MATHunesa
Jurnal Ilmiah Matematika *Volume 14 No 02*
e-ISSN : 2716-506X | p-ISSN : 2301-9115 *Tahun 2026*

PEMODELAN PREVALENSI PENYAKIT DIARE BALITA MENGGUNAKAN REGRESI ROBUST DI PROVINSI JAWA TIMUR

Ninda Kartika Putri
Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, Indonesia
e-mail:nindakartika.22020@mhs.unesa.ac.id

Danang Ariyanto
Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya, Surabaya, Indonesia
e-mail:danangariyanto@unesa.ac.id*

Abstrak

Regresi linear berganda merupakan salah satu metode statistik yang digunakan dalam analisis regresi linier berganda untuk memodelkan dan menyelidiki hubungan antar satu variabel dependen dan dua atau lebih variabel independen. Pendugaan parameter pada analisi regresi linier berganda umumnya menggunakan Metode Kuadrat Terkecil (MKT). Namun, metode ini memiliki asumsi kenormalan galat yang seringkali tidak terpenuhi apabila terdapat pencilan (outlier) dalam data. Regresi robust merupakan metode yang efisien untuk menganalisis data yang mengandung outlier. Penelitian ini bertujuan untuk mengetahui adanya outlier yang mengganggu persamaan regresi linear berganda, mengetahui hasil estimasi regresi robust dengan metode Estimator M (Maximum Likelihood Type), Estimator S (Scale Type), dan Estimator LMS (Median Least Square Type) serta mengetahui perbandingan antara ketiga estimasi regresi robust berdasarkan nilai AIC (Akaike Information Criterion). Data yang digunakan dalam penelitian ini, data prevalensi penyakit diare di Provinsi Jawa Timur tahun 2024, data lain meliputi vitamin A, penduduk miskin, posyandu dan sarana air minum. Berdasarkan hasil penelitian dan pembahasan untuk mengatasi outlier dan juga untuk mengetahui perbandingan hasil estimasi M, estimasi S dan estimasi LMS dapat disimpulkan bahwa diperoleh model regresi robust estimasi LMS memiliki nilai AIC terbaik dibandingkan model lainnya.
Kata Kunci: Regresi Robust, Outlier, M-Estimator, S-Estimator, LMS-Estimator

Abstract

Multiple linear regression is a statistical method used in multiple linear regression analysis to model and investigate the relationship between one dependent variable and two or more independent variables. Parameter estimation in multiple linear regression analysis typically uses the. However, this method assumes normality of errors, which is often not met when there are outliers in the data. Robust regression is an efficient method for analyzing data containing outliers. This study aims to identify the presence of outliers that disrupt multiple linear regression equations, to determine the results of robust regression estimation using the M-Estimator (Maximum Likelihood Type), the S-Estimator (Scale Type), and the LMS Estimator (Median Least Squares Type), and to compare the three robust regression estimates based on the AIC (Akaike Information Criterion) value. The data used in this study include data on the prevalence of diarrhea in East Java Province in 2024, as well as other data such as vitamin A, the poor population, posyandu, and drinking water facilities. Based on the research results and discussion regarding the handling of outliers and the comparison of the results of the M-estimator, S-estimator, and LMS-estimator, it can be concluded that the robust regression model using the LMS-estimator has the best AIC value compared to the other models.)
Keywords: *Robust Regression, Outliers, M-Estimator, S-Estimator, LMS-Estimator*

PENDAHULUAN

Diare masih menjadi salah satu masalah kesehatan masyarakat yang serius secara global, terutama di negara berkembang seperti Indonesia. Sejak tahun 2013, diare tercatat sebagai penyebab kematian kedua pada balita setelah Infeksi Saluran Pernapasan Akut. Setiap tahunnya, terdapat 1,7 miliar kasus penyakit diare pada anak yang menyebabkan kematian sekitar 443.832 anak dibawah usia 5 tahun. Selain itu, diare juga menjadi salah satu penyebab utama malnutrisi, yang membuat anak rentan terhadap penyakit infeksi lainnya di masa mendatang. Di Indonesia, berdasarkan data Kementrian Kesehatan Republik Indonesia tahun 2022, jumlah penderita diare pada

semua kelompok umur tercatat sebanyak 7.350.708 kasus dengan cakupan pelayanan kesehatan sebanyak 2.473.081 kasus (33,6%). Pada kelompok balita, jumlah kasus diare mencapai 3.690.984 kasus, namun hanya 879.596 kasus (23,8%) yang memperoleh pelayanan kesehatan.

Provinsi Jawa Timur merupakan salah satu provinsi dengan populasi terbesar di Indonesia. Besarnya populasi ini berdampak pada berbagai aspek kesehatan masyarakat, termasuk tingginya angka kejadian diare. Prevalensi diare pada balita di Jawa Timur sebesar 51,6% yang menduduki peringkat ketiga di Indonesia.

Implementasi Sanitasi Total Berbasis Masyarakat (STBM) juga menunjukkan variasi yang signifikan dengan tidak meratanya cakupan program di berbagai Kabupaten/Kota. Tingginya kejadian diare pada balita dipengaruhi oleh berbagai faktor, di antaranya status gizi dan kondisi sosial ekonomi serta lingkungan kesehatan. Balita dengan status gizi buruk memiliki daya tahan tubuh lemah sehingga rentan terhadap infeks termasuk infeksi saluran pencernaan yang menyebabkan diare. Anak dengan gizi buruk memiliki risiko lebih besar terinfeksi diare, karena asupan nutrisi yang tidak mncukupi sehingga dapat menyebabkan penurunan daya tahan tubuh. Faktor kemiskinan juga menjadi kendala utama karena keterbatasan akses terhadap layanan kesehatan serta penerapan perilaku hidup bersih dan sehat. Upaya pencegahan diare dapat dilakukan melalui pemberian vitamin A, yang terbukti mampu menurunkan kejadian diare dan Infeksi Saluran Pernafasan Akut (ISPA). Selain itu, posyandu memiliki peran penting dalam upaya pencegahan dan pemantauan kesehatan balita, penyakit ini termasuk dalam lima prioritas utama layanan posyandu. Faktor lingkungan, seperti ketersediaan sarana air minum sesuai standar aman juga berperan penting dalam menurunkan resiko terjadinya diare.

Untuk menganalisis faktor yang mempengaruhi diare pada balita, hubungan antarvariabel dianalisis secara kuantitatif menggunakan metode statistika yang mampu menjelaskan keterkaitan antara variabel yaitu analisis regresi. Analisis regresi dapat digunakan untuk menjelaskan hubungan antara variabel terikat dan satu atau lebih variabel bebas. Salah satu metode yang umum digunakan untuk mengestimasi parameter model regresi adalah Metode Kuadrat Terkecil yang meminimumkan jumlah kuadrat galat. MKT memberikan hasil pendugaan parameter yang efisien apabila asumsi-asumsi model regresi terpenuhi. Namun, dalam praktiknya data empiris sering kali tidak sepenuhnya memenuhi asumsi tersebut. Salah satu permasalahan yang sering dijumpai adalah adanya pencilan atau nilai ekstrem pada data pengamatan yang penyimpangan distribusi galat dari asumsi normal serta memengaruhi kestabilan hasil estimasi parameter.

