

Regresi Kuantil pada Data Jumlah Kematian Bayi di Jawa Timur Pada Tahun 2019

Ajeng Hananingrum*, Anneke Iswani Achmad

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*ajenghananingrum@gmail.com, annekeiswani11@gmail.com

Abstract. Quantile regression is a regression method introduced by Koenker and Basset in 1978. This regression divides data into several group that thought to have different parameter values on specific quantile. Quantile regression can be used on data distribution is not homogeneous and asymmetric. Quantile regression is also unaffected by outlier data. The estimate parameters is conducted using the Least Absolute Deviation (LAD) by giving τ as weight to positive error and giving $(1 - \tau)$ as weight to negative error. The multiplicity between weight and error called Loss Function. Optimum results obtained when loss function has a minimum. To minimize loss function is used the simplex method. The study was done with quantile regression modelling using data of the number of infant deaths in east Java in 2019 with τ used at 0,1 0,25 0,5 0,75, and 0.9. The result show that there are differences in value parameters in each quantile and each quantile has significantly different variable effects.

Keywords: Quantile Regression, Least Absolute Deviation (LAD), Loss Function, Simplex Method.

Abstrak. Regresi kuantil adalah metode regresi yang diperkenalkan oleh Koenker dan Bassett pada tahun 1978. Regresi ini membagi data menjadi beberapa kelompok yang diduga mempunyai perbedaan nilai parameter pada kuantil-kuantil tertentu. Regresi kuantil dapat digunakan pada data yang distribusinya tidak homogen dan tidak simetris. Bahkan regresi kuantil tidak terpengaruh oleh data *outlier*. Estimasi parameter dilakukan menggunakan metode *Least Absolute Deviation* (LAD) dengan memberi pembobot τ pada sisaan positif dan $(1 - \tau)$ pada sisaan negatif. Perkalian antara pembobot dan sisaan dinamakan *Loss Function*. Hasil optimum didapat saat *Loss Function* memiliki nilai minimum. Untuk meminimumkan *Loss Function* maka digunakan metode simpleks. Pada penelitian ini dilakukan pemodelan regresi kuantil menggunakan data jumlah kematian bayi di Jawa Timur Tahun 2019 dengan τ yang digunakan yaitu 0,1 0,25 0,5 0,75 dan 0,9. Hasil regresi kuantil menunjukkan terdapat perbedaan nilai parameter pada masing-masing kuantil dan setiap kuantil memiliki variabel berpengaruh secara signifikan yang berbeda-beda.

Kata Kunci: Regresi Kuantil, *Least Absolute Deviation* (LAD), *Loss Function*, Metode Simpleks.

1. Pendahuluan

Salah satu metode statistika yang sering digunakan sebagai alat analisis yaitu analisis regresi. Analisis regresi merupakan kajian terhadap hubungan satu variabel yang disebut sebagai variabel yang diterangkan (variabel terikat) dengan satu atau dua variabel yang menerangkan (variabel bebas) (Gujarati, 2006). Untuk membuat model regresi tersebut maka perlu dilakukan estimasi terhadap parameter-parameter regresi. Salah satu metode yang umum digunakan yaitu Metode Kuadrat Terkecil (MKT) atau *Ordinary Least Square* (OLS).

Estimasi parameter dengan MKT dilakukan dengan meminimumkan jumlah kuadrat sisaan yaitu dengan menurunkan jumlah kuadrat sisaan terhadap masing-masing parameter kemudian disamakan dengan nol. Metode ini akan menghasilkan nilai estimasi parameter yang bersifat *Best Linear Unbiased Estimator* (BLUE) jika memenuhi asumsi-asumsi regresi klasik. Namun metode ini sangat sensitif terhadap pelanggaran asumsi klasik tersebut sehingga nilai parameter yang dihasilkan tidak lagi bersifat BLUE.

