Jurnal Matematika UNAND
Vol. **10** No. **1** Hal. 71 – 78
Edisi Januari 2021
ISSN : 2303–291X
e-ISSN : 2721–9410

REGRESI ROBUST MM-*ESTIMATOR* UNTUK MEMODELKAN JUMLAH KEMATIAN BALITA DI PROVINSI JAWA TIMUR TAHUN 2017

ATIKAH RAHMAH PUTRI, HAZMIRA YOZZA*, FERRA YANUAR
Program Studi S1 Matematika,
Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Andalas,
Kampus UNAND Limau Manis Padang, Indonesia.
email : atikahrahmahputri@gmail.com, hazmirayozza@sci.unand.ac.id,
ferrayanuar@sci.unand.ac.id

Diterima 15 Desember 2020 Direvisi 29 Desember 2020 Dipublikasikan 12 Januari 2021

Abstrak. Dengan berakhirnya MDGs, PBB kembali membuat agenda pembangunan yaitu SDGs dengan salah satu targetnya yaitu mengakhiri kematian balita yang dapat dicegah, dengan seluruh negara menurunkan Angka Kematian Balita 25 per 1000 kelahiran hidup. Provinsi Jawa Timur merupakan salah satu provinsi yang memiliki angka kematian absolut yang masih tinggi yaitu sebanyak 4.464 balita meninggal pertahun. Untuk menganalisis faktor yang mempengaruhi jumlah kematian balita digunakan analisis regresi *robust* MM-*estimator*. MM-*Estimator* merupakan gabungan dari S-*Estimator* dan M-*Estimator* sehingga metode ini mempunyai efisiensi yang tinggi dan nilai resisten yang tinggi pula, sehingga metode MM-*Estimator* merupakan metode yang terbaik dibandingkan metode regresi *robust* lainnya. Hasil analisis menunjukkan bahwa variabel bebas yang mempe- ngaruhi jumlah kematian balita di Provinsi Jawa Timur adalah jumlah kasus *pneumonia* pada balita.

Kata Kunci: Regresi *Robust* MM-*Estimator*, Jumlah Kematian Balita, *Outlier*

1. Pendahuluan

Pada tahun 2000, kepala negara beserta perwakilan 189 negara anggota Perserikatan Bangsa-Bangsa (PBB) menyepakati Deklarasi Millenium atau *Millennium Development Goals* (MDGs). Deklarasi tersebut memiliki delapan target, salah satunya adalah menurunkan 2/3 jumlah kematian anak yang berusia dibawah lima tahun (balita) pada tahun 2015. Berakhirnya *Millenium Development Goals* (MDGs), PBB kembali membuat agenda pembangunan yaitu *Sustainable Development Goals* (SDGs) yang merupakan lanjutan dari MDGs dengan salah satu targetnya yaitu pada tahun 2030, mengakhiri kematian bayi dan anak di bawah lima tahun yang dapat dicegah, dengan seluruh negara menurunkan Angka Kematian Neonatal seti-

*penulis korespondensi

daknya hingga 12 per 1000 kelahiran hidup dan Angka Kematian Balita 25 per 1000 kelahiran hidup. [10]

Provinsi Jawa Timur merupakan salah satu provinsi yang memiliki angka kematian absolut yang masih tinggi yaitu sebanyak 4.464 balita meninggal pertahun [5]. Salah satu cara untuk menurunkan angka kematian absolut balita tersebut adalah dengan mengetahui segala masalah yang terkait dengan faktor yang diperkirakan mempengaruhi angka kematian absolut balita tersebut. Oleh karena itu, perlu dianalisis faktor yang mempengaruhi jumlah kematian balita agar masyarakat dan pemerintah dapat memiliki upaya untuk menurunkan jumlah kematian balita. Salah satu analisis yang dapat dilakukan untuk menganalisis faktor yang mempengaruhi jumlah kematian balita adalah dengan regresi *robust*. Regresi *robust* merupakan estimasi yang lebih efisien dalam menangani *outlier* pada data tanpa harus membuang *outlier* tersebut dalam analisis.