Dalam konteks analisis regresi, *outlier* merupakan pengamatan yang memiliki karakteristik berbeda dari data pengamatan lainnya. Keberadaan *outlier* tidak selalu memberikan dampak yang signifikan terhadap model karena terdapat *outlier* yang tidak berpengaruh terhadap hasil estimasi. Namun, apabila *outlier* bersifat berpengaruh, maka dapat mengubah struktur model regresi dan menyebabkan estimasi parameter menjadi bias. Oleh karena itu, untuk menangani data yang mengandung *outlier* sehingga diperlukan pendekatan regresi *robust*. Metode ini di rancang untuk menghasilkan pendugaan parameter yang lebih stabil dan tidak sensitif terhadap keberadaan pencilan.

Dengan demikian, penelitian ini menerapkan metode regresi *robust* guna menghasilkan estimasi parameter yang lebih stabil dan tidak sensitif terhadap pencilan. Sehingga regresi *robust* menjadi alternatif metode analisis yang tepat untuk mengatasi permasalahan *outlier* dalam data yang sering ditemui dalam penelitian dan tidak selalu disebabkan oleh kesalahan pengukuran. Terdapat beberapa metode yang dapat digunakan untuk mengidentifikasi *outlier* antara lain *scatterplot, boxplot, leverage values*, DfFITS (*Difference in Fit Statndarized*), *cook's distance* dan DfBETA(s) (*Difference in Beta*), dan *internal studentization* yang memiliki perhitungan hampir sama dengan R-student.

KAJIAN TEORI

Analisis Deskriptif

Analisis deskriptif adalah metode yang digunakan untuk pengumpulan atau menyajikan

data penelitian sehingga memberikan gambaran awal mengenai karakteristik setiap variabel (Ghozali, 2018). Dalam penelitian ini, analisis deskriptif digunakan untuk melihat sebaran data prevalensi penyakit diare balita serta variabel-variabel yang diduga memengaruhinya. Statistik seperti nilai minimum, maksimum, rata-rata, median, kuartil dan standar deviasi diperlukan untuk mengidentifikasi pola distribusi data dan mendeteksi adanya nilai ekstrem atau *outlier*. Hal tersebut menjadi bagian penting dalam proses analisis model menggunakan regresi *robust*. Dengan demikian, analisis deskriptif berperan sebagai langkah awal sebelum dilakukan pemodelan regresi. Berikut adalah rumus yang digunakan:

Nilai rata-rata: $\mu = \frac{1}{N} \sum_{i=1}^{n} x_i$

Nilai kuartil pertama (Data ke-): $Q_1 \frac{1(n+1)}{4}$

Nilai kuartil kedua (Data ke-): $Q_2 \frac{2(n+1)}{4}$

Nilai kuartil ketiga (Data ke-): $Q_3 \frac{3(n+1)}{4}$

Standar deviasi: $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}}$

di mana:

μ: nilai rata-rata
x_i: data ke-i
n: banyaknya data x_i
Q: nilai kuartil
s: standar deviasi

Regresi Linier Berganda

Regresi linier berganda adalah satu cara prediksi yang menggunakan persamaan garis lurus untuk menjelaskan hubungan antara variabel terikat dan satu atau lebih variabel bebas (Algifari, 2000). Terdapat dua jenis persamaan untuk menyatakan hubungan antara variabel independen *X* dengan variabel dependen *Y* yaitu analisis linear sederhana dan analisis regresi linear berganda. Walpole dkk., (2011) mengatakan bahwa untuk menyatakan hubungan *Y* dan *X* menggunakan regresi linear berganda dengan *k* variabel independen dapat dinyatakan dalam persamaan berikut:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + e_i \qquad i = 1,2,\ldots,n$$

di mana:

Y_i: variabel dependen ke-*i*
$\beta_0, \beta_1, \ldots, \beta_k$: parameter regresi
x_{ki}: variabel independen dengan *k = 1,2,…, j*
e_i: variabel galat/ residual
n: banyaknya data observasi

Asumsi Regresi Linier Berganda

Metode regresi linier berganda terdapat beberapa asumsi yang harus dipenuhi, asumsi tersebut antara lain:

1. Nilai rata-rata kesalahan pengganggu nol, yaitu E (ε_i) = 0, untuk $i = 1,2,\ldots,n$
2. Varian (ε_i) = E $(\varepsilon_i^2) = \sigma^2$
3. Tidak ada autokorelasi antara kesalahan pengganggu (galat/*error*), kovarian $(\varepsilon_i, \varepsilon_j)$ = 0, $i \neq j$
4. Variabel bebas $x_1, x_2, \ldots, x_k$) konstanta dalam sampling yang terulang dan bebas terhadap kesalahan pengganggu ε_i
5. Tidak ada multikolinieritas variabel bebas *X*
6. $\varepsilon_i \sim N(0; \sigma^2)$, artinya kesalahan pengganggu mengikuti distribusi normal dengan rata-rata 0 dan varian σ^2 dalam regresi linier berganda, proses pendugaan parameter dilakukan dengan menggunaan Metode Kuadrat Terkecil (Kutner dkk., 2004).

Metode Kuadrat Terkecil

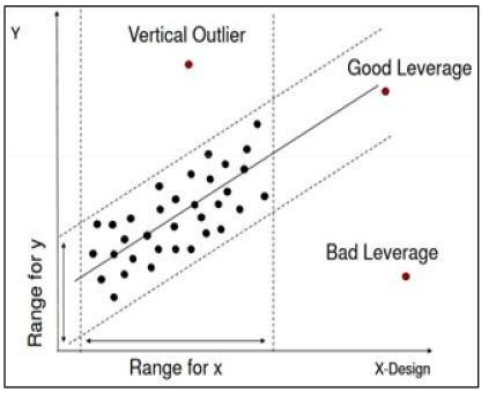
Salah satu metode pendugaan parameter regresi linier berganda adalah Metode Kuadrat Terkecil (MKT), yaitu dengan cara meminimumkan Jumlah Kuadrat Sisaan:

$$JKS = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \cdots - \beta_k x_{ik})^2$$

Data *Outlier*

Outlier menyebabkan hal-hal berikut (Soemartini, 2007):

1. Residual yang besar dari model yang terbentuk
2. Varians pada data tersebut menjadi besar
3. Taksiran interval memiliki rentang yang hebat



Gambar 1. Tipe Data Outlier

Sumber: (*Verardi*, 2008)

Data *outlier* harus dilihat terhadap posisi dan sebaran data yang lainnya. Menurut (Soemartini, 2007) terdapat beberapa metode untuk menentukan batasan *outlier* dalam sebuah analisis yaitu (1) *Scatterplot*, (2) *Boxplot*, (3) *Leverage values*, *DfFITS* (*Difference in Fit Statndarized*), *Cook's Distance*, dan *DfBETA(s)* (*Difference in Beta*), (4) *Internal Studentization* yang memiliki perhitungan hampir sama dengan *R-student*. Pada penelitian ini akan menggunakan ketiga metode dengan lebih terperinci, yaitu (1) *Scatterplot*, (2) *Boxplot*, (3) *Leverage Values* untuk menentukan batasan *outlier*.

