estimator.

Model regresi linier berganda jumlah kematian ibu nifas di Jawa Tengah

Pada pemodelan regresi berganda jumlah kematian ibu nifas di Jawa Tengah, menggunakan metode stepwise. Hasil

yang didapatkan bahwa, hanya variabel X_3 yaitu prosentase ibu hamil yang melakukan kunjungan pertama (K1) yang berpengaruh signifikan terhadap jumlah kematian ibu Nifas. Adapun hasil estimasi model regresi linier berganda pada pemodelan jumlah kematian ibu nifas adalah sebagai berikut:

$$\hat{y} = 146,093577 - 1,396391x_3$$

Variabel X_3 mempunyai koefisien parameter yang bernilai negatif. Hal ini menunjukkan semakin tinggi variabel X_3 , akan menyebabkan berkurangnya jumlah kematian Ibu Nifas. Pengujian signifikansi parameter adalah sebagai berikut:

Hipotesis

$H_0 : \beta_3 = 0$ (model sesuai atau variabel X_3 berpengaruh terhadap Y)

$H_1 : \text{terdapat } \beta_3 \neq 0$ (model sesuai atau variabel X_3 berpengaruh terhadap Y)

Dikarenakan hanya 1 variabel yang digunakan, maka pengujian signifikansi parameter menggunakan uji F saja. Diperoleh nilai F_{hitung} sebesar 5.484904 dengan kriteria uji H_0 ditolak jika $F_{hitung} > F_{(\alpha,k,n-p)}$ dengan $k = 1$, $n - p = 35 - 2 = 33$ diperoleh nilai dari tabel $F_{(5\%,3,31)} = 4.13$. Sehingga dapat disimpulkan bahwa model sesuai, artinya variabel X_3 berpengaruh terhadap Y. Pada model tersebut diperoleh nilai R-square sebesar 14.25209%, dengan nilai MSE sebesar 20.4177.

Selanjutnya dilakukan pengujian asumsi model. Pengujian asumsi model meliputi Normalitas, Autokorelasi dan Heterokedastisitas. Pengujian normalitas menggunakan Kolmogorov smirnov (KS) Test, pengujian autokorelasi menggunakan uji Durbin Watson (DW) dan pengujian heterokedastisitas menggunakan uji Gleitser. Adapun hasil pengujian disajikan di Tabel 1.

Tabel 1. Pengujian Asumsi Model regresi Linier Berganda

Pengujian	Hasil Pengujian	Keputusan	Kesimpulan
Normalitas	Nilai KS test sebesar 0.10784, dengan nilai KS table sebesar 0.202	H ₀ diterima	Residual berdistribusi Normal
Autokorelasi	Nilai DW test sebesar 1.9291 dengan nilai nilai d_U sebesar 1.5191 dan nilai d_L sebesar 1.4019	H ₀ diterima	Tidak ada Autokorelasi antar residual
Heterokedastisitas	Diperoleh nilai t hitung sebesar -1.169155, dengan nilai t tabel sebesar $t_{(5\%,33)} = 1.6929$.	H ₀ diterima	Variansi residual bersifat homokedastisitas

Berdasarkan tabel 1 tersebut dapat disimpulkan bahwa model regresi jumlah kematian ibu nifas yang dipengaruhi oleh prosentase kunjungan ibu hamil, dapat digunakan. Akan tetapi menghasilkan R-square yang cukup kecil. Langkah selanjutnya adalah pendeteksian adanya pencilan

Pendeteksian Pencilan

Untuk pendeteksian pencilan pada paper ini menggunakan DFFITS test. Terdapat pencilan jika nilai $|DFFITS| > 1$ (Cohen, 2003). Berdasarkan hasil *output*, didapatkan bahwa data ke-4, yaitu Kota Banjarnegara dengan nilai $|DFFITS|$ sebesar 1.290558 dan data ke-15, yaitu Kota Grobogan dengan nilai $|DFFITS|$ sebesar 2.776831 yang lebih besar dari 1. Hal ini mengindikasikan bahwa data tersebut merupakan pencilan.

Regresi Robust M-Estimator

Selanjutnya karena adanya pencilan, dilakukan pemodelan regresi robust M-estimator untuk memodelkan Jumlah kematian Ibu Nifas. Variabel prediktor yang digunakan mengikuti pemodelan regresi berganda yang signifikan.