2. Landasan Teori

2.1. *Outlier dan Pengamatan Berpengaruh*

Outlier adalah data yang tidak mengikuti pola umum pada model regresi yang dihasilkan [9]. Adanya *outlier* dapat disebabkan oleh kesalahan ketika memasukkan data, kegagalan menspesifikasi adanya data hilang dalam program komputer, adanya data yang bukan berasal dari populasi yang diambil sebagai sampel, dan variabel dalam populasi memiliki nilai ekstrim serta tidak berdistribusi normal. Akibat dari adanya *outlier* diantaranya variansi pada data menjadi lebih besar, interval dan *range* pada data menjadi lebar, dan *mean* tidak dapat menunjukkan nilai yang sebenarnya. [4]

Salah satu cara untuk mendeteksi *outlier* adalah dengan *R-student*. Perhitungan R-*student* pada pengamatan ke-i dapat dinyatakan secara umum sebagai [8]

$$t_i = \frac{e_i}{s_{(i)}(1 - h_{ii})^{1/2}} \tag{2.1}$$

Jika terdapat *outlier* pada data, analisis dapat dilakukan dengan mengeluarkan *outlier* yang telah teridentifikasi kemudian tetap menggunakan metode kuadrat terkecil dalam menduga koefisien regresi [9]. Namun akan lebih bijaksana untuk tidak mengabaikan *outlier* begitu saja. Adakalanya *outlier* dapat memberikan informasi yang tidak bisa diberikan oleh titik data lainnya. *Outlier* seperti ini biasanya dikategorikan sebagai pengamatan berpengaruh [8].

Pengamatan yang berpengaruh adalah pengamatan yang baik secara individu atau bersama-sama dengan beberapa pengamatan lainnya, memiliki pengaruh yang jauh lebih besar terhadap berbagai nilai estimasi (koefisien, *standar error*, nilai-t, dsb) dibandingkan pengaruh sebagian besar pengamatan lainnya [1]. Pengamatan berpengaruh memiliki dampak pada model koefisien regresi yang dihasilkan karena dalam beberapa kasus, estimasi parameter lebih bergantung pada pengamatan berpengaruh daripada mayoritas data. Oleh karena itu, perlu dilakukan deteksi pengamatan berpengaruh dan menilai dampaknya terhadap model.

Beberapa metode pendeteksian pengamatan berpengaruh antara lain [8]:

(1) *Cooks Distance* digunakan untuk mengidentifikasi apakah suatu pengamatan berpengaruh terhadap semua estimasi koefisien regresi. Perhitungan *Cook's Distance* dapat dinyatakan secara umum dengan

$$D_i = \frac{(\hat{\boldsymbol{\beta}}_{(\boldsymbol{i})} - \hat{\boldsymbol{\beta}})'\mathbf{M}(\boldsymbol{\beta}_{(\boldsymbol{i})} - \hat{\boldsymbol{\beta}})}{c}, i = 1, 2, \cdots, n \tag{2.2}$$

Suatu data pengamatan dikatakan pengamatan berpengaruh terhadap koefisien regresi jika nilai $D_i > 4/n$.

(2) DFFITS digunakan untuk mengetahui apakah suatu pengamatan ke-i berpengaruh terhadap model regresi yang ditinjau dari nilai dugaan y-nya. Perhitungan DFFITS dapat dinyatakan secara umum sebagai

$$DFFITS_i = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}} \tag{2.3}$$

Suatu pengamatan ke-i dikatakan berpengaruh terhadap nilai fitnya apabila pengamatan tersebut memiliki nilai $|DFFITS_i| > 2\sqrt{\frac{p}{n}}$.

(3) DFBETAS digunakan untuk menyatakan pengaruh suatu pengamatan ke-i terhadap koefisien ke-j. DFBETAS dapat dihitung dari

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j \hat{\beta}_{j(i)}}{\sqrt{s^2_{(i)} C_{jj}}} \tag{2.4}$$

Suatu pengamatan ke-i dikatakan berpengaruh terhadap koefisien ke-j apabila pengamatan tersebut memiliki nilai $|DFBETAS_{j,i}| > \frac{2}{\sqrt{n}}$.

(4) COVRATIO adalah suatu ukuran yang menggambarkan pengaruh pengamatan ke-i terhadap ketelitian estimasinya. Nilai COVRATIO dihitung dari

$$COVRATIO_i = \frac{(s^2_{(i)})^p}{(MS_{Res})^p} \frac{1}{1 - h_{ii}} \tag{2.5}$$

Pengamatan ke-i dikatakan sebagai pengamatan berpengaruh terhadap ketelitian estimasinya jika nilai $COVRATIO_i > 1 + \frac{3p}{n}$ atau $COVRATIO_i < 1 - \frac{3p}{n}$.

2.2. *Regresi Robust*

Regresi robust diperkenalkan oleh Andrews pada tahun 1972 dan merupakan solusi alternatif untuk mengestimasi parameter regresi ketika terdapat *outlier* yang mempengaruhi model regresi. Metode ini penting untuk menganalisa data yang dipengaruhi oleh *outlier* dengan hasil model yang resisten terhadap *outlier*. Resisten artinya tidak dipengaruhi oleh perubahan besar pada sebagian kecil data atau perubahan sedikit pada sebagian besar data. [7]

Perbedaan antara metode kuadrat terkecil dengan regresi *robust* yaitu pada metode kuadrat terkecil data diberi bobot yang sama dalam mengestimasi parameter sedangkan pada regresi *robust* data diberi bobot yang berbeda dimana bobot yang lebih kecil diberikan pada data yang terdeteksi sebagai *outlier*. Karena bobot yang diberikan berbeda maka pada pendugaan digunakan metode kuadrat terkecil terboboti (*Weighted Least Square*). Dalam pengaplikasikannya, pendugaan parameter dilakukan dalam beberapa iterasi sehingga dengan kata lain. pendugaan parameter model dilakukan menggunakan IRLS (*Iteratively Reweighted Least Square*).

Beberapa metode *robust* yang dapat digunakan dalam analisis regresi adalah sebagai berikut.

2.2.1. *Robust M-Estimator*

Salah satu cara mengestimasi koefisien parameter pada regresi *robust* adalah M-*Estimator*. M-*Estimator* diperkenalkan oleh Huber pada tahun 1973. Pada umumnya regresi *robust* M-*Estimator* merupakan estimasi yang meminimumkan fungsi ρ (fungsi objektif) yang dapat ditulis sebagai berikut[8]

$$\hat{\beta} = arg\min_{\beta} \sum_{i=1}^{n} \rho(u_i) = arg\min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) \tag{2.6}$$

dengan $u_i = \dfrac{e_i}{s}$ dan s adalah skala estimasi *robust* dengan formula sebagai berikut

$$s = \frac{MAD}{0.6745} = \frac{median|e_i - median(e_i)|}{0.6745}$$

untuk fungsi ρ digunakan fungsi objektif *Tuckey Bisquare* dengan $c = 4.685$ yang dinyatakan sebagai

$$\rho(u_i) = \begin{cases} \frac{c^2}{6}\left[1 - \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2\right], & |u_i| \leqslant c \\ \frac{c^2}{6}, & |u_i| > c \end{cases} \tag{2.7}$$

Fungsi pembobot *Tuckey Bisquare* dinyatakan sebagai berikut

$$w_i = w(u_i) = \frac{\psi(u_i)}{u_i} = \begin{cases} \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & |u_i| \leqslant c \\ 0, & |u_i| > c \end{cases} \tag{2.8}$$

Untuk meminimumkan fungsi ρ pada Persamaan (2.6) digunakan turunan parsial pertama fungsi objektif ρ terhadap β_j dimana $j = 0, 1, \cdots, k$ yang selanjutnya disamakan dengan 0, sehingga diperoleh [6]

$$\sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s}\right) = 0, \tag{2.9}$$

dengan $\psi(u_i) = \rho'(u_i)$ dan x_{ij} adalah pengamatan ke-i pada variabel bebas ke-j dan $x_{i0} = 1$.