1. *Scatterplot*

Scatterplot digunakan untuk mendeteksi suatu data yang mngandung pencilan atau tidak dengan melakukan plot antara data dengan observasi ke-i ($i = 1,2, \dots n$). Selain itu, jika sudah diperoleh model regresi maka dapat dilakukan dengan plot antara residual (e) dengan nilai prediksi $Y(\hat{Y})$. Jika terdapat satu atau beberapa data yang terletak jauh dari pola kumpulan data keseluruhan, maka hal ini mengindikasikan adanya pencilan dalam data (Soemartini, 2007).

2. *Boxplot*

Boxplot digunakan untuk mendeteksi pencilan dalam suatu data yang mempergunakan nilai kuartil dan jangkauan untuk mendeteksi *outlier*. Kuartil 1,2, dan 3 akan membagi data yang telah diurutkan sebelumnya menjadi empat bagian. Jangkauan IQR (*Interquartile Range*) didefinisikan sebagai selisih kuartil 1 terhadap 3 (Soemartini, 2007).

Boxplot mempunyai komponen diantaranya adalah:

1) Nilai observasi terkecil
2) Kuartil terendah atau kuartil pertama (Q_1), yang memotong 25% dari data terendah
3) Median (Q_2) atau nilai pertengahan
4) Kuartil tertinggi atau kuartil ketiga (Q_3), yang memotong 25% dari data tertinggi
5) Nilai observasi terbesar

Suatu nilai dikatakan *outlier* jika Q_3+(1,5 x IQR) < outlier ≤ Q_3+(3 x IQR) atau Q_1 – (1,5 x IQR) > *outlier* ≥ Q_1 – (3 x IQR). Suatu nilai dikatakan ekstrem jika lebih besar dari Q_3 + (3 x IQR) atau lebih kecil dari Q_1 – (3 x IQR)

3. *Leverage Values*

Metode ini mengukur pengaruh suatu observasi terhadap besarnya estimasi parameter antara lain dapat dilihat dari jarak nilai *X* semua observasi Menurut (Wijaya, 2009). Nilai *leverage* dapat ditentukan sebagai berikut: *leverage* $(h_{ii}) = \frac{1}{2} + \frac{(x_i - \hat{x})^2}{(n-1)S_x^2}$

di mana:

h_{ii} : *leverage* kasus ke$-i$

n : banyaknya data

x_i : nilai untuk kasus ke$-i$

$\hat{x}$: *mean* dari *X*

S_x^2 : kuadrat n kasus dari simpangan x_i terhadap mean

Regresi Robust

Konsep pertama kali dipopulerkan oleh *Sir Francis Galton* pada akhir abad ke-19 dalam jurnal yang berjudul "*Regression towards mediocrity in hereditary stature*", yang dimuat dalam *Journal of the Anthropological Institute*, volume 15, hal 246-263,

tahun 1885 (Galton, 1885). Regresi *robust* merupakan metode untuk menganalisa data yang mengandung *outlier* sehingga dihasilkan model yang bersifat *robust* atau tahan (*resistance)* terhadap pengaruh *outlier.* Berbeda denggan regresi linier biasa menggunakan metode *Ordinary Least Squares* yang sensitif terhadap *outlier* karena meminimalkan jumlah kuadrat residual, regresi *robust* menggunakan metode estimasi yang mampu mengurangi pengaruh pengamatan ekstrem terhadap hasil estimasi parameter. Secara umum, bentuk model regresi *robust* sama dengan regresi linier biasa, yang dapat dinyatakan dalam persamaan berikut:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$$
$$i = 1,2,\ldots,n$$

Estimasi yang *resistan*, artinya tidak terlalu terpengaruh oleh *outlier,* baik berupa perubahan besar di sebagian kecil data maupun perubahan kecil pada sebagian besar data (Widodo dan Dewayanti, 2016). Meskipun tahan terhadap *outlier*, prosedur *robust* tetap memiliki efisiensi yang cukup tinggi sekitar 90%-95% pada data berdistribusi normal. Regresi *robust* ini dikembangkan oleh *Rousseeuw dan Leroy* pada tahun 1987 sebagai pendekatan sistematis untuk mendeteksi dan menangani data *outlier*. Metode ini dirancang untuk menangani pencilan (*outlier*) tanpa memerlukan identifikasi manual. Analisis regresi *robust* ini tidak membuat galat model menjadi normal namun model yang dihasilkan oleh metode ini memiliki tingkat keakuratan yang lebih tinggi dari model yang dihasilkan oleh Metode Kuadrat Terkecil (Romdi dkk., 2015). Estimasi parameter dengan regresi *robust* dapat dibentuk persamaan sebagai berikut (Chen, 2002):

$$\boldsymbol{\beta_{(n)robust} = (X'WX)^{-1}X'WY}$$

di mana:

Y: variabel tak bebas

X: variabel bebas

$\beta_{(n)robust}$: koefisien regresi robust (*i=1,2,…,n*)

W: matriks pembobot

Menurut (Chen, 2002) metode estimasi regresi diantaranya:

1. Estimasi *Least Median Squares* (LMS) merupakan metode dengan *high breakdown point* hingga 50%.
2. Estimasi *Least Trimmed Squares* (LTS) merupakan metode dengan *high breakdown point* hingga 50%.
3. Estimasi *Method of Moment* (MM) merupakan metode yang menggabungkan estimasi S dan estimasi M.
4. Estimasi *Scale* (S) merupakan metode dengan *high breakdown point* hingga 50%. Dan memiliki nilai *breakdown point* yang sama dengan estimasi LTS.
5. Estimasi *Maximmum Likelihood Type* (M) merupakan metode yang sederhana, baik dalam perhitungan maupun secara teoritis. Estimasi ini menganalisis data dengan mengasumsikan bahwa Sebagian besar yang terdeteksi pencilan adalah variabel independen. Metode ini memiliki nilai *breakdown point* sebesar $\frac{1}{n}$.

Estimasi Parameter Menggunakan Estimator M (*Maximum Likelikood Type*)

Menurut (Susanti dkk., 2013) tahapan dalam regresi *robust* menggunakan estimasi M sebagai berikut:

1. Menghitung estimasi parameter β menggunakan MKT sehingga didapatkan $\hat{y}_i$
2. Menghitung nilai residual dengan $e_i = y_i - \hat{y}_i$
3. Menghitung Nilai $\hat{\sigma}$
$$\hat{\sigma} = \frac{MAD}{0{,}6745} = \frac{median|e_i - median(e_i)|}{0{,}6745}$$
4. Menghitung nilai skala residual (u_i)
$$u_i = \frac{(y_i - \hat{y}_i)}{\hat{\sigma}} = \frac{e_i}{\hat{\sigma}}$$
5. Menghitung nilai fungsi pembobot $w_i = w(u_i)$ dengan fungsi pembobot *Huber* konstanta yang digunakan adalah *c* = 1.345,
$$w(u_i) = \begin{cases} 1, & |u_i| \le c \\ \dfrac{c}{|u_i|}, & |u_i| > c \end{cases}$$
6. Mengestimasi nilai $\hat{\beta}_m$ menggunakan bobot w_i sehingga diperoleh estimasi M satu tahap. Pada setiap iterasi ke-t dihitung residual $e_i^{(t-1)}$ dan menggunakan pembobot $w_i^{(t-1)} = w(u_i^{(t-1)})$ dari iterasi sebelumnya sehingga didapatkan estimasi parameter $\hat{\beta}_m$ yang baru
7. Melakukan langkah 2 sampai langkah hingga didapatkan estimasi parameter $\hat{\beta}_m = (X'WX)^{-1}X'WY$ yang konvergen.
8. Melakukan langkah 2 sampai 5 hingga didapatkan β_{robust} yang konvergen.