Berdasarkan hal diatas maka munculah regresi median yaitu regresi yang dilakukan dengan membagi data menjadi 2 bagian. Regresi median dilakukan dengan meminimumkan $\sum_{i=1}^n |\varepsilon_i|$ atau disebut juga metode *Least Absolute Deviation* (LAD) (Irevanie, 2017). Hal ini mempertimbangkan apabila data berbentuk lonceng atau tidak simetris. Namun pendekatan regresi median hanya membagi data menjadi 2 kelompok. Maka berkembanglah menjadi regresi kuantil yaitu regresi yang dapat membagi data menjadi lebih dari 2 kelompok.

Regresi kuantil diperkenalkan oleh Kankoe dan Basset pada tahun 1978 sebagai pengembangan dari regresi median. Regresi kuantil membagi data menjadi beberapa kelompok yang diduga mempunyai perbedaan nilai parameter pada kuantil-kuantil tertentu. Selain itu regresi kuantil sangat berguna ketika digunakan pada data yang distribusinya tidak homogen dan tidak simetris. Bahkan regresi kuantil tidak terpengaruh oleh outlier (Widodo, 2016). Regresi kuantil menggunakan metode LAD dengan memberikan pembobot yang berbeda pada $\sum_{i=1}^n |\varepsilon_i|$ yaitu diberikan pembobot τ pada sisaan positif dan pembobot $(1-\tau)$ pada sisaan negatif, dengan τ merupakan kuantil.

Dalam rangka meningkatkan kesehatan keluarga, maka indikator kesehatan ibu dan anak merupakan hal yang perlu diperhatikan. Maka angka kematian ibu dan angka kematian anak perlu diperhatikan. Salah satu upaya untuk mewujudkan kesejahteraan anak adalah dengan pemeliharaan dan perlindungan kesehatan anak sejak dalam kandungan sampai sesudah dilahirkan. Hasil Survey Demografi dan Kesehatan Indonesia (SDKI) Tahun 2017 menunjukkan Angka Kematian Bayi sebesar 24 per 1000 kelahiran hidup (KPPPA, 2018). Angka ini menunjukkan penurunan dari tahun-tahun sebelumnya. Meskipun demikian, pemerintah mempunyai target dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 yaitu penurunan angka kematian bayi menjadi 16 kematian per 1000 kelahiran hidup. Kematian bayi dipengaruhi oleh persalinan ditolong oleh tenaga kesehatan dan bayi dengan berat bayi lahir rendah/BBLR (Tarigan, 2017). Oleh karena itu, pada penelitian ini akan dibahas mengenai penerapan regresi kuantil pada jumlah kematian bayi. Sementara data yang digunakan merupakan data jumlah kematian bayi, jumlah persalinan ditolong oleh tenaga kesehatan dan jumlah bayi dengan berat bayi lahir rendah/BBLR di Jawa Timur pada tahun 2019.

Berdasarkan latar belakang diatas maka identifikasi masalah dalam penelitian ini adalah “Bagaimana model regresi kuantil pada data jumlah kematian bayi di Jawa Timur Tahun 2019?”. Sehingga didapat tujuan dari penelitian ini adalah untuk mengetahui model regresi kuantil pada data jumlah kematian bayi di Jawa Timur Tahun 2019.

2. Landasan Teori

Regresi kuantil diperkenalkan pertama kali oleh Koenker dan Bassett pada tahun 1978. Regresi ini berguna pada berdistribusi data tidak homogen dan tidak berbentuk standar. Bentuk tidak standar antara lain tidak simetris, terdapat ekor dalam sebaran (Balami, 2017). Sehingga model regresi kuantil dapat dituliskan sebagai berikut :

$$Y_i = \beta_{0,\tau} + \beta_{1,\tau}X_{1,i} + \dots + \beta_{k,\tau}X_{k,i} + \varepsilon_{i,\tau}, i = 1, \dots, n$$

dimana,

$$\hat{\beta}_\tau = \arg \min_{\beta} \left\{ \tau \sum_{t; y_t \geq x_t} |y_t - x_t \beta| + (1 - \tau) \sum_{t; y_t < x_t} |y_t - x_t \beta| \right\}$$

atau

$$\hat{\beta}_\tau = \arg \min_{\beta} \sum_{i=1}^n \rho_\tau(\varepsilon_i)$$

dimana

$$\rho_\tau(\varepsilon_i) = \text{Loss Function} = \begin{cases} \tau \varepsilon_i & , \text{ untuk } \varepsilon_i \geq 0 \\ -(1 - \tau) \varepsilon_i & , \text{ untuk } \varepsilon_i < 0 \end{cases}$$

Untuk mencari solusi bagi persamaan diatas tidak dapat dilakukan secara analitik, melainkan dapat dilakukan dengan numerik menggunakan metode *simplex*.

Untuk menduga nilai parameter regresi kuantil dengan metode simpleks diperlukan tabel simpleks atau disebut juga dengan tabulasi simpleks. Berikut merupakan bentuk umum tabel simpleks untuk menduga parameter regresi kuantil.

Tabel 1. Bentuk Umum Tabel Simpleks

	D _j	0	0	...	0	τ	...	τ	1- τ	...	1- τ		
C _b	V _b	x ₀	x ₁	...	x _k	q ₁	...	q _n	r ₁	...	r _n	solusi	ratio
τ	q ₁	a _{ij}										y ₁	
⋮	⋮											⋮	
τ	q _n											y _n	
Z _j													
D _j - Z _j													

Keterangan:

D_j = baris koefisien fungsi tujuan

C_b = kolom koefisien variabel basis

V_b = kolom variabel basis pada iterasi tersebut

Solusi = nilai ruas kanan atau y_i dimana i = 1 . . . n

Langkah-langkah untuk menduga parameter regresi kuantil sebagai berikut

1. Mengubah masalah optimasi linier kedalam bentuk standar dengan menambahkan variabel slack, surplus dan artificial lalu masukan pada tabel simpleks.
2. Menghitung nilai Z_j yaitu $Z_j = \sum a_{ij} \times C_b$
3. Menentukan kolom kunci yaitu dengan untuk masalah minimalisasi pilih D_j - Z_j dengan hasil negatif paling besar. Sedangkan masalah maksimalisasi pilih D_j - Z_j dengan hasil positif paling besar.
4. Menghitung nilai ratio yaitu membagi solusi dengan nilai kolom kunci yang bersesuaian $ratio = \frac{y_i}{a_{ij} \text{ kolom kunci}}$
5. Menentukan baris kunci yaitu untuk masalah minimalisasi dan maksimalisasi dengan memilih ratio positif paling kecil.
6. Nilai elemen merupakan nilai perpotongan antara kolom kunci dan baris kunci.
7. Membentuk tabel iterasi kesatu
Variabel kolom kunci (variabel masuk) akan menjadi variabel basis pada iterasi selanjutnya
8. Hitung nilai baris pada iterasi kesatu

Untuk baris yang merupakan baris kunci pada tabel iterasi sebelumnya yaitu

$$\text{baris baru} = \frac{\text{baris kunci lama}}{\text{elemen}}$$

Sementara untuk baris lainnya (yang bukan baris kunci pada tabel iterasi sebelumnya) dengan cara

$$\text{baris} = \text{baris lama} - (\text{baris baru} \times \text{nilai kolom kunci berkaitan})$$

9. Hitung nilai $D_j - Z_j$ lalu pilih kolom yang memiliki nilai negatif paling besar sebagai kolom kunci pada iterasi pertama.
10. Hitung ratio lalu pilih nilai ratio positif terkecil sebagai baris kunci pada iterasi pertama.
11. Kembali lakukan langkah 7

Pada masalah minimalisasi proses iterasi akan berhenti saat semua nilai $D_j - Z_j > 0$. Sedangkan pada masalah maksimalisasi proses iterasi akan berhenti saat semua nilai $D_j - Z_j < 0$.

Berdasarkan penelitian yang dilakukan oleh Tarigan (2017) persalinan ditolong oleh tenaga kesehatan dan bayi dengan berat bayi lahir rendah/BBLR berpengaruh terhadap kelangsungan hidup bayi. Berdasarkan hal tersebut maka dalam penelitian ini variabel-variabel yang digunakan adalah kematian bayi, persalinan ditolong tenaga kesehatan dan bayi dengan berat bayi lahir rendah/BBLR. Berikut ini merupakan pengertian variabel-variabel tersebut.

1. Kematian bayi adalah kejadian kematian yang terjadi pada periode sejak bayi lahir sampai bayi belum berusia tepat satu tahun (Tarigan, 2017).
2. Berdasarkan definisi operasional Kemenkes, persalinan ditolong tenaga kesehatan adalah ibu bersalin yang mendapatkan pertolongan persalinan oleh tenaga kesehatan yang memiliki kompetensi kebidanan (dokter kandungan dan kebidanan, dokter umum dan bidan).
3. Bayi dengan berat bayi lahir rendah/BBLR merupakan berat bayi lahir kurang dari 2500 gram, akan membawa risiko kematian, gangguan pertumbuhan dan perkembangan anak, termasuk dapat berisiko menjadi pendek jika tidak ditangani dengan baik (KPPPA, 2018)

3. Hasil Penelitian dan Pembahasan

Regresi kuantil digunakan untuk membuat model regresi pada kuantil-kuantil tertentu. Pemodelan regresi kuantil pada penelitian ini dilakukan pada kuantil 0,1 0,25 0,5 0,75 dan 0,9. Berikut merupakan nilai estimasi parameter regresi kuantil pada masing-masing kuantil.

Tabel 2. Estimasi Nilai Parameter

Kuantil	Parameter		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
0,1	5.63025	0.00445	- 0.06047
0,25	-4.10939	0.00389	0.04785
0,50	-8.15590	0.00301	0.12178
0,75	17.85269	0.00277	0.11554
0,9	46.94847	0.00198	0.11478

Berikut ini merupakan nilai p-value pada masing-masing estimasi parameter regresi kuantil pada kuantil 0,1 0,25 0,5 0,75 dan 0,9.

Tabel 3. Nilai P-Value Pada Masing-Masing Parameter

Kuantil	Parameter		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
0,1	0.78070	0.00462*	0.13159
0,25	0.53941	0.00000*	0.00074*
0,5	0.37511	0.00007*	0.00000*
0,75	0.04549*	0.00010*	0.00000*
0,9	0.00000*	0.00344*	0.00000*

Ket: Pada taraf signifikansi 5%

Setiap kuantil memiliki model regresi kuantil masing-masing maka masing-masing kuantil memiliki koefisien determinasi tersendiri. Koefisien determinasi ini menunjukkan kebaikan model tersebut. Berikut koefisien determinasi pada masing-masing kuantil.

Tabel 4. Nilai Koefisien Determinasi

	Kuantil				
	0,1	0,25	0,5	0,75	0,9
$R^2(\tau)$	25,21	34,22	44,84	50,94	62,63
%	%	%	%	%	%

Berdasarkan Tabel 4 dapat dijelaskan bahwa koefisien determinasi tiap kuantil berbeda. Berdasarkan kuantil-kuantil diatas, kuantil 0,9 menjadi kuantil dengan koefisien determinasi paling tinggi sebesar 62,63%. artinya variabel jumlah persalinan ditolong tenaga kesehatan (X_1) dan variabel jumlah bayi dengan berat bayi lahir rendah/BBLR (X_2) dapat menjelaskan variabel jumlah kematian bayi (Y) paling baik pada data dibawah kuantil 0,9 yaitu sebesar 62,63% dan sisanya sebesar 37,37% disebabkan oleh variabel lain yang tidak diteliti dalam penelitian ini.

4. Kesimpulan

Berdasarkan hasil analisis yang telah dibahas dapat disimpulkan bahwa

Pada regresi kuantil terdapat perbedaan nilai parameter pada masing-masing kuantil. Berikut model regresi kuantil masing-masing kuantil

Kuantil 0,1

$$\hat{y}_{0,1} = 5.63025 + 0.00445 X_1 - 0.06047 X_2$$

Kuantil 0,25

$$\hat{y}_{0,25} = -4.10939 + 0.00389 X_1 + 0.04785 X_2$$

Kuantil 0,5

$$\hat{y}_{0,5} = -8.15590 + 0.00301 X_1 + 0.12178 X_2$$

Kuantil 0,75

$$\hat{y}_{0,75} = 17.85269 + 0.00277 X_1 + 0.11554 X_2$$

Kuantil 0,9

$$\hat{y}_{0,9} = 46.94847 + 0.00198 X_1 + 0.11478 X_2$$

Interpretasi untuk model regresi kuantil ke-0,1 adalah jika data dipartisi dalam 100 bagian maka untuk 10 data pertama dapat menggunakan model regresi kuantil ke-0,1 yaitu saat variabel jumlah persalinan ditolong tenaga kesehatan (X_1) naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,1}$) bertambah sebesar 45 satuan. Untuk variabel jumlah bayi dengan berat bayi lahir rendah/BBLR (X_2), saat variabel X_2 naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,1}$) berkurang sebesar 605 bayi.

Interpretasi untuk model regresi kuantil ke-0,25 adalah jika data dipartisi dalam 100 bagian maka untuk 25 data pertama dapat menggunakan model regresi kuantil ke-0,25 yaitu saat variabel jumlah persalinan ditolong tenaga kesehatan (X_1) naik 10000 satuan dan variabel lain bernilai tetap maka jumlah kematian bayi ($\hat{y}_{0,25}$) bertambah sebesar 39 bayi. Untuk variabel jumlah bayi dengan berat bayi lahir rendah/BBLR (X_2), saat variabel X_2 naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,25}$) bertambah sebesar 479 bayi.

Interpretasi untuk model regresi kuantil ke-0,5 adalah jika data dipartisi dalam 100 bagian maka untuk 50 data pertama dapat menggunakan model regresi kuantil ke-0,5 yaitu saat variabel jumlah persalinan ditolong tenaga kesehatan (X_1) naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,5}$) bertambah sebesar 30 bayi. Untuk variabel jumlah bayi dengan berat bayi lahir rendah/BBLR (X_2), saat variabel X_2 naik 10000

satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,5}$) bertambah sebesar 1217 bayi.

Interpretasi untuk model regresi kuantil ke-0,75 adalah jika data dipartisi dalam 100 bagian maka untuk 75 data pertama dapat menggunakan model regresi kuantil ke-0,75 yaitu saat variabel jumlah persalinan ditolong tenaga kesehatan (X_1) naik 10000 satuan dan variabel lain bernilai tetap maka jumlah kematian bayi ($\hat{y}_{0,75}$) bertambah sebesar 28 bayi. Untuk variabel jumlah bayi dengan berat bayi lahir rendah/BBLR (X_2), saat variabel X_2 naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,75}$) bertambah sebesar 1155 bayi.

Interpretasi untuk model regresi kuantil ke-0,9 adalah jika data dipartisi dalam 100 bagian maka untuk 90 data pertama dapat menggunakan model regresi kuantil ke-0,9 yaitu saat variabel jumlah persalinan ditolong tenaga kesehatan (X_1) naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,9}$) bertambah sebesar 20 bayi. Untuk variabel jumlah bayi dengan berat bayi lahir rendah/BBLR (X_2), saat variabel X_2 naik 10000 satuan dan variabel lain bernilai tetap maka variabel jumlah kematian bayi ($\hat{y}_{0,9}$) bertambah sebesar 1148 bayi.

Setiap kuantil memiliki variabel berpengaruh secara signifikan yang berbeda-beda. Pada kuantil 0,1 hanya variabel jumlah persalinan ditolong tenaga kesehatan yang berpengaruh secara signifikan terhadap variabel jumlah kematian bayi di Jawa Timur tahun 2019. Pada kuantil 0,25 0,5 0,75 dan 0,9 semua variabel bebas yaitu variabel jumlah persalinan ditolong tenaga kesehatan dan variabel jumlah bayi dengan berat bayi lahir rendah/BBLR berpengaruh secara signifikan terhadap variabel jumlah kematian bayi di Jawa Timur tahun 2019.

5. Saran

Untuk melakukan estimasi parameter terdapat beberapa metode selain metode simpleks, sehingga peneliti menyarankan untuk melakukan estimasi parameter menggunakan metode selain metode simpleks seperti metode interior point atau metode *smoothing*.

Untuk penggunaan variabel, penulis menyarankan untuk memperbanyak variabel bebas agar koefisien determinasi yang diperoleh semakin baik.

Pada penelitian ini, peneliti menggunakan model linier dari regresi kuantil. Namun selain model linier, regresi kuantil dapat dilakukan menggunakan model nonlinier. Sehingga peneliti menyarankan untuk penelitian selanjutnya dapat menggunakan metode *nonlinear quantile regression*.

Daftar Pustaka

- [1] Balami, A. M. (2017). *Estimasi Parameter Regresi Kuantil Pada Kasus Demam Berdarah Dengue di Kota Surabaya*. Tesis tidak dipublikasikan. Surabaya: Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Institut Teknologi Sepuluh Nopember.
- [2] Dinas Kesehatan Provinsi Jawa Timur. (2020). *Profil Kesehatan Provinsi Jawa Timur Tahun 2019*. Surabaya: Diskes Provinsi Jawa Timur.
- [3] Gujarati, D. N. (2006). *Ekonometrika Dasar*. Jakarta: Erlangga.
- [4] Hapsery, A. (2017). *Regresi Kuantil Berbasis Model Rekursif dan Estimasi Sparsity untuk Analisis Publikasi Dosen ITS di Scopus*. Tesis tidak dipublikasikan. Surabaya: Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Institut Teknologi Sepuluh Nopember.
- [5] Irevanie, R. S. (2017). *Perbandingan Metode Quantile Regression (QR) dan Geographically Weighted Regression (GWR) Pada Data Angka Harapan Hidup di Indonesia*. Tesis tidak dipublikasikan. Surabaya: Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Institut Teknologi Sepuluh Nopember.
- [6] Kementerian Pemberdayaan Perempuan dan Perlindungan Anak. (2018). *Profil Anak Indonesia 2018*. Jakarta: KPPPA.

- [7] Khairunisa. (2015). *Estimasi Parameter Analisis Regresi Kuantil Menggunakan Metode Simpleks atau Metode Interior Point*. Skripsi tidak dipublikasikan. Medan: Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sumatera Utara.
- [8] Koenker, R., & Basset, J. (1978). Regression Quantiles. *Econometrica*, **46**(1), 33-50.
- [9] Kurniawan, K., & Yuniarto, B. (2016). *Analisis Regresi: Dasar dan Penerapannya Dengan R*. Jakarta: Kencana.
- [10] Tarigan, I.U., Afifah, T., & Simbolon, Demsa. (2017). Faktor-Faktor yang Berhubungan Dengan Pelayanan Bayi di Indonesia: Pendekatan Analisis Multilevel. *Jurnal Kesehatan Reproduksi*, **8**(1),103-118.
- [11] Utami, A. R. (2018). *Regresi Kuantil Pada Data yang Mengandung Pencilan*. Skripsi tidak dipublikasikan. Lampung: Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.
- [12] Widodo, E. & Andani, F. P. P. (2016). *Regresi Kuantil Median Untuk Mengatasi Heteroskedastisitas*. Laporan Penelitian. Yogyakarta: Direktorat Penelitian dan Pengabdian Masyarakat, Universitas Islam Indonesia.