Persamaan (2.9) dapat ditulis dalam notasi matriks sebagai berikut

$$\mathbf{X}'\mathbf{W}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{W}\mathbf{Y} \tag{2.10}$$

Persamaan (2.10) diselesaikan secara iteratif dengan *Iteratively Reweight Least Square* yang dinyatakan sebagai berikut

$$\hat{\boldsymbol{\beta}}^{(\boldsymbol{q}+\mathbf{1})} = (\mathbf{X}'\mathbf{W}^{(q)}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{W}^{(q)}\mathbf{Y}). \tag{2.11}$$

Proses iterasi dihentikan jika $\hat{\beta}$ konvergen, dengan kata lain jika $|\hat{\beta}_j^{(q+1)} - \hat{\beta}_j^{(q)}|$ cukup kecil atau sama dengan 0, untuk $j = 0, 1, \cdots, k$. [3]

2.2.2. *Robust S-Estimator*

S-*Estimator* merupakan metode yang diperkenalkan oleh Rousseeuw dan Yohai pada tahun 1984. S-*estimator* didasarkan pada skala residual M-*estimator*. S-*estimator* dapat didefinisikan sebagai [9]:

$$\hat{\beta}_s = \min_{\hat{\beta}} s(e_1, \cdots, e_n), \tag{2.12}$$

yang artinya $\hat{\beta}$ diperoleh dengan menentukan estimator skala *robust* s yang minimum dengan e_i $(i = 1, 2, \cdots, n)$ merupakan residual ke-i dimana:

$$s = \sqrt{\frac{1}{nK} \sum_{i=1}^{n} w_i e_i^2}, \tag{2.13}$$

dengan $K = 0.1995$, w_i adalah fungsi pembobot *Tuckey Biquare* dan estimasi s awal sebagai berikut

$$s = \frac{MAD}{0.6745} = \frac{median|e_i - median(e_i)|}{0.6745}.$$

2.2.3. *Robust MM-Estimator*

MM-*estimator* adalah metode yang pertama kali diperkenalkan oleh Yohai pada tahun 1987. MM-*estimator* merupakan langkah untuk mengestimasi parameter regresi menggunakan S-*estimator* yang meminimumkan *scale* dari galat pada M-*estimator* kemudian dilanjutkan dengan M-*estimator*. MM-*estimator* adalah solusi dari[2]:

$$\hat{\beta} = arg \min_{\beta} \sum_{i=1}^{n} \rho(u_i) = arg \min_{\beta} \sum_{i=1}^{n} \rho \left(\frac{y_i - \sum_{j=0}^{k} x_{ij}\beta_j}{s} \right), \tag{2.14}$$

dimana x_{ij} adalah pengamatan ke-i pada variabel bebas ke-j, s adalah standar deviasi yang diperoleh dari galat dari S-*estimator* dan ρ adalah fungsi objektif *Tuckey Bisquare*.

3. Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data jumlah kematian balita di kabupaten/kota Provinsi Jawa Timur tahun 2017. Data pada penelitian ini didapatkan dari Profil Kesehatan Provinsi Jawa Timur tahun 2017.

Variabel tak bebas dalam penelitian ini adalah jumlah kematian balita di 38 kabupaten/kota di Provinsi Jawa Timur tahun 2017. Variabel bebas yang digunakan dalam penelitian ini adalah jumlah kasus *pneumonia* pada balita (X_1), persentase penanganan komplikasi kebidanan (X_2), persentase berat bayi lahir rendah (BBLR) (X_3), persentase balita balita mendapat vitamin A (X_4), jumlah kasus gizi buruk pada balita (X_5), persentase pemberian imunisasi BCG (X_6), dan persentase pemberian imunisasi Hb (X_7).

