UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 1 Nomor 5 Tahun 2023 Halaman 377-384

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Application Of Quantile Regression Method to Data Containing Outliers for Crime Rate in Jabodetabek

Arssita Nur Muharromah, Zamahsary Martha*,Dony Permana, Tessy Octavia Mukhti
Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia
*Corresponding author: zamahsarymartha@fmipa.unp.ac.id

Submitted : 25 Juli 2023
Revised : 19 Oktober 2023
Accepted : 31 Oktober 2023

ABSTRACT

The problem of crime is increasingly widespread in Indonesia. The crime rate in Jabodetabek is the second highest in Indonesia. Jabodetabek faces the problem of high crime rates which can affect other problems such as social problems, poverty, education, and others. This research aims to find out what factors cause high levels of crime and how to overcome high levels of crime, which is one of the law enforcement efforts, namely prevention before crime occurs (preventive efforts). In this research, the data contains outliers, so the appropriate method for this research is quantile regression. Quantile regression is useful for overcoming the problem of unfulfilled assumptions in classical regression, namely the phenomenon of heteroscedasticity and quantile regression can model data that contains outliers. The approach to the quantile regression method is to separate or divide the data into certain parts or quantiles where it is suspected that there are differences in estimated values. In this study five quantiles were used, namely 0.05; 0.25; 0.50; 0.75; and 0.95. From the analysis results, it is known that a good and stable parameter estimation model is found at the 0.95 quantile and it can be concluded that the variables that have a significant influence on the crime rate are at the 0.05 quantile; 0.75 ; and 0.95. Meanwhile, at the 0.25 and 0.50 quantiles, there are no variables that have a significant effect due to the possible influence of other factors that are not included in the model. The pattern of changes between variables in the model shows a linear relationship when the data is below the 0.05 quantile; 0.75; and 0.95.

Keywords: *Crime Rate, Heteroscedasticity, LAD, Quantile Regression*

I. PENDAHULUAN

Kejahatan merupakan permasalahan sosial yang sering terjadi di Indonesia. Menurut BPS (2022) dalam kurun waktu 4 tahun terakhir dari 2018-2021 jumlah tindak kejahatan cenderung mengalami penurunan, tetapi tingkat kejahatan per 100.000 penduduk (*crime rate*) di Indonesia masih dalam kategori yang tinggi. Tingkat kejahatan merupakan angka yang menggambarkan kerawanan suatu kejahatan pada suatu wilayah dan dalam waktu tetentu. Sedangkan, definisi kejahatan itu sendiri merupakan perbuatan individu yang melanggar dan bertentangan dengan apa yang sudah ditetapkan dalam hukum yang berlaku yakni Kitab Undang-Undang Hukum Pidana (KUHP). Faktor individu melakukan tindak kejahatan yaitu permasalahan sosial, rendahnya ekonomi, pengangguran, lingkungan, pendidikan, serta taraf kesejahteraan. Jabodetabek merupakan peringkat kedua dengan tingkat kejahatan tertinggi di Indonesia. Tingkat kejahatan di Jabodetabek tercatat pada Kepolisian Daerah Metropolitan Jakarta Raya (Polda Metro Jaya), dimana Jabodetabek merupakan kota besar sebagai pusat perekonomian masyarakat, yang termasuk tinggi kepadatan penduduk karena banyak masyarakat yang melakukan urbanisasi, sehingga kurangnya lapangan pekerjaan menyebabkan menjamurnya kemiskinan. Oleh sebab itu, masyarakat tidak mampu memenuhi kebutuhan hidupnya yang dapat memicu tindak kejahatan seseorang untuk mencuri, merampok, dan tindak kejahatan yang lainnya.

Analisis regresi merupakan salah satu metode statistika yang digunakan sebagai alat analisis untuk menyelesaikan suatu permasalahan. Permasalahan dalam penelitian ini yaitu tingginya tingkat kejahatan di Jabodetabek yang dapat di selesaikan menggunakan analisis regresi. Penelitian ini bertujuan untuk mengetahui faktor apa saja yang menyebabkan tingginya tingkat kejahatan dan mengatasi tingginya tingkat kejahatan merupakan salah satu upaya penegakan hukum yakni pencegahan sebelum terjadinya kejahatan (upaya preventif). Analisis regresi yakni hubungan antara satu variabel dependen dengan satu atau dua variabel independen (Gujarati, 2006). Untuk membuat model regresi perlu dilakukan estimasi terhadap parameter regresi yaitu menggunakan Metode Kuadrat Terkecil (MKT) atau *Ordinary Least Square*

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

(OLS). Estimasi dengan MKT dilakukan untuk meminimumkan jumlah kuadrat sisaan yaitu terhadap masing-masing parameter yang disamakan dengan nol. Jika memenuhi asumsi-asumsi regresi klasik, maka metode ini akan menghasilkan nilai estimasi parameter yang berifat *Best Linear Unbiased Estimator* (BLUE). Data pada penelitian ini bersumber dari publikasi BPS (2023) apabila dilihat lebih detail, akan terlihat bahwa data mengandung *outlier*. *Outlier* ini mengindikasikan bahwa datanya mengalami pelanggaran asumsi. Apabila terjadi pelanggaran asumsi klasik maka parameter yang dihasilkan tidak bersifat BLUE. Menurut Irevanie (2017) regresi median merupakan regresi yang dilakukan untuk membagi data menjadi dua bagian untuk meminimumkan error atau disebut metode *Least Absolute Deviation* (LAD). Perkembangan dari regresi median ini adalah regresi kuantil dimana regresi yang membagi data menjadi lebih dari dua bagian karena pada regresi median hanya membagi data menjadi dua bagian.

Metode untuk menyelesaikan permasalahan *outlier* pada data tingkat kejahatan dapat dilakukan dengan menggunakan metode regresi kuantil. Di karenakan terdapat *outlier* yang menyebabkan adanya pelanggaran asumsi dan terjadinya gejala heteroskedastisitas atau data tidak normal karena keberagaman data. Metode regresi kuantil merupakan salah satu metode regresi yaitu dengan pendekatan untuk memisahkan maupun membagi data menjadi beberapa bagian atau kuantil tertentu dimana karena dicurigai terdapatnya perbedaann nilai dugaan pada data. Regresi kuantil pertama kali diperkenalkan oleh Koenker dan Basset (1978), dimana regresi kuantil bertujuan untuk meminimumkan error yang tidak simetris dan dapat menghilangkan gejala heteroskedastisitas pada data (Koenker & Hallock, 2001).

II. METODE PENELITIAN

A. Data dan Peubah Penelitian

Penelitian ini menggunakan data sekunder. Data yang digunakan dalam penelitian ini adalah jumlah tindak kejahatan (satuannya adalah jumlah) kemudian diolah menjadi tingkat kejahatan (menggunakan rumus yaitu jumlah tindak kejahatan dibagi dengan jumlah penduduk di suatu wilayah dan dikalikan 100.000 penduduk), hal ini menjadikan satuan tingkat kejahatan adalah jumlah kejahatan setiap 100.000 penduduk. Tingkat kejahatan di Jabodetabek tahun 2022 terdiri dari 14 Kabupaten/Kota. Data bersumber dari BPS Provinsi DKI Jakarta, BPS Provinsi Banten, dan BPS Provinsi Jawa Barat Dalam Angka tahun 2023. Variabel penelitian yang digunakan merupakan skala numerik yakni Tingkat Kejahatan (Y), Tingkat Pengangguran Terbuka (X_1), Indeks Pembangunan Manusia (X_2), Rasio Jenis Kelamin (X_3), dan Rata-rata Lama Sekolah (X_4).

B. Teknik Analisis Data

Pada penelitian ini teknik analisis data yang digunakan adalah analisis statistika deskriptif dan analisis regresi kuantil dengan bantuan *software Rstudio* dan *Geoda.* Adapun langkah-langkah dalam penelitian ini sebagai berikut:

1. Melakukan statistik deskriptif dan eksplorasi data untuk melihat ringkasan data penelitian.
2. Melakukan uji Heteroskedastisitas.

Heteroskedastisitas merupakan suatu kondisi atau gejala dimana sebaran atau variansi dari data hasil pengamatannya tidak konstan di sepanjang observasi. Menurut Ghozali (2009:49) yang dikutip oleh Andriani, S (2017), misalkan terdapat model regresi linear dengan *n*-variabel independen:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_n X_{ni} + e_i$$

Dimana:

Y_i : nilai pengamatan variabel dependen ke-i
α : konstanta
$\beta_{1,2,n}$: Parameter model regresi ke-1,2,..n
$X_{n,i}$: nilai pengamatan ke-i pada variabel independen ke-n
ε_i : nilai galat/sisaan pengamatan ke-i

Diasumsikan error variansi atau σ_i^2 adalah sebagai berikut:

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi}) \approx \sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi}$$

Dimana:

σ_i^2 : error variansi (ketidaksamaan variansi dari vektor error ($var(\varepsilon_i)$))
a_j : nilai konstanta a_j bernilai nol (j=1,2,3..m)
Z_{mi} : konstanta regresi bernilai satu yaitu variabel independen (1,2,..m)

Dimana σ_i^2 merupakan fungsi linear dari Z jika α_1=α_2=α_3=…=α_1=0, maka $\sigma_i^2 = \alpha_1$ merupakan konstanta untuk menguji apakah σ_i^2 homoskedastisitas.

Menurut Ghozali (2009:49) yang dikutip oleh Andriani, S (2017) langkah-langkah yang digunakan untuk mendeteksi heteroskedastisitas dengan menggunakan uji *Breusch Pagan Godfrey* sebagai berikut :

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 1 Nomor 5 Tahun 2023 Halaman 377-384
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

a. Melakukan regresi OLS dengan persamaan $Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_n X_{ni} + e_i$ sehingga memperoleh nilai error atau residualnya dan mencari $\sigma_i^2 = \frac{\sum e_i^2}{n}$
Dimana:
σ_i^2 : error variansi (ketidaksamaan variansi dari vektor error ($var(\varepsilon_i)$))
e_i : nilai error atau residual pada pengamatan
n : banyaknya pengamatan

b. Mencari p_i yang didefinisikan sebagai $p_i = \frac{e_i^2}{\sigma^2}$
Dimana:
p_i : variabel dependen baru berupa p pada pengamatan ke-i
e_i^2 : nilai error atau residual pada pengamatan ke-i
σ^2 : error variansi

c. Regresi p_i terhadap variabel Z sebagai berikut :
$p_i = \alpha_0 + \alpha_1 Z_i + v_i$
Dimana:
p_i : varaiabel dependen baru p pada pengamatan ke-i
α_0, α_1: koefisien regresi
Z_i : variabel independen
v_i : error variansi pada regresi baru ke-i

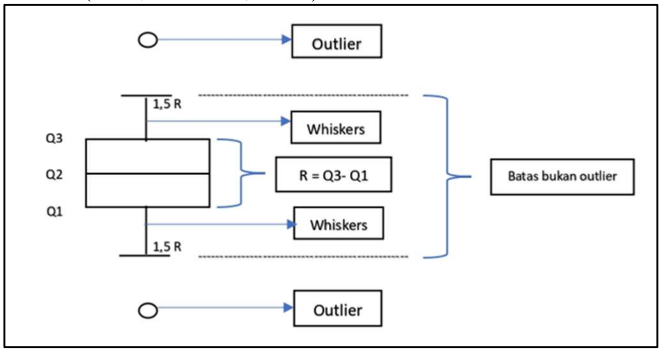
d. Mendapatkan ESS(*Explained sum of square*) dari persamaan hasil regresi p_i dengan persamaan $\phi = \frac{1}{2}(ESS)$. Apabila residual pada persamaan terdistribusi normal maka $\frac{1}{2}(ESS)$ akan mengikuti distribusi chi-square (χ^2) sebagai berikut ini: $\phi = \frac{1}{2}(ESS) \approx \chi^2_{df}$
Dimana:
ESS : jumlah kuadrat yang dijelaskan atau jumlah kuadrat deviasi nilai prediksi dari nilai rata-rata variabel dependen.

Langkah-langkah hipotesisnya adalah sebagai berikut:

a. Perumusan Hipotesis
H_0: $\sigma(\varepsilon_i) = \sigma^2$ (variansi error bersifat homoskedastisitas)
H_1 : $\sigma(\varepsilon_i) \neq \sigma^2$ (variansi error tidak bersifat homoskedastisitas /heteroskedastisitas)
b. Menentukan taraf signifikansi α
c. Statistik Uji *chi-square* χ^2_{df}
d. Kriteria Pengujian : H_0 ditolak jika $\phi_{hitu} > \chi^2_{df,tabel}$: df =(m-1)

3. Mengidentifikasi adanya *outlier*

Terdapatnya *outlier* merupakan salah satu hal kasus yang dapat membuat nilai taksiran pada metode regresi linear menjadi kurang akurat. Tetapi pada metode regresi kuantil bisa membatasi pengaruh dari *outlier* tersebut sehingga menghasilkan penaksiran pada model yang lebih akurat. Pada penelitian ini untuk mendeteksi adanya *outlier* menggunakan *boxplot*, *boxplot* merupakan suatu *box* atau kotak berbentuk bujur sangkar yang menampilkan distribusi data berdasar lima rangkuman (nilai minimum, kuartil pertama, median, kuartil ketiga, dan nilai maksimum) sehingga dapat memudahkan mendeteksi *outlier*(Janu, M. N. P., 2020).



Gambar 1. *Boxplot*

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Dalam *boxplot* terdapat nilai jangkauan (IQR/R) yang didefinisikan sebagai selisih kuartil 1 terhadap kuartil 3, atau $IQR = Q_3 - Q_1$. Data *outlier* dapat ditentukan yaitu nilai yang kurang dari 1,5R terhadap Q_1 dan nilai *outlier* yang kurang dari 1,5R terhadap Q_3 diidentifikasikan data *outlier* (Soemartini, 2007).

4. Model Regresi Kuantil

Metode Regresi kuantil adalah perkembangan dari model regresi pada kuantil bersyarat dimana distribusi kuantil dari variabel dependen yang dinyatakan sebagai fungsi dari kovariat yang akan diamati. Pendekatan regresi kuantil yaitu membagi data menjadi beberapa bagian atau kuantil yang diduga mempunyai perbedaan nilai parameter pada kuantil tertentu. Pendekatan ini dapat menduga fungsi kuantil dari sebaran bersyarat variabel dependen pada setiap nilai kuantil sesuai dengan kuantil yang diinginkan untuk diteliti (Chen, 2005). Pada penelitian ini dilakukan pemodelan pada data Tingkat Kejahatan (Y), Tingkat Pengangguran Terbuka (X_1), Indeks Pembangunan Manusia (X_2), Rasio Jenis Kelamin (X_3), dan Rata-rata Lama Sekolah (X_4) yang berskala numerik menjadi model regresi kuantil dengan persaamaan berikut:

$$Q_\tau(Y|X) = \beta_0(\tau) + \beta_1(\tau)X_1 + \beta_2(\tau)X_2 + \beta_3(\tau)X_3 + \beta_4(\tau)X_4$$

Keterangan:
$Q_\tau(Y|X)$: Fungsi kuantil ke - τ dari variabel Y dengan syarat X
τ : Indeks kuantil dengan $\tau \epsilon (0,1)$
$\beta_i(\tau)$: Koefisien regresi ke-*i* pada kuantil ke - τ, dengan i= 0,1,2,..,n

Regresi kuantil sangat disarankan untuk menganalisis sejumlah data yang tidak simetris dan memiliki distribusi yang tidak homogen yang sifatnya *robust* terhadap data *outlier*. Menurut Koenker (2005:14) pendekatan *direct, rank-score,* dan *resampling* digunakan untuk menaksirkan interval pada regresi kuantil. Regresi kuantil berguna untuk data yang tidak homogen dan tidak berbentuk simetris serta terdapat ekor dalam sebarannya (Balami, 2017). Pada regresi linear berlaku E(y|x_i) =$x_i^T\beta$, sedangkan dalam regresi kuantil $Q_\tau(y|x_i) = x_i^T\beta_\tau$ diuraikan menjadi persamaan berikut:

$$Y_i = \beta_{0,\tau} + \beta_{0,\tau} X_{1,i} + \ldots + \beta_{k,\tau} X_{k,i} + \varepsilon_{i,\tau} \text{ , i= 1, 2, …, n} \quad (1)$$

Dimana:
Y_i : nilai pengamatan variabel dependen ke-i
$\beta_{0,\tau}$: Parameter model regresi pada kuantil ke- τ
$\beta_{k,\tau}$: koefisien variabel independen ke-k pada kuantil ke- τ
$X_{k,i}$: nilai pengamatan ke-i pada variabel independen ke-k pada kuantil ke- τ
$\varepsilon_{i,\tau}$: nilai galat/sisaan pengamatan ke-i pada kuantil ke-τ
Atau dalam notasi matriks pada kuantil ke- τ dapat ditulis sebagai berikut:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \cdots & \cdots & \cdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_{0\tau} \\ \beta_{1\tau} \\ \vdots \\ \beta_{p\tau} \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_{1\tau} \\ \varepsilon_{2\tau} \\ \vdots \\ \varepsilon_{n\tau} \end{bmatrix} \quad (2)$$

Bentuk umum model regresi kuantil dari matriks pada persamaan berikut ini menurut Buhai (2015):

$$Y_i = X\beta_\tau + \varepsilon_\tau \text{ , } 0 < \tau < 1 \quad (3)$$

Misalkan, terdapat data $y_1, y_2, y_3, \ldots y_n$ dan τ merupakan fungsi distribusi kumulatif atau kuantil ke - τ dari y sehingga dapat dituliskan sebagai berikut:

$$F_y\ (y) = F_y = P(Y \leq y) \quad (4)$$

$$Q_\tau(Y) = \inf \{y : F(y) \geq \tau\} \quad (5)$$

Kemudian terdapat kuantil ke – τ dari F(y) yang diperoleh dengan meminimumkan fungsi tersebut terhadap Q sebagai berikut:

$$\tau \int_{y>q} |y - q|\, dF_Y(y) + (1-\tau) \int_{y>q} |y - q|\, dF_Y\,(y) \quad (6)$$

Apabila Y adalah fungsi X yang diketahui memiliki fungsi probabilitas $F_{Y|X}$(y), maka kuantil ke - τ dari fungsi tersebut dapat dituliskan menjadi $Q_{Y_i}\ (\tau|x)$ merupakan fungsi dari X dan diselesaikan dengan persamaan berikut ini:

$$min_q \tau \int_{i=1,y>q}^{n} |y - q|\, dF_Y(y) + (1-\tau) \int_{i=1,y>q}^{n} |y - q|\, dF_Y\,(y) \quad (7)$$

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 1 Nomor 5 Tahun 2023 Halaman 377-384
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Di dalam regresi kuantil terdapat fungsi kuantil bersyarat ke- τ yang mempertimbangkan penduga β_τ sehingga memperoleh solusi untuk permasalahan tersebut yang ditulis sebagai berikut:

$$\beta(\tau) = min_\beta\{\tau \sum_{i=1,y>x}^{n} |y - \boldsymbol{x}^T\boldsymbol{\beta}| + (1 - \tau)\sum_{i=1,y>x}^{n} |y - \boldsymbol{x}^T\boldsymbol{\beta}|\} \tag{8}$$

Metode algoritma simpleks yang telah dikembangkan oleh Barrodale & Robert pada tahun 1974 merupakan salah satu metode estimasi parameter secara numerik untuk regresi kuantil. Pada metode algoritma simpleks ini akan memberikan hasil yang lambat dan memberikan solusi pada beberapa jenis data khususnya pada data dengan jumlah *outlier* yang besar (Chen & Wei, 2005).

5. Uji Signifikansi Parameter

Dalam analisis regresi kuantil, uji t (parsial) merupakan uji signifikansi parameter yang dilakukan pada analisis regresi kuantil (Gujarati, 2004). langkah dalam uji t sebagai berikut :

a. Perumusan Hipotesis
   H_0 : β_i $(\tau) = 0$ (variabel independen(X) tidak berpengaruh secara signifikan terhadap Tingkat Kejahatan (Y))
   H_1 : β_i $(\tau) \neq 0$ (variabel independen(X) berpengaruh secara signifikan terhadap Tingkat Kejahatan (Y))
b. Statistik Uji : $T_{hitung} = \frac{\hat{\beta}_i - \beta_i}{s_{\hat{\beta}_i}}$
   Keterangan:
   $\hat{\beta}_i$: Nilai estimasi parameter β_i
   β_i : nilai parameter di hipotesis H_0
   $S_{\hat{\beta}_i}$: Standar deviasi $\hat{\beta}_i$
c. Kriteria Pengujian : Tolak H_0, jika $t_{hitung} \geq t_{tabel}$ atau $p_{value} < \alpha$. jika H_0 maka suatu faktor tersebut berpengaruh secara signifikan terhadap Variabel Tingkat Kejahatan.

6. Uji Koefisien Determinasi

Menurut Priyatno (2008) uji koefisien determinasi berguna untuk mengukur seberapa besar persentase variasi pada variabel independen pada model linear berganda dalam menjelaskan variasi variabel dependen. Dalam regresi kuantil, koefisien determinasi dapat dipakai adalah R^2pseudo (Koenker, 1999). Nilai dari R^2 dapat diperoleh dengan menggunakan persamaan:

$$R^2(\tau) = 1 - \frac{\sum_{y_i \geq \hat{y}_i} |y_i - \hat{y}_i| + \sum_{y_i \geq \hat{y}_i} (1-\tau)|y_i - \hat{y}_i|}{\sum_{y_i \geq \bar{y}_i} |y_i - \bar{y}_i| + \sum_{y_i \geq \bar{y}_i} (1-\tau)|y_i - \bar{y}_i|} \tag{9}$$

Dimana $\hat{y}_i = \alpha_\tau + \beta_\tau x$ merupakan kuantil ke $-\tau$ yang sesuai dengan observasi *i*, sedangkan $\bar{y}_i = \beta_\tau$ merupakan nilai dari intersep model yang sesuai.

III. HASIL DAN PEMBAHASAN

A. Statistik Deskriptif

Statistik deskriptif berguna untuk melihat ringkasan data secara umum mengenai variabel penelitian yang terdapat dalam data penelitian. Tabel 1 menyajikan statistik deskriptif untuk setiap variabel penelitian.

Tabel 1. Statistik deskriptif variabel penelitian

Variabel	Minimum	Q1	Median	Rata-Rata	Q3	Maksimum
Y	18,32	46,75	114,34	991,42	214,27	12235,09
X1	5,630	7,115	7,960	8,107	8,725	10,780
X2	71,20	75,71	81,33	79,19	82,37	85,21
X3	100,1	101,5	102,0	102,3	103,0	105,6
X4	8,340	9,805	10,985	10,636	11,515	11,840

(Sumber: *Output Software Rstudio*).

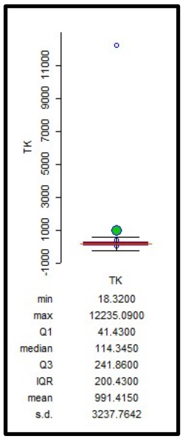
Pada Tabel 1 berdasarkan data publikasi BPS (2023), terlihat bahwa pada variabel Tingkat Kejahatan (Y) dengan tingkat kejahatan terendah yaitu pada Kota Depok sebesar 18,32 yang setara dengan 18 artinya adalah terdapat 18 kasus kejahatan setiap 100.000 penduduk dan tingkat kejahatan tertinggi terdapat pada Kabupaten Kepulauan Seribu yaitu sebesar 12.235,09 yang setara dengan 12.235 kasus kejahatan setiap 100.000 penduduk. Hal ini mengindikasikan terdapatnya *outlier* karena nilai observasi nya jauh berbeda dengan observasi lainnya. Dimana tingginya tingkat kejahatan menggambarkan tingginya pula kerawanan pada suatu daerah tersebut, dan sebaliknya.

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 1 Nomor 5 Tahun 2023 Halaman 377-384
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

B. Identifikasi *Outlier* dan Gejala Heteroskedastisitas

Mengidentifikasikan *outlier* pada data penelitian ini menggunakan *boxplot* yaitu pada Gambar 2 berikut ini.



(Sumber: *Output Software Geoda*).

Gambar 2. *Outlier* pada Tingkat Kejahatan di Jabodetabek

Berdasarkan Gambar 2 terlihat bahwa terdapatnya *outlier* di dalam *boxplot* yakni pada Kabupaten Kepulauan Seribu sebesar 12235,09 yang menyebabkan munculnya *outlier* pada tingkat kejahatan di Jabodetabek.

Tabel 2. Uji BP

Nilai BP	Df	*p-value*
9.8091	4	0.04377

(Sumber: *Output Software Rstudio*).

Pengujian heteroskedastisitas dalam penelitian ini menggunakan uji *Breusch Pagan Godfrey* terlihat pada tabel 2 bahwa nilai BP adalah sebesar 9,8091 dan *p-value* adalah 0,04377. Hal ini menunjukkan tolak $H_{0,}$ karena nilai *p-value* (0,04377) < α(0,05). Dapat disimpulkan bahwa terjadi gejala heteroskedastisitas.

C. Estimasi Parameter Regresi Kuantil

Dalam pemodelan regresi kuantil digunakan bantuan *software Rstudio* untuk memperoleh estimasi parameter. Hasil estimasi parameter untuk model regresi kuantil dapat dilihat pada tabel 3 pada kuantil 0, 05; 0,25; 0,50; 0,75; dan 0,95.

Tabel 3. Estimasi Nilai Parameter

Kuantil (τ)	Parameter				
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
0,05	12109.538	-53.5810	-24.9043	-89.4909	-45.6289
0,25	16398.291	-32.6710	-17.9877	-129.911	-124.412
0,50	21008.907	-23.2493	-19.9509	-168.952	-172.295
0,75	424757.58	-1615.324	-1691.722	-2814.730	1220.416
0,95	429006.85	-1455.400	-1777.928	-2852.694	1723.533

(Sumber: *Output Software Rstudio*).

Berdasarkan estimasi parameter pada tabel 2, diperoleh model regresi kuantil untuk kuantil 0,05; 0,25; 0,50; 0,75; dan 0,95 adalah sebagai berikut :

$y_{0,05}$ = 12109.538 - 53.5810X_1 - 24.9043 X_2 - 89.4909 X_3 - 45.6289 X_4

$y_{0,25}$ = 16398.291 - 32.6710 X_1 - 17.9877 X_2 - 129.911 X_3 - 124.412 X_4

$y_{0,50}$ = 21008.907 - 23.2493 X_1 - 19.9509 X_2 - 168.952 X_3 - 172.295 X_4

$y_{0,75}$ = 424757.58 - 1615.324 X_1- 1691.722 X_2- 2814.730 X_3 +1220.416 X_4

$y_{0,95}$ = 429006.85- 1455.400 X_1 - 1777.928 X_2- 2852.694 X_3+1723.533 X_4

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume 1 Nomor 5 Tahun 2023 Halaman 377-384
ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

Tabel 4. Nilai P-Value setiap Parameter

Kuantil (τ)	Parameter				
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
0,05	0.00047*	0.00442*	0.16174	0.00103*	0.48824
0,25	0.04212*	0.47116	0.72689	0.05015	0.53508
0,50	0.06659	0.72120	0.78954	0.07384	0.55414
0,75	0.00032*	0.00772*	0.01252*	0.00150*	0.57583
0,95	0.00000*	0.00000*	0.00000*	0.00000*	0.03057*

(Sumber: *Output Software Rstudio*).

Berdasarkan Tabel 4 menunjukkan nilai *p-value* untuk setiap estimasi parameter model regresi kuantil untuk kuantil 0,05; 0,25; 0,50; 0,75; dan 0,95. Diperoleh variabel yang signifikan pada 0,05 adalah X_1 dan X_3, pada kuantil 0,075 adalah X_1, X_2, dan X_3, dan pada kuantil ke 0,95 seluruh variabel signifikan. Sedangkan pada kuantil 0,25 dan 0,50 tidak terdapat variabel yang signifikan, hal ini kemungkinan terjadi dikarenakan pengaruh faktor-faktor lain yang tidak terdapat dalam penelitian yang berpengaruh terhadap tingkat kejahatan.

D. Koefisien Determinasi

Setiap kuantil memiliki model regresi kuantil masing-masing, maka masing-masing kuantil memiliki koefisien determinasi tersendiri. Koefisien determinasi ini menunjukkan kebaikan model tersebut. Pada Tabel 5 menunjukkan koefisien determinasi pada masing-masing kuantil.

Tabel 5. Koefisien Determinasi

Koefisien Determinasi	Kuantil 0,05	0,25	0,50	0,75	0,95
R^2	41,48%	34,23%	29,18%	35,99%	84,02 %

(Sumber: *Output Software Rstudio*).

Berdasarkan Tabel 5 menjelaskan bahwa koefisien determinasi pada setiap kuantil berbeda. Kuantil 0,95 menjadi kuantil dengan koefisien determinasi tertinggi yaitu 84,02% yang memiliki makna bahwa variabel Tingkat Pengangguran Terbuka (X_1), Indeks Pembangunan Manusia (X_2), Rasio Jenis Kelamin (X_3), dan Rata-rata Lama Sekolah (X_4) dapat dijelaskan variabel Tingkat Kejahatan (Y) paling baik pada data dibawah kuantil 0,95 yaitu 84,02% dan sisanya sebesar 15,98% disebabkan oleh variabel lain yang tidak diteliti paada penelitian ini.

IV. KESIMPULAN

Dengan menggunakan regresi kuantil menunjukkan pemodelan yang baik untuk mengatasi permasalahan data yang tidak simetris, gejala heteroskedastisitas, dan adanya *outlier*. Setiap kuantil mempunyai variabel yang berpengaruh secara signifikan yang berbeda. Pada kuantil 0,05 variabel yang berpengaruh signifