

PEMODELAN REGRESI ROBUST LEAST TRIMMED SQUARE (LTS)

(Studi Kasus : Faktor -Faktor yang Mempengaruhi Penyebaran Penyakit Malaria di Indonesia)

M. Y. Matdoan^{1*)}, M. W. Talakua²⁾, M. S. Noya Van Delsen³⁾
Program Studi Statistika, Fakultas MIPA Universitas Pattimura
e-mail : keepyahya@gmail.com*

Abstract

Ordinary Least Square (OLS) method is based on a normal distribution, so it is not appropriate to analyze a number of data that are not symmetrical or contain outliers. Therefore, the Robust Least Trimmed Square (LTS) regression is used. This study compares the robust regression of LTS with OLS in the case of factors affecting malaria in Indonesia. The results obtained that by using the OLS method obtained by 0.681 and robust LTS regression of 0.732. so we get the best model, the LTS robust regression model. Further results were obtained that the main factors causing the spread of malaria in Indonesia were poverty, health personnel and clean water.

Keywords : Robust LTS Regression, Outliers, Malaria

1. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk menggambarkan hubungan antara dua atau lebih variabel, sehingga suatu variabel dapat diprediksikan dari variabel yang lain. Variabel yang ditaksir nilainya disebut variabel respon (y), sedangkan variabel penaksir disebut sebagai variabel prediktor (x).

Pendekatan standar untuk mendapatkan nilai dugaan parameter dari model regresi linier adalah metode *Ordinary Least Square* (OLS). Dalam analisis regresi parametrik (khususnya regresi klasik), Gujarati (2004) menyatakan bahwa ada beberapa asumsi yang harus dipenuhi dalam melakukan estimasi, yaitu tidak terjadi multikolinieritas antar variabel

prediktor, tidak terjadi otokorelasi dan heteroskedastisitas, serta residual berdistribusi normal. Jika semua asumsi tersebut terpenuhi, maka hasil estimasi dengan menggunakan metode OLS bersifat *Best Linear Unbiased Estimator* (BLUE).

Estimasi parameter dengan menggunakan metode OLS diperoleh dengan meminimumkan jumlah kuadrat sisaan (residual) dan berdasarkan pada distribusi mean, dimana nilai mean menunjukkan ukuran pemusatan dari suatu distribusi data. Oleh sebab itu, pendekatan dengan menggunakan metode OLS tidak dapat merepresentasikan keseluruhan data. Untuk mengatasi masalah tersebut, maka dikembangkanlah metode regresi *Robust Least Trimmed Square* merupakan metode yang dapat digunakan sebagai alternatif untuk menyelesaikan permasalahan data yang fluktuatif dan mengandung outliers.

Malaria adalah penyakit infeksi yang disebabkan oleh parasit *Plasmodium* yang hidup dan berkembang biak dalam sel darah merah manusia, ditularkan oleh nyamuk (*Anopheles*) betina dan dapat menyerang semua orang, baik laki-laki maupun perempuan pada semua golongan umur dari bayi, anak-anak maupun orang dewasa. Penyakit malaria secara langsung dapat menyebabkan seseorang sering sakit, daya tahan tubuh rendah serta menyebabkan kematian.

Di Indonesia penyakit malaria ditemukan tersebar luas di berbagai daerah dengan derajat infeksi yang bervariasi. Kejadian tersebut disebabkan karena adanya permasalahan-permasalahan teknis seperti pembangunan yang tidak peduli terhadap kesehatan lingkungan, mobilitas penduduk dari daerah endemis malaria, wilayah terpencil dengan kondisi lingkungan yang tidak baik dan minimnya akses pelayanan kesehatan serta kondisi ekonomi masyarakat yang rendah. Jumlah Penderita positif penyakit malaria di Indonesia pada tahun 2015 sebanyak 217.025 jiwa, Berdasarkan data dari Kementerian Kesehatan Republik Indonesia bahwa jumlah penderita penyakit malaria sudah mengalami penurunan. Namun, penurunan penderita penyakit malaria di Indonesia masih tetap harus diwaspadai, karena adanya disparitas (kesenjangan) penyebaran penyakit malaria antar daerah di Indonesia. Sehingga, diduga datanya mengandung pencilan (*outliers*). Oleh karena itu, perlu dilakukan penelitian lebih lanjut mengenai faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia.

2. Metode Penelitian

2.1 Sumber data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Kementerian Kesehatan Republik Indonesia dan Badan Pusat Statistika (BPS) pada tahun 2015. Data yang diambil mengenai presentase *annual parasite incidence* (API) malaria di Indonesia beserta faktor-faktor yang diduga mempengaruhi.