Tahap-tahap yang dilakukan adalah:

(1) Mendeskripsikan data yang akan digunakan dalam penelitian.

(2) Analisis data dengan Metode Kuadrat Terkecil.
(3) Melakukan pendeteksian *outlier* dengan R-*student*.
(4) Melakukan pendeteksian Pengamatan berpengaruh dengan *Cook's Distance*, DFFITS, DFBETAS, dan COVRATIO.
(5) Membentuk model regresi *Robust* MM-*Estimator*.
(6) Menganalisis faktor yang mempengaruhi jumlah kematian balita.

4. Pembahasan

4.1. *Pendeteksian Outlier dan Pengamatan Berpengaruh*

Dengan menggunakan metode kuadrat terkecil biasa, diperoleh persamaan regresi:

$$\begin{aligned}\widehat{y} &= 24.933 + 0.0.022X_1 + 0.822X_2 + 4.885X_3 \\ &\quad -1.997X_4 + 0.039X_5 + 0.089X_6 + 0.704X_7,\end{aligned} \tag{4.1}$$

dengan R^2 yang diperoleh sebesar 67.8%. Dengan menggunakan model tersebut, dilakukan pendeteksian *outlier*.

Pada pendeteksian *outlier* dengan R-student diperoleh bahwa pengamatan ke $7, 26$, dan 28 merupakan *outlier* karena memiliki nilai $|t_i| > 2.02453$. Hasil perhitungan R-students dapat dilihat pada tabel berikut.

Tabel 1. Pendeteksian *Outlier*

Data ke	t_i
7	$\mathbf{-3.46182}$
26	**2.22785**
28	$\mathbf{-2.03491}$

Selanjutnya akan dideteksi apakah pengamatan ke $7, 26$ dan 28 merupakan pengamatan berpengaruh yang dapat dilihat pada tabel berikut.

Tabel 2. Pendeteksian Pengamatan Berpengaruh

Data ke	D_i	$DFFITS_i$	$COVRATIO_i$
7	0.54861 *	-2.65796 *	0.03031 *
26	0.53600 *	2.22863 *	0.57526
28	0.15442 *	-1.17702 *	0.51894

Tabel 3. Pendeteksian Pengamatan Berpengaruh Lanjutan

Data ke	$DB_{0,i}$	$DB_{1,i}$	$DB_{2,i}$	$DB_{3,i}$	$DB_{4,i}$	$DB_{5,i}$	$DB_{6,i}$	$DB_{7,i}$
7	0.096	-2.056 *	0.230	-0.225	-0.345 *	0.576 *	0.327 *	1.043 *
26	1.485 *	-0.311	-1.063 *	0.193	-0.933 *	1.254 *	0.193	-1.23 *
28	-0.678 *	0.225	-0.489 *	0.402 *	0.794 *	0.170	0.006	-0.451 *

*pengamatan berpengaruh

Dari Tabel 2 dan Tabel 3 dapat dilihat, pengamatan ke $7, 26$ dan 28 merupakan pengamatan berpengaruh paling tidak terhadap tujuh indikator dari empat metode pendeteksian pengamatan berpengaruh.

4.2. *Analisis Faktor-Faktor yang Mempengaruhi Jumlah Kematian Balita*

Dengan menggunakan metode MM-*Estimator*, diperoleh persamaan regresi:

$$\begin{aligned}\widehat{y} = 4.748348 + 0.028699X_1 + 0.831235X_2 + 5.707692X_3 \\ -1.520916X_4 - 0.003984X_5 - 0.035122X_6 + 0.458624X_7. \qquad (4.2)\end{aligned}$$

Selanjutnya, dilakukan uji signifikansi parameter regresi *robust* untuk semua variabel bebas yang dapat dilihat pada tabel berikut.