Estimasi Parameter

Berdasarkan *output* yang dihasilkan, dibutuhkan sebanyak 10 iterasi untuk mendapatkan nilai $\hat{\beta}_j$ yang konvergen. Selain itu, nilai R^2 cenderung meningkat dan nilai MSE cenderung menurun di tiap

iterasinya. Hal ini mengindikasikan bahwa regresi robust M-estimator pembobot *Andrew* yang digunakan mendapatkan model yang lebih baik dibandingkan regresi linier berganda. Model regresi *robust* M-estimator yang didapatkan adalah sebagai berikut:

$$\hat{y} = 169.856932 - 1.636701x_3$$

Sama halnya pada model regresi linier berganda, model regresi robust M-estimator menghasilkan koefisien parameter untuk X_3 bernilai negatif. Hal ini menunjukkan bahwa prosentase ibu hamil yang melakukan kunjungan pertama mempunyai hubungan yang berbanding terbalik dengan jumlah kematian ibu Nifas. Langkah selanjutnya adalah pengujian signifikansi parameter. Pengujian signifikansi parameter hanya dilakukan melalui uji F, karena hanya menggunakan 1 variabel prediktor. Berikut pengujian hipotesisnya

Hipotesis

H₀ : $\beta_3 = 0$ (model regresi robust M-estimator sesuai)

H₁ : terdapat $\beta_3 \neq 0$ (model regresi robust M-estimator tidak sesuai)

Diperoleh nilai F_{hitung} sebesar 9.165186 dengan kriteria uji H₀ ditolak jika $F_{hitung} > F_{(\alpha,k,n-p)}$ dengan $k = 1$, $n - p = 35 - 2 = 33$ diperoleh nilai dari tabel $F_{(5\%,3,31)} = 4.13$. Sehingga dapat disimpulkan bahwa model sesuai, artinya variabel X_3 berpengaruh terhadap Y pada model regresi Robust M-estimator. Pada model tersebut diperoleh nilai R-square

sebesar 21.74% dengan nilai MSE sebesar 15.02766. Hal ini menunjukkan bahwa model regresi robust M-estimator mampu menangani pencilan pada data dan menghasilkan nilai akurasi model yang lebih baik. Hal ini ditunjukkan dengan meningkatnya nilai R-square dan menurunnya nilai MSE. Antara regresi klasik dengan regresi robust memberikan interpretasi model yang sama yaitu hubungan antara variabel prosentase ibu hamil melakukan kunjungan pertama berbanding terbalik dengan jumlah kematian ibu Nifas. Hal ini menunjukkan bahwa meningkatnya kunjungan pertama (K1) pada ibu hamil mampu menekan atau mengurangi jumlah kematian ibu nifas di provinsi Jawa Tengah.

KESIMPULAN

Model regresi linier berganda pada jumlah kematian ibu Nifas dengan variabel yang signifikan adalah sebagai berikut:

$$\hat{y} = 146.093577 - 1.396391x_3$$

Asumsi dalam model regresi tersebut terpenuhi. Sedangkan pada pendeteksian pencilan, data mengandung pencilan sehingga performansi model kurang bagus. Untuk menangani digunakan regresi Robust M-estimator. Diperoleh model sebagai berikut:

$$\hat{y} = 169.856932 - 1.636701x_3$$

Nilai R-square dalam model regresi Robust M-estimator sebesar 21.74% dengan nilai MSE sebesar 15.02766.

Ucapan Terima Kasih

Paper ini merupakan salah satu luaran penelitian tahun 2020 dengan sumber pendanaan Selain APBN Undip untuk riset madya .

DAFTAR PUSTAKA

- [1] Draper, N.R. dan Smith, H. 1992. Analisis Regresi Terapan. Edisi Kedua. Diterjemahkan oleh : Bambang Sumantri. Jakarta : Gramedia Pustaka Utama. Terjemahan dari : Applied Regression Analysis.
- [2] Montgomery, D.C. dan Peck, E.A. 2012. *Introduction To Linier Regression Analysis. Fifth Edition.* New York : John Wiley and Sons, Inc.
- [3] Gujarati, D.N. 1997. *Ekonometri Dasar.* Diterjemahkan oleh : Sumarno Zain. Jakarta : Erlangga. Terjemahan dari : Basic Econometrics.
- [4] Fox, J. 2002. Robust Regression : Appendix to An R and S-Plus Companion to Applied Regression.
- [5] Chen, C. 2002. *Robust Regression and Outlier Detection with The ROBUSTREG Procedure,* paper 265-27. SAS Institute Inc., Lary, NC.
- [6] Stuart, C. 2011. *Robust Regression.* England : Durham University.
- [7] Cohen, J., Cohen, P., West, S.G., dan Aiken, L.S. 2003. *Applied Multiple Regression Correlation Analysis for The Behavioral Sciences.* Third Edition. New Jersey : Lawrence Erlbaum Associates.
- [8] [Dinkes Jateng] Dinas Kesehatan Provinsi Jawa Tengah. 2018. *Profil Kesehatan Provinsi Jawa Tengah Tahun 2017.* <http://dinkesjatengprov.go.id/> (diunduh 30 Desember 2018).