2.2 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini terdiri dari respon dan prediktor. Variabel respon yang digunakan yaitu *Annual parasite incidence* (API) malaria (Y) dan variabel prediktor yang digunakan yaitu presentase akses sanitasi layak (X_1), presentase rumah layak huni (X_2), presentase penduduk miskin (X_3), rasio tenaga dokter (X_4), presentase akses air bersih (X_5) dan rasio puskesmas (X_6).

2.3 Langkah Analisis

Langkah-langkah yang dilakukan dalam penelitian ini adalah sebagai berikut :

- a. Membuat gambaran umum tentang pertumbuhan ekonomi di Provinsi Maluku.
- b. Membuat *Scatter Plot* antara variabel prediktor dengan variabel respon untuk mengetahui pola data.
- c. Memodelkan Persentase pertumbuhan ekonomi dan faktor-faktor yang mempengaruhinya dengan menggunakan spline linier 1 knot, 2 knot dan 3 knot.
- d. Memilih titik knot optimal dengan menggunakan metode *Generalized Cross Validation* (GCV) dimana titik knot optimal berkaitan dengan GCV minimum.
- e. Menghitung nilai *Koefisien Determinasi* (R^2) dan *Mean Squared Error* (MSE).

- f. Melakukan interpretasi model dan menarik kesimpulan.

3. Hasil dan Pembahasan

A. Analisis Deskriptif dan Identifikasi Data

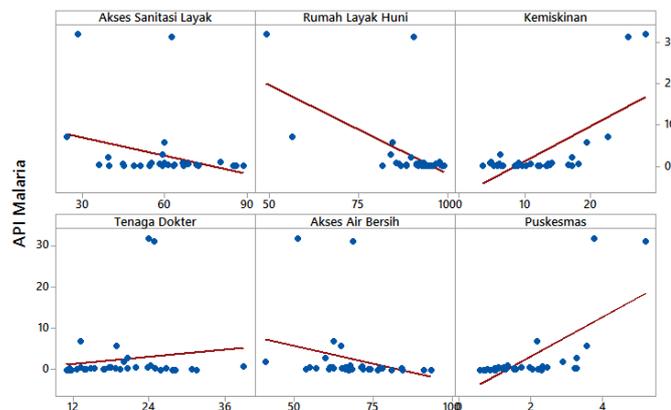
Hasil statistik deskriptif pada kasus penyakit malaria dan faktor-faktor yang mempengaruhi penyebaran penyakit malaria disajikan pada Tabel 1 berikut :

Tabel 1. Statistika Deskriptif

Deskriptif	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
Mean	2,63	59	89,8	11,7	20,1	68,6	1,86
Min	0	23,9	49,2	3,6	10,9	41,0	0,58
Maks	31,9	89,2	99,2	28,4	39,1	93,4	5,20
B. Range	31,9	65,3	49,9	24,7	28,2	52,3	4,62
Std Dev	7,51	15,3	10,4	6,1	6,8	11,0	1,07
Variansi	56,4	236	108,3	38,2	46,4	121,7	1,15

Identifikasi Pola Hubungan dan Pengujian Outlier

Studi kasus pada penelitian ini adalah faktor-faktor penyebab penyebaran penyakit malaria di Indonesia, dimana akan diteliti faktor penyebab penyebaran penyakit malaria (X) terhadap banyaknya presentase penderita penyakit malaria (Y). Adapun pola hubungan antara variabel respon (Y) dengan variabel prediktor (X) dapat dilihat pada Gambar 1 berikut :



Gambar 1 Scatterplot antara variabel X dan variabel Y

Pada Gambar 1 *scatterplot* di atas, menunjukkan bahwa pola hubungan antara presentase penderita malaria (Y) dengan faktor-faktor penyebab penyakit malaria (X) memiliki pola tidak jelas dan sebaran data menyebar. Pemeriksaan hubungan antar variabel secara statistik dilakukan dengan menggunakan *Korelasi Pearson*.

Pengujian hipotesis korelasi adalah sebagai berikut :

H_0 : Tidak ada hubungan antar kedua variabel

H_1 : Terdapat hubungan antar kedua variabel

Koefisien korelasi antar variabel respon dengan variabel prediktor disajikan dalam Tabel 2.

Tabel 2. Koefisien korelasi antar variabel

Var	Y	X ₁	X ₂	X ₃	X ₄	X ₅
X ₁	[-0,2] (0,08)					
X ₂	[-0,5] (0)*	[0,71] (0)*				
X ₃	[0,70] (0)*	[-0,44] (0)*	[-0,66] (0)*			
X ₄	[0,13] (0,44)	[0,28] (0,09)	[0,14] (0,40)	[-0,11] (0,52)		
X ₅	[-0,2] (0,14)	[0,70] (0)*	[0,54] (0)*	[-0,43] (0,01)*	[0,31] (0,06)	
X ₆	[0,68] (0)*	[-0,39] (0,02)*	[-0,47] (0)*	[0,54] (0)*	[0,24] (0,15)	[-0,40] (0,01)*

Ket : [] Korelasi Pearson,

() *P-value*, *) Sig pada $\alpha = 5\%$

Berdasarkan Tabel 2 diatas, menunjukkan bahwa terdapat tiga variabel yang berkorelasi positif terhadap *annual parasite incidence* (API) malaria (Y) yaitu presentase rumah layak huni (X₂), variabel presentase penduduk miskin (X₃) dan variabel rasio puskesmas (X₆). Hal ini menunjukkan bahwa jika terjadi peningkatan pada variabel presentase rumah layak huni (X₂), variabel presentase penduduk miskin (X₃) dan variabel rasio puskesmas (X₆) akan mengakibatkan semakin tingginya *annual parasite incidence* (API) malaria (Y) begitu sebaliknya. Sementara itu, variabel presentase akses sanitasi layak (X₁), variabel rasio tenaga dokter (X₄) dan variabel presentase akses air bersih (X₅) memiliki korelasi negatif artinya jika terjadi peningkatan pada variabel tersebut maka akan mengakibatkan penurunan pada *annual parasite incidence* (API) malaria (Y). Selanjutnya akan untuk memastikan ada atau tidaknya kasus multikolinieritas, dapat dilihat dari nilai *Variance*

Inflation Factors (VIF) dari setiap variabel prediktor. Berikut pada Tabel 3, menunjukkan nilai *Variance Inflation Factors* (VIF) dari masing-masing variabel prediktor yang digunakan dalam penelitian ini.

Tabel 3. Nilai *Variance Inflation Factor* (VIF)

Variabel Prediktor	Nilai VIF
X ₁	3,05
X ₂	3,01
X ₃	2,20
X ₄	1,48
X ₅	2,24
X ₆	1,99

Berdasarkan Tabel 3, menunjukkan bahwa nilai VIF untuk semua variabel prediktor adalah lebih kecil dari 10 yang artinya tidak terjadi multikolinieritas diantara variabel prediktor. Pengujian data outlier dengan menggunakan metode *DFFITs*. Pengamatan dinyatakan sebagai outlier jika

$$|DFFITs| > 2\sqrt{(k+1)/n} = 0,907.$$

Tabel 4. Pengujian Outlier

Propinsi	<i>DFFITs</i>
Nusa Tenggara Timur	1,69655
Sulawesi Tenggara	1,28412
Papua Barat	6,41416
Papua	4,55268

Berdasarkan Tabel 4, dapat dilihat bahwa terdapat empat Propinsi yang dinyatakan sebagai outlier yaitu Propinsi Nusa Tenggara Timur (NTT), Propinsi Sulawesi Tenggara, Propinsi Papua Barat dan Propinsi Papua. Apabila analisis regresi dilakukan dengan menggunakan metode OLS, akan mengakibatkan varians dari penduga parameter besar, sehingga tidak efisien. Oleh karena itu, digunakan regresi kuantil dan regresi *robust* LTS untuk menentukan parameter dari faktor-faktor yang mempengaruhi peningkatan presentase penderita penyakit malaria di Indonesia

C. Estimasi Parameter

1. Estimasi Parameter dengan Menggunakan Metode *Ordinary*

Least Square (OLS)

Hasil Estimasi metode *Ordinary Least Square* (OLS) dapat dilihat pada Tabel 5 berikut :

Tabel 5. Estimasi Paramete dengan *Ordinary Least Square* (OLS)

Parameter	Estimasi	P- Value	R ²
β_0	2,211	0,865	
β_1	0,095	0,314	
β_2	-0,255	0,071	
β_3	0,451	0,028	0,681
β_4	0,024	0,870	
β_5	0,091	0,417	
β_6	3,073	0,008	

Berdasarkan Tabel 5, dapat dilihat bahwa terdapat dua variabel independen yang berpengaruh signifikan terhadap banyaknya *Annual Parasite Incidence* (API) Malaria di Indonesia yaitu variabel presentase penduduk miskin (X_3) dan variabel rasio puskesmas (X_6). Hasil Estimasi tersebut dapat dimodelkan sebagai berikut :

$$\hat{Y} = 2,211 + 0,095X_1 - 0,255X_2 + 0,451X_3 + 0,024X_4 + 0,091X_5 + 3,073X_6$$

2. Estimasi Parameter dengan Metode Regresi *Robust LTS*

Metode regresi *robust LTS* dapat digunakan untuk mengatasi pengamatan *outlier*.