Estimasi Parameter Menggunakan Estimator S (*Scale Type*)

Menurut (Susanti dkk., 2013) Prosedur pendugaan parameter dengan penduga S adalah sebagai berikut:

1. Mengestimasi koefisien regresi menggunakan MKT
2. Menghitung nilai residual dengan $e_i = y_i - \hat{y}_i$
3. Menghitung nilai estimasi skala robust standar deviasi sisaan $\hat{\sigma}_s$

$$\hat{\sigma}_s = \sqrt{\frac{n\sum_{i=1}^{n}(e_i^2) - \left(\sum_{i=1}^{n} e_i\right)^2}{n(n-1)}}$$

4. Menghitung nilai skala residual (u_i)

$$u_i = \frac{(y_i - \hat{y}_i)}{\hat{\sigma}} = \frac{e_i}{\hat{\sigma}}$$

5. Menghitung nilai fungsi pembobot w_i berdasarkan persamaan

$$w_i(u_i) = \begin{cases} \left(\left(1 - \left(\frac{u_i}{c}\right)^2\right)\right)^2 ; & |u_i| \le c \\ 0, & ; \quad |u_i| > c \end{cases}$$

6. Nilai c = 1,547 adalah suatu konstanta yang ditetapkan untuk memberikan efisiensi estimasi sebesar 95% pada fungsi pembobot Tukey Bisquare
7. Mengestimasi nilai $\hat{\beta}_s$
8. Melakukan langkah 2 sampai 6 sehingga diperoleh nilai
   $\widehat{\boldsymbol{\beta}}_{\boldsymbol{s}} = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY}$ yang konvergen.

Estimasi Parameter Menggunakan Estimator LMS (*Least Median Squares Type*)

Prosedur pendugaan parameter dengan penduga LMS adalah sebagai berikut:

1. Mengambil *m* himpunan bagian contoh berdasarkan kombinasi banyaknya *n* pengamatan dan parameter *k+1*

$$m =_n C_{k+1} = \frac{n!}{(k+1)!\,(n-(k+1))!}$$

2. Membentuk model regresi untuk himpunan bagian J_z di mana J_z merupakan persamaan model pada z, dengan z: 1,2,…, m. Persamaan model J_z adalah
   $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + e_i$
3. Menduga parameter β dengan MKT
4. Mengevaluasi model regresi yang menggunakan intersep pada persamaan (2.6) untuk setiap himpunan bagian contoh yang terbentuk dengan prosedur:
   a. Menghitung penduga parameter β dengan MKT
   b. Menghitung nilai penduga galat tanpa menggunakan β_0
   c. Mengurutkan nilai e_i^* sehingga $e_i^* \le e_{(2)}^* \le \cdots \le e_{(n)}^*$
   d. Membentuk kelas interval dari e_i^* masing-masing kelas interval berisi h pengamatan sebesar $h = \left[\frac{n}{2}\right] + \left[\frac{\rho+1}{2}\right]$
   e. Menghitung nilai intersep baru $\beta_0^* = \frac{1}{2}$ (jumlah batas atas dan batas bawah dari lebar kelas interval terkecil)
   f. Menyatakan nilai penduga koefisien regresi dengan nilai intersep yang baru, sehingga didapat persamaan regresi yang baru untuk setiap subset contoh.
5. Menentukan penduga LMS, $\hat{\beta}_{LMS}$ berdasarkan nilai $Q_z = Q_{LMS}$ yang minimum untuk setiap $J_z, z: 1,2,\ldots,m$.

Fungsi Pembobot Robust

Fungsi obyektif merupakan fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust* Fungsi pembobot dibagi menjadi dua yaitu fungsi pembobot huber

$$w_i = w(\boldsymbol{u_i}) = \frac{\psi(\boldsymbol{u_i})}{\boldsymbol{u_i}} = \begin{cases} 1, & |u_i| \le c \\ \dfrac{c}{|u_i|}, & |u_i| > c \end{cases}$$

dan Fungsi Pembobot Tukey bisquare.

$$\psi(u_i) \quad \rho'(u_i)\frac{\partial(\rho(u_i))}{\partial(u_i)} \begin{cases} u_i - \left[1 - \left(\frac{u_i^2}{c}\right)\right]^2, & |u_i| \le c \\ 0, & |u_i| > c \end{cases} \quad 3.2.1$$
$$(2.33)$$

Breakdown Point

Dalam analisis data, diperlukan suatu ukuran untuk menilai ketahanan suatu estimator terhadap pengaruh pencilan atau kontaminasi data yang dapat dijelaskan melalui ukuran yang disebut *Breakdown point*. *Breakdown point* didefinisikan sebagai presentase dari *outlier* yang dapat menyebabkan nilai estimator menjadi sangat besar. Berdasarkan definisi tersebut, pada kasus univariat median memiliki breakdown point sebesar 50%, sedangkan mean memiliki *breakdown point* sebesar 0. Nilai *breakdown poin*t menunjukkan seberapa kuat estimator dalam menghadapi data ekstrem.

Pengujian Signifikansi Parameter

1. Pengujian Signifikansi Simultan

Menurut Ratnasari (2017) uji simultan dilakukan untuk mengetahui apakah variabel independen secara simultan berpengaruh terhadap variabel dependen. Statistik uji didefinisikan sebagai berikut:

$$\mathrm{F_{hitung}} = \frac{\mathrm{SSR}/(\mathrm{p}-1)}{\mathrm{SSE}/(\mathrm{N}-\mathrm{P})}$$

$SSR = \sum(\hat{Y}_i - \widehat{Y})^2$ dan $SSE = \sum(Y_i - \hat{Y}_i)^2$

Jika $F > X^2_{\alpha,J}$ atau *p-value* < 0.05 maka H_0 ditolak, sehingga dapat disimpulkan bahwa variabel independen *X* pengaruh simultan terhadap variabel dependen *Y*. Jika $F \leq X^2_{\alpha,J}$ atau *p-value* ≥ 0.05 berarti H_0 diterima, sehingga dapat disimpulkan bahwa variabel independen *X* tidak terdapat pengaruh simultan terhadap variabel dependen *Y*.

2. Pengujian Signifikansi Parsial

Pengujian ini dilakukan jika pada pengujian parameter secara keseluruhan, terdapat minimal satu variabel berpengaruh terhadap model. Pengujian ini melihat hasil dari Wald Test

Statistik uji yang digunkan adalah Uji Wald didefinisikan sebagai berikut:

$$W_i = \frac{\beta_i}{SE(\beta_i)}$$

Jika thitung > ttabel ,dan p-value 0.05 maka H0 ditolak, sehingga dapat disimpulkan bahwa variabel independen X berpengaruh terhadap variabel dependen Y. Jika thitung < ttabel dan p-value 0.05 maka H0 diterima, sehingga dapat disimpulkan bahwa variabel X tidak berpengaruh terhadap variabel dependen Y.