Tabel 4. Uji Signifikansi Parameter MM-*Estimator*

Variabel	**MM-*Estimator***
Jumlah Kasus Pneumonia (X_1)	7.844*
Persentase Penanganan Komplikasi Kebidanan (X_2)	1.523
Persentase Berat Bayi Lahir Rendah (X_3)	1.059
Persentase Balita Mendapat Vitamin A (X_4)	-0.646
Jumlah Kasus Gizi Buruk (X_5)	-0.045
Persentase Imunisasi BCG (X_6)	-0.135
Persentase Imunisasi Hb (X_7)	1.176

Variabel bebas dikatakan berpengaruh terhadap variabel tak bebas jika $|t_0| > t_{\alpha/2,n-(k+1)}$ dengan n adalah banyak data yaitu 38, k adalah banyak variabel bebas yaitu 7, dan $\alpha = 0.05$ diperoleh $t_{\alpha/2,n-(k+1)} = t_{0.025,30} = 2.042$. Pada Tabel 4 dapat dilihat variabel bebas X_1 berpengaruh signifikan terhadap variabel tak bebas.

Dari hasil uji signifikansi, selanjutnya dilakukan estimasi parameter pada variabel bebas yang signifikan, diperoleh model

$$\widehat{y} = 27.119 + 0.028X_1, \qquad (4.3)$$

dengan R^2 yang diperoleh sebesar 71.24%.

Berdasarkan model tersebut terlihat bahwa semakin tinggi jumlah kasus *pneumonia* pada balita (X_1) maka juga semakin tinggi jumlah kematian balita. Pada model dapat dilihat bahwa jika presentase kasus *pneumonia* pada balita meningkat 1% maka jumlah kematian akan meningkat sebesar 0.028.

5. Kesimpulan

Berdasarkan hasil penelitian diambil kesimpulan :

(1) Model regresi *robust* MM-*estimator* yang menggambarkan faktor yang mempengaruhi jumlah kematian balita di Provinsi Jawa Timur tahun 2017 adalah

$$Y = 27.119 + 0.028X_1. \qquad (5.1)$$

(2) Berdasarkan regresi *robust* MM-*estimator*, faktor yang mempengaruhi jumlah kematian balita di Jawa Timur pada tahun 2017 adalah *pneumonia* pada balita.

6. Ucapan Terima kasih

Penulis mengucapkan terimakasih kepada Ibu Izzati Rahmi HG M.Si, Ibu Radhiatul Husna M.Si dan Ibu Monika Rianti Helmi M.Si yang telah memberikan masukan dan saran sehingga paper ini dapat diselesaikan dengan baik.

Daftar Pustaka

[1] Belsley, David A., Kuh, Edwin and Welsch, Roy E. 2004. *Regression Diagnostics.* New York: John Wiley and Sons.

[2] Chen, C. 2002. *Robust Regression and Outlier Detection with the ROBUSTREG Procedure. Paper 265-27.* North Carolina: SAS Institute.

[3] Draper, N.R snd Smith, H. 1992. *Applied Regression Analysis, Second Edition.* John Wiley and Sons, Inc. New York.

[4] Ghozali, I. 2009. *Aplikasi Analisis Multivariate dengan Program SPSS.* Semarang : UNDIP.

[5] Kementrian Kesehatan RI. 2018. *Profil Kesehatan Jawa Timur 2017.* Surabaya.

[6] Maronna, R., Douglas Martin and Yohai. 2006. *Robust Statistics.* John Wiley and Sons. England.

[7] Mashitah, A.W., Indriani D. 2013. Metode Regresi Robust Regression On Ordered Statistics pada Data Tersensor Kiri dengan Outlier. *Jurnal Biometrika dan Kependudukan.*

[8] Montgomery, D. C., Peck, E. A. 1992. *Introduction to Linear Regression Analysis, 2nd edition.* John Wiley & Sons, Inc. New York.

[9] Rousseuw, P. J., dan Leroy, A.M. 1987. *Robust Regression and Outlier Detection.* John Wiley dan Sons. New York.

[10] World Health Organizations. 2015. *Health in 2015 : From MDGs to SDGs.* Switzerland.