Tabel 6. Estimasi Parameter dengan *Robust LTS*

Parameter	Estimasi	P- Value	R ²
β_0	-1,324	0,231	
β_1	-0,001	0,774	
β_2	0,018	0,177	
β_3	0,035	0,000	
β_4	0,029	0,000	0,732
β_5	-0,014	0,022	
β_6	0,069	0,204	

Berdasarkan Tabel 6, dapat dilihat bahwa terdapat tiga variabel independen yang berpengaruh secara signifikan terhadap banyaknya jumlah penderita malaria di Indonesia yaitu variabel presentase kemiskinan (X_3), variabel rasio tenaga kesehatan (X_4) dan variabel presentase akses air bersih (X_5). Pemodelan metode regresi *robust LTS* sebagai berikut :

$$\hat{Y} = -1,324 - 0,001X_1 + 0,018X_2 + 0,035X_3 + 0,029X_4 - 0,014X_5 + 0,069X_6$$

4. Kesimpulan

Berdasarkan hasil dan pembahasan, maka diperoleh kesimpulan dari penelitian ini adalah sebagai berikut :

1. Hasil perbandingan metode estimasi yang diperoleh yaitu metode Regresi *Robust Least trimmed Square* (LTS) lebih baik dibandingkan dengan metode *Ordinary Least Square* (OLS) pada kasus faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia. Hal ini, ditunjukkan dengan nilai R^2 dari model Regresi *Robust Least Trimmed Square* (LTS) sebesar 0,732 dan model *Ordinary Least Square* (OLS) sebesar 0,681.
2. Faktor-faktor yang mempengaruhi penyebaran penyakit malaria di Indonesia dengan menggunakan model terbaik, pada model regresi *robust LTS* adalah faktor kemiskinan, faktor tenaga kesehatan, dan faktor air bersih. Oleh karena itu, Pemerintah beserta seluruh elemen masyarakat perlu mengevaluasi dan melakukan proses pencegahan agar terhindar dari penyebaran penyakit malaria.

DAFTAR PUSTAKA

- Badan Pusat Statistika. (2014). *Hasil Survei Sosial Ekonomi Nasional Tahun 2014*. Jakarta
- Budiantara, I.N. (2011), *Penelitian Bidang Regresi Spline Menuju Terwujudnya Penelitian Statistika yang Mandiri dan Berkarakter*. Seminar Nasional FMIPA Universitas Pendidikan Ganesha.
- Datukramat, Mayulu dan Masi. (2013). *Menganalisis Hubungan Antara Sanitasi Lingkungan Dengan Kejadian Malaria Pada Murid Sekolah Dasar Di Kabupaten Bolaang Mongondow Utara*. journal keperawatan (e-Kp) Fakultas Kedokteran Universitas Sam Ratulangi.
- Ghozali, Imam. 2013. *Aplikasi analisis Multivariat dengan program SPSS*. Edisi ketujuh. Semarang
- Gujarati, D. N. (2004). *Basic Econometrics, 4th Edition*. New York: McGrahill. Co.
- Hardle, W. (1990). *Applied Nonparametric Regression*. Cambridge University Press. New York.
- Hastie, T. J. dan Tibshirani, R. J. (1990). *Generalized Additive Models*. Chapman and Hall. New York. London.
- Indriani dan Yuyun. (2010). *Analisis faktor-faktor yang mempengaruhi angka kejadian penyakit tropis malaria dengan metode klasifikasi pohon* (Studi kasus Kabupaten

- Sorong Selatan, Provinsi Papua Barat). Jurnal, Institut Teknologi Sepuluh Nopember
- Kementrian Kesehatan Republik Indonesia. (2014). *Profil kesehatan Indonesia*, Pusat data dan Informasi Kementrian Kesehatan. Jakarta.
- Lin W, Zongwu, Li. (2015). *Optimal smoothing in nonparametric conditional quantile derivative function estimation*. Journal of Econometrics ScienceDirect vol. 188, hal. 502-513.
- Mirontoneng, A.R, Ismanto, A.Y dan Malara, R. (2014). *Analisis faktor-faktor yang berhubungan kejadian malaria pada anak di wilayah kerja PKM Tona kecamatan tahuna kabupaten Sangihe*. Program Studi Ilmu Keperawatan Fakultas kedokteran Universitas Sam Ratulangi Manado.
- Susilowati, P, A. (2013). *Analisis regresi pada prevalensi malaria di Provinsi Maluku Utara, Malauku, Papua Barat dan Papua dengan faktor yang mempengaruhinya*. Skripsi, Jurusan Statistika Institut Teknologi Sepuluh Nopember.
- Werner, Mark. (2003). *Identification of Multivariate Outliers in Large Data Sets*, University of Colorado at Denver.
- Willems, G. and Aelst, S.V. (2005). *Fast and Robust Bootstrap for LTS* Computational Statistics and Data Analysis, No. 48, hal. 703-715.