***Akaike information criterion* (AIC)**

$$AIC = \left(\frac{2p}{n}\right) + ln\left(\frac{JKG}{n}\right)$$

di mana :

p: banyak parameter

n: banyak pengamatan

dalam membandingkan dua atau lebih model, model dengan AIC lebih rendah dinilai lebih baik.

METODE

Teknik pengumpulan data dalam penelitian ini menggunakan data sekunder. yang diperoleh dari dua website yang berbeda, website resmi Dinas Kesehatan Jawa Timur diperoleh 38 data Kabupaten/Kota terkait penyakit diare pada balita. Sementara itu website resmi Badan Pusat Statistik Jawa Timur diperoleh 38 data Kabupaten/Kota yaitu prevalensi diare balita. Data tersebut tersedia dalam publikasi statistik tahunan, serta data yang diambil hanya tahun 2024.

Rancangan Penelitian

Prosedur penelitian untuk mencapai tujuan penelitian yang ditetapkan,yaitu:

a. Melakukan studi literatur
   Pada tahap ini yang sedang dilakukan adalah tinjauan pustaka, atau menemukan penjelasan tentang metode yang digunakan dalam bacaan sebelumnya sehingga dapat mendukung metode regresi *robust* untuk pemodelan penyakit diare pada balita di Jawa Timur tahun 2024.
b. Pengumpulan data dan analisis deskriptif
   Analisis deskriptif merupakan metode yang digunakan untuk mengolah, menyajikan dan mendeskripsikan data sehingga menghasilkan informasi yang relevan. Data diperoleh dari website resmi Badan Pusat Statistik Jawa Timur dan Dinas Kesehatan Jawa Timur.
c. Menganalisis identifikasi *outlier*
   Terlepas dari apakah data mengandung *outlier,* evaluasi akan dilakukan. Dalam ulasan ini, ketiga tersebut akan digunakan secara lebih rinci, yaitu:
   1) Mengamati pola hubungan antara variabel dependen dan variabel independen menggunakan *scatterplot*
   2) Mendeteksi *outlier* pada variabel dependen menggunakan *boxplot*
   3) Mengidentifikasi pengamatan yang memiliki pengaruh tinggi berdasarkan variabel independen menggunakan *leverage value*
d. Menganalisis menggunakan metode regresi *robust*
   Metode regresi *robust* digunakan untuk mengurangi pengaruh residual ekstrem melalui pembobotan. Semua tahapan yang dilakukan sebelumnya merupakan persyaratan untuk mendapatkan model penyakit diare pada balita di Jawa Timur tahun 2024. Penelitian ini menggunakan metode regresi yang tahan terhadap *outlier* dengan tiga estimasi, yaitu *M-Estimator*, *S-Estimator*, dan *LMS-Estimator* yang memiliki karakteristik *breakdown point* berbeda. M-estimator memiliki *breakdown point* yang relatif rendah. S-Estimator memiliki *breakdown point* yang tinggi mendekati 50%, sehingga lebih tahan terhadap pencilan dan LMS-Estimator memiliki *breakdown point* sangat tinggi sebesar 50%.

e. Menganalisis validitas menggunakan uji signifikan parameter

Tahap ini bertujuan untuk mengidentifikasi pengaruh variabel independen terhadap variabel dependen, baik secara simultan maupun parsial. Berdasarkan parameter yang signifikan, dilakukan beberapa tahapan pengujian sebagai berikut:

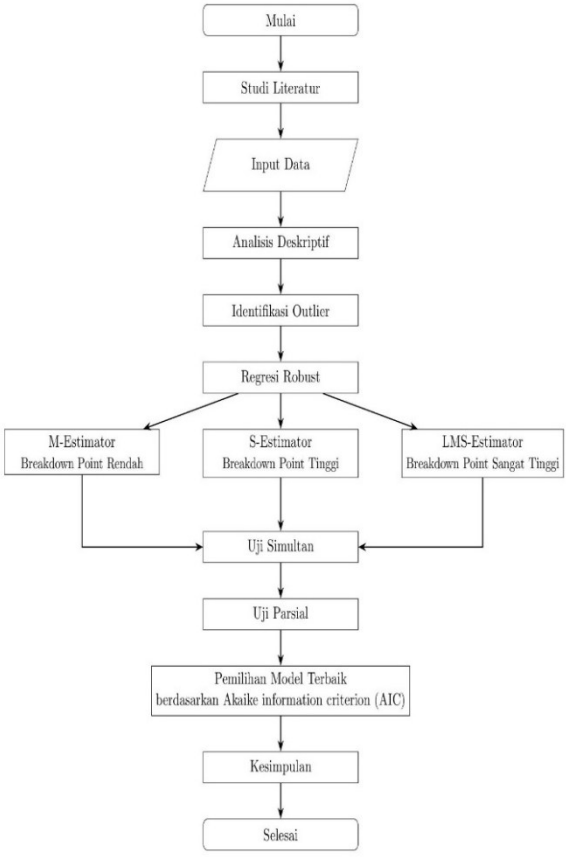
1) Melakukan uji simultan
2) Melakukan uji parsial

f. Mencari model terbaik berdasarkan AIC (*Akaike information criterion*)

Tahap ini bertujuan untuk memilih model yang paling baik berdasarkan nilai AIC dan hasil pengevaluasian model, sehingga didapatkan model yang paling tepat dalam menjelaskan prevalensi penyakit diare balita di Provinsi Jawa Timur.

g. Kesimpulan

Tahap terakhir dari penelitian ini adalah penarikan kesimpulan yang diharapkan dapat menjadi referensi bagi penelitian selanjutnya. Diagram alir penelitian sebagai berikut:".



Gambar 2. Diagram Alir Penelitian

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini adalah data tahun 2024 dengan wilayah penelitian adalah Provinsi Jawa Timur. Terdapat lima data yang digunakan dalam penelitian: (1) Penyakit diare balita (Y), (2) balita yang mendapatkan vitamin A (X_1), (3) penduduk miskin (X_2), (4) posyandu (X_3), dan (5) sarana air minum (X_4). Data yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik Jawa Timur dan Dinas Kesehatan Jawa Timur yang digunakan dalam penelitian ini ditunjukkan pada tabel dibawah ini sebagai berikut:

Tabel 1. Data Penelitian

Y	X_1	X_2	X_3	X_4
0,0126	27369	847	73.03	76
0,0219	43263	1134	80.05	82
0,0170	38044	860	73.75	48
⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
0,4363	91426	1738	159.27	109

Analisis Deskriptif

Analisis deskriptif bertujuan untuk memperoleh nilai serta mendeskripsikan data dari setiap variabel penelitian Berikut disajikan tabel yang digunakan untuk menggambarkan data pada setiap variabel:

Tabel 2. Hasil Analisis Deskriptif

	Min	Max	$\bar{X}$	σ
Y	0,002954	0,067130	0,026316	0,0170
X1	5075	175166	51659	41735,5278
X2	169	2895	1205	708,2287
X3	6,59	240,14	104,81	65,3349
X4	1,00	360,00	80,71	68,5108

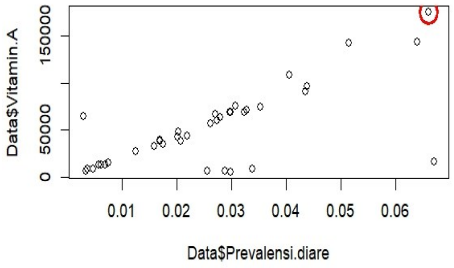
Variabel prevalensi diare balita (Y) menunjukkan distribusi yang relatif seimbang dengan variasi antar wilayah yang kecil, sehingga tingkat kasus cenderung stabil meskipun tetap ada perbedaan daerah. Variabel jumlah balita yang mendapatkan vitamin A (X_1) memiliki variasi yang sangat besar, mencerminkan perbedaan jumlah populasi dan cakupan layanan kesehatan. Variabel jumlah posyandu (X_2) relatif seimbang, namun masih terdapat ketimpangan ketersediaan fasilitas antar wilayah. Variabel jumlah penduduk miskin (X_3) juga menunjukkan variasi yang cukup tinggi,

mengindikasikan adanya kesenjangan sosial ekonomi. Sementara itu, variabel sarana air minum layak (X_4) memiliki variasi besar yang menunjukkan ketimpangan akses air bersih antar wilayah.

Identifikasi Data Outlier

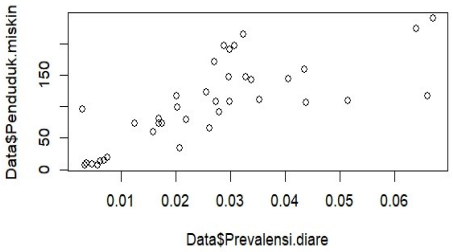
1. *Scatterplot*

   Scatterplot digunakan untuk memeriksa keberadaan *outlier* dalam data. Grafik berbentuk scatterplot yang diperoleh dengan memplot data terhadap observasi ke-i (i=1,2,3,…,n) menghasilkan gambar sebagai berikut:



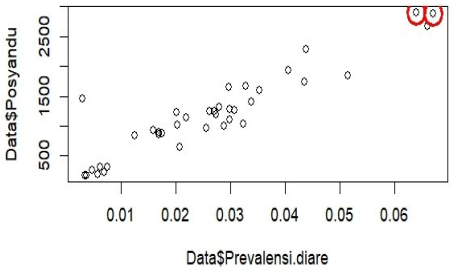
Gambar 3. *Scatterplot* Y dan X_1

Berdasarkan gambar diatas, terlihat adanya kecenderungan pola hubungan positif antara prevalensi diare dan jumlah balita yang mendapatkan vitamin A, yang ditunjukkan oleh pola titik meningkat. Sebagian besar titik pengamatan berada pada rentang nilai vitamin A yang relatif rendah hingga sedang. Selain itu, terdapat satu titik pengamatan yang terlihat menyimpang dari pola umum data dan ditandai dengan lingkaran merah pada grafik.



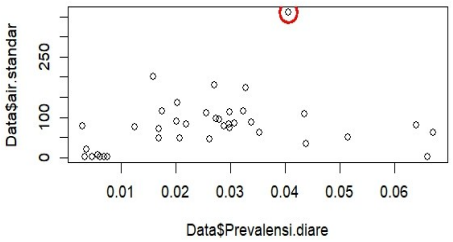
Gambar 4. *Scatterplot* Y dan X_2

Berdasarkan gambar diatas terlihat adanya kecenderungan pola hubungan positif antara prevalensi diare dan jumlah penduduk miskin. Namun, penyebaran data masih cukup bervariasi, dan relatif merata serta tidak terdapat titik pengamatan yang berada jauh dari pola umum. Kondisi ini menunjukkan bahwa data pada variabel tersebut relatif stabil.



Gambar 5. *Scatterplot* Y dan X_3

Berdasarkan gambar diatas, terlihat adanya kecenderungan pola hubungan positif antara prevalensi penyakit diare dan jumlah posyandu, terdapat dua titik pengamatan yang terlihat menyimpang dari pola umum data, ditandai dengan lingkaran merah pada grafik.

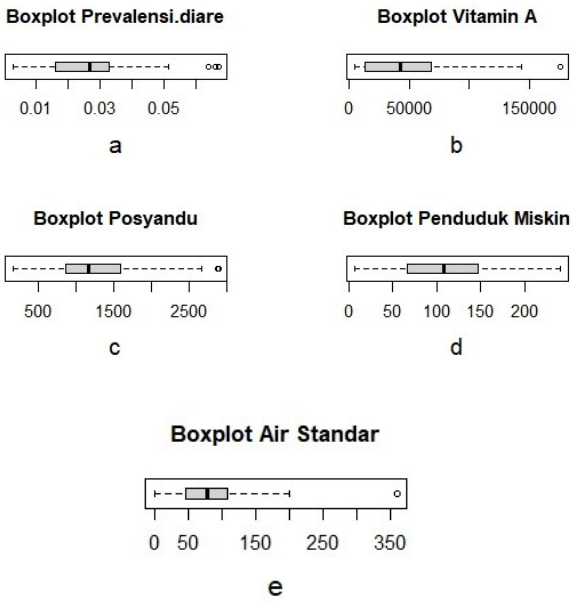


Gambar 6. *Scatterplot* Y dan X_4

Berdasarkan gambar diatas, terlihat adanya kecenderungan pola hubungan negatif antara prevalensi diare dan jumlah sarana air minum sesuai standar, Pola penyebaran data terlihat cukup bervariasi, namun secara umum menunjukkan tren menurun. Selain itu, terdapat satu titik pengamatan yang terlihat menyimpang dari pola umum data dan ditandai dengan lingkaran merah pada grafik.

2. *Boxplot*

   Boxplot bertujuan untuk mendeteksi nilai outlier. Berikut adalah hasil boxplot yang dilakukan dengan mempergunakan nilai kuartil dan jangkauan:



Gambar 7. *Boxplot*

Berdasarkan gambar diatas, yang menunjukkan hasil metode *boxplot* dengan pendekatan IQR (*Interquartile Range*), diperoleh bahwa variabel prevalensi penyakit diare ditemukan tiga *outlier* dengan nilai 0,0671. 0,0640, dan 0,0661 dari total 38 observasi. Variabel jumlah balita mendapatkan vitamin A, teridentifikasi satu *outlier* dengan nilai 175166 dari total 38 observasi. Variabel jumlah penduduk miskin, tidak teridentifikasi *outlier*, variabel jumlah posyandu, teridentifikasi *outlier*, dengan nilai 2888 dan 2895 dari total 38 observasi, Selanjutnya, variabel jumlah air minum sesuai standar, teidentifikasi satu *outlier* dengan nilai 360 dari total 38 observasi. Nilai tersebut berada di luar batas sebaran data berdasarkan aturan IQR, sehingga menunjukkan adanya satu wilayah dengan nilai yang lebih tinggi dibandingkan pengamatan lainnya

3. *Leverage values*

Nilai leverage digunakan untuk mengukur sejauh mana suatu pengamatan memiliki nilai variabel independen yang menyimpang dari pusat sebaran data. Nilai *cutoff* dihitung menggunakan rumus = $\frac{2(p+1)}{n} = \frac{2.(4+1)}{38} = \frac{10}{38} = 0{,}2632$

Tabel 3. Hasil *Leverage Value*s

No	h_{ii}
7	0,6046
14	0,5616
27	0,3311
37	0,3685

Hasil Estimasi Parameter Menggunakan Estimator M (*Maximum Likelikood Type*)

Model regresi *robust* dengan estimasi parameter menggunakan M-Estimator tetap mampu menghasilkan estimasi parameter yang stabil dan andal meskipun terdapat beberapa pengamatan yang teridentifikasi sebagai *outlier*. Selanjutnya diperoleh hasil estimasi parameter yang telah konvergen setelah proses iterasi, sebagai berikut:

Tabel 4. Hasil Estimasi M

	Nilai Estimasi (β)	Standar Error	*t-value*	*p-value*
Intercept	-2,660283 x 10^{-4}	0,0014	-0,1909	-0,0003
X1	3,434693 x 10^{-8}	0,0000	1,3837	0,0000
X2	1,754433 x 10^{-5}	0,0000	8,9110	0,0000
X3	5,971926 x 10^{-5}	0,0000	3,5458	0,0001
X4	-2,525549 x 10^{-5}	0,0000	-2,3162	0,0000

Berdasarkan tabel diatas, diperoleh bentuk persamaan regresi *robust* estimasi M sebagai berikut:

$$Y = -2{,}660283 \text{ x } 10^{-4} + 3{,}434693 \text{ x } 10^{-8} X_1 + 1{,}754433 \text{ x } 10^{-5} X_2 + 5{,}971926 \text{ x } 10^{-5} X_3 - 2{,}525549 \text{ x } 10^{-5} X_4$$

Berdasarkan hasil uji simultan, didapatkan hasil F = 589,8 > $F_{(0,05;4;33)}$ = 2,66 dengan p-*value* = 0,000 < α = 0,05 maka H_0 ditolak, sehingga model regresi signifikan secara simultan layak digunakan untuk analisis. Berdasarkan hasil uji Wald parsial yang diperoleh dari nilai *t-value,* variabel X_2 sebesar 8,9110, X_3 sebesar 3,5458, dan X_4 sebesar -2,3162. Nilai $|t| > t_{(0,025;33)}$ = 2,0345 nilai ketiga variabel tersebut lebih besar dari nilai $t_{tabel,}$ sehingga pada taraf signifikansi $\alpha = 0{,}05$, H_0 ditolak. Hal ini menunjukkan bahwa $X_2, X_3,$ dan X_4 berpengaruh signifikan terhadap variabel dependen. Sementara itu, variabel X_1 memiliki nilai *t-value* sebesar 1,3837 yang lebih kecil dari nilai t_{tabel}, sehingga H_0 gagal ditolak. Dengan demikian, dapat disimpulkan bahwa variabel X_1 tidak berpengaruh signifikan terhadap variabel dependen.

Hasil Estimasi Parameter Menggunakan Estimator S (*Scale Type*)

Model regresi *robust* dengan estimasi parameter menggunakan S-Estimator adalah metode dengan *high breakdown point*. Selanjutnya diperoleh hasil estimasi parameter yang konvergen setelah proses iterasi, sebagai berikut:

Tabel 5. Hasil Estimasi S

	Nilai Estimasi (β)	Standar Error	*t-value*	*p-value*
Intercept	2,945 x 10^{-4}	4,166 x 10^{-4}	0,707	0,4846
X1	3,784 x 10^{-7}	3,870 x 10^{-8}	9,776	2,84 x 10^{-1}
X2	4,658 x 10^{-6}	1,576 x 10^{-6}	2,957	0,0057
X3	-2,030 x 10^{-5}	1,018 x 10^{-5}	-1,994	0,0544
X4	3,658 x 10^{-7}	5,466 x 10^{-6}	0,067	0,9470

Berdasarkan tabel diatas, diperoleh bentuk persamaan regresi robust estimasi S sebagai berikut:

$$Y = 2{,}945 \text{ x } 10^{-4} + 3{,}784 \text{ x } 10^{-7}X_1 + 4{,}658 \text{ x } 10^{-6}X_2 - 2{,}030 \text{ x } 10^{-5}X_3 + 3{,}658 \text{ x } 10^{-5}X_4$$

Berdasarkan hasil uji simultan, didapatkan hasil F = 4118 > $F_{(0,05;4;33)}$ = 2,66 dengan p-*value* = 0,0000 < α = 0,05 maka H_0 ditolak, sehingga model regresi signifikan secara simultan layak digunakan untuk analisis dan dapat disimpulkan bahwa setidaknya satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Berdasarkan hasil uji Wald parsial yang diperoleh dari nilai *t-value,* variabel X_1 sebesar 9,776 dan X_2 sebesar 2,957. Nilai $|t|$ > $t_{(0,025;33)}$ = 2,0345, nilai kedua variabel tersebut lebih besar dari nilai $t_{tabel,}$ sehingga pada taraf signifikansi $\alpha = 0{,}05$, H_0 ditolak. Hal ini menunjukkan bahwa variabel X_1, dan X_2 berpengaruh signifikan terhadap variabel dependen. Sementara itu, variabel X_3 memiliki nilai *t-value* sebesar -1,994 dan X_4 sebesar 0,067 yang keduanya lebih kecil dari nilai t_{tabel}, sehingga H_0 gagal ditolak. Dengan demikian, dapat disimpulkan bahwa variabel X_3 dan X_4 tidak berpengaruh signifikan terhadap variabel dependen.

Hasil Estimasi Parameter Menggunakan Estimator LMS(*Least Median Squares Type*)

Tabel 6. Hasil Estimasi LMS

	Nilai Estimasi (β)	Standar Error	*t-value*	*p-value*
Intercept	4,993 x 10^{-4}	6,634 x 10^{-4}	0,753	0,4585
X1	2,410 x 10^{-7}	2,281 x 10^{-8}	10,562	6,71 x10^{-11}
X2	8,687 x 10^{-6}	1,435 x 10^{-6}	6,054	2,14 x 10^{-6}
X3	1,576 x 10^{-5}	8,829 x 10^{-5}	1,785	0,0860
X4	-1,167 x 10^{-5}	5,12 x 10^{-6}	-2,278	0,0312

Model regresi *robust* dengan estimasi parameter menggunakan LMS-Estimator memiliki *breakdown point* yang tinggi, karena metode ini meminimalkan median dari kuadrat residual, sehingga lebih tahan terhadap pengaruh pencilan

Berdasarkan tabel 6 diatas, diperoleh bentuk persamaan regresi *robust* estimasi LMS sebagai berikut:

$$Y = 4{,}993 \text{ x } 10^{-4} + 2{,}410 \text{ x } 10^{-7}X_1 + 8{,}687 \text{ x } 10^{-6}X_2 + 1{,}576 \text{ x } 10^{-5}X_3 - 1{,}167 \text{ x } 10^{-5}X_4$$

Berdasarkan hasil uji simultan, didapatkan hasil F = 652,3 > $F_{(0,05;4;33)}$ = 2,66 dengan p-*value* = 0,000 < α = 0,05 maka H_0 ditolak, sehingga dapat disimpulkan bahwa setidaknya ada satu variabel independen yang berpengaruh signifikan terhadap variabel dependen. Berdasarkan hasil uji Wald parsial yang diperoleh dari nilai *t-value,* variabel X_1 sebesar 10,562, X_2 sebesar 6,0454 dan X_4 sebesar -2,278. Nilai $|t|$ > $t_{(0,25;33)}$ = 2,0345 nilai kedua variabel tersebut lebih besar dari nilai $t_{tabel,}$ sehingga pada taraf signifikansi $\alpha = 0{,}05$, H_0 ditolak. Hal ini menunjukkan bahwa X_1, X_2, dan X_4 berpengaruh signifikan terhadap variabel dependen. Sementara itu, variabel X_3 memiliki nilai *t-value* sebesar 1,785 yang artinya lebih kecil dari nilai t_{tabel}, sehingga H_0 gagal ditolak. Dengan demikian, dapat disimpulkan bahwa X_3 tidak berpengaruh signifikan terhadap variabel dependen.

Akaike Information Criteria

Pemilihan model terbaik dilakukan dengan membandingkan semua model yang diperoleh. Berikut merupakan hasil model terbaik yang dilihat berdasarkan nilai AIC:

Model	AIC
Estimator M (Maximum Likelihood Type)	0,317066
Estimator S (Scale Type)	-49,65224
Estimator LMS (Least Median of Squares Type)	-512,9729

Berdasarkan tabel diatas hasil perbandingan estimasi M, estimasi S dan estimasi LMS, didapatkan bahwa estimasi LMS lebih disarankan untuk digunakan dibandingkan dengan metode estimasi M dan estimasi S. Hal tersebut dikarenakan nilai AIC yang paling rendah sebesar -512,9729 dengan tiga variabel yang signifikan yaitu vitamin A, penduduk miskin, dan air standar aman.

PENUTUP

SIMPULAN

Berdasarkan hasil analisis dan pembahasan yang diperoleh, dapat disimpulkan bahwa:

1. Model prediksi berdasarkan estimasi parameter menggunakan analisis regresi *robust* Estimator M

$$Y = -2{,}660283 \text{ x } 10^{-4} + 3{,}434693 \text{ x } 10^{-8}X_1 + 1{,}754433 \text{ x } 10^{-5}X_2 + 5{,}971926 \text{ x } 10^{-5}X_3 - 2{,}525549 \text{ x } 10^{-5}X_3$$

2. Model prediksi berdasarkan estimasi parameter menggunakan analisis regresi *robust* Estimator S

$$Y = 2{,}945 \text{ x } 10^{-4} + 3{,}784 \text{ x } 10^{-7}X_1 + 4{,}658 \text{ x } 10^{-6}X_2 - 2{,}030 \text{ x } 10^{-5}X_3 + 3{,}658 \text{ x } 10^{-5}X_4$$

3. Model prediksi berdasarkan estimasi parameter menggunakan analisis regresi *robust* Estimator LMS:

$$Y = 4{,}993 \text{ x } 10^{-4} + 2{,}410 \text{ x } 10^{-7}X_1 + 8{,}687 \text{ x } 10^{-6}X_2 + 1{,}576 \text{ x } 10^{-5}X_3 - 1{,}167 \text{ x } 10^{-5}X_4$$

4. Berdasarkan hasil perbandingan nilai *Akaike Information Criteria* pada analisis regresi *robust* M-Estimator, S-Estimator dan LMS-Estimator data penyakit diare balita di Provinsi Jawa Timur Tahun 2024 diketahui bahwa model dengan menggunakan LMS-Estimator lebih baik digunakan dari pada M-Estimator dan S-Estimator hal tersebut dikarenakan nilai AIC paling rendah sebesar -512,9729.

SARAN

Penelitian selanjutnya disarankan untuk menggunakan data dengan cakupan wilayah yang lebih luas dari beberapa provinsi di Indonesia serta rentang tahun pengamatan yang lebih Panjang, sehingga diperoleh hasil estimasi yang lebih representatif dan stabil. Selain itu, dapat mengembangkan mtode estimasi *robust* lain seperti LTS (*Least Trimmed Squares Type*). Penambahan variabel lain yang relevan untuk meningkatkan kualitas model yang dihasilkan.

DAFTAR PUSTAKA

Algifari, A. (2000). Analisis Regresi, Teori, Kasus dan Solusi. *Penerbit BPFE Yogyakarta*.

Badan Pusat Statistik. (2025). Indikator makro sosial ekonomi Jawa Timur triwulan II2025.https://jatim.bps.go.id/id/publication/2025/09/30/b9f5eda44af3f531ab330256/indikator-makro-sosial-ekonomi-jawa-timur-triwulan-ii-2025.html Accessed:25-10-15

Chen, C. (2002). Robust Regression and Outlier Detection with the Robustreg Procedure. SUGI paper 265-27. SAS Institute : Cary, NC.

Dinas Kesehatan Provinsi Jawa Timur. (2023). *Profil Kesehatan Provinsi Jawa Timur Tahun 2023.*

Dinas Kesehatan Provinsi Jawa Timur. (2024). Profil Kesehatan Provinsi Jawa Timur tahun 2024. https://dinkes.jatimprov.go.id/userfile/dokumen/PROFIL%20KESEHATAN%20PROVINSI%20JAWA%20TIMUR%20TAHUN%202024.pdf. Accessed: 25-10-15

Galton, F. (1885). Regression towards mediocrity in hereditary stature. The Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246-263

Ghozali, I. (2018). Aplikasi Analisis Multivariate SPSS 25. Semarang: Badan Penerbit Universitas Diponegoro. Acc

Kutner, M. H., Neter, J., Nachtsheim, C., & Wasserman, W. (2004). *Applied Linear Regression Models* (4th ed.). McGraw-Hill.

Ratnasari. (2017). Pemodelan Konsumsi dan Estimasi Impor Daging Sapi untuk Keperluan Rumah Tangga di indonesia Tahun 2016 Menggunakan Regresi Robust. Skripsi Sekolah Tinggi Ilmu Statistik.

Romdi, Wahyuningsih S, dan Yuniarti D. (2015). Regresi Robust Linear Sederhana dengan Menggunakan Estimasi MM (Method of Memont). Jurnal Eksponensial. Vol.6. No. 2. Universitas Mulawarman. Samarinda.

Susanti, Y., Pratiwi, H., Sulistijowati H., S., & Liana, T. (2013). M Estimation, S Estimation, and MM Estimation in Robust Regression. *International Journal of Pure and Applied Mathematics*, *91*(3), 349-360.

Soemartini. (2007). Pencilan (*Outlier*). Jurnal Penelitian Universitas Padjadjaran, Jatinangor, 5.

Walpole, R.E., Raymond, H.M., Shaton, L.M., dan Keying, Y. (2011*). Probability And Statistics for Engineers and Scientist,* (9th ed), Lifland et al, New York.

Widodo, E., & Dewayanti, A. A. (2016). Perbandingan Metode Estimasi LTS. Estimasi

M, dan Estimasi MM pada Regresi Robust, Yogyakarta: Universitas Islam Indonesia.

Wijaya, S. (2009). Taksiran Parameter pada Model Regresi Robust. Skripsi Universitas Indonesia.

Verardi, V. (2008). Robust Statistic In Stata. Belgium: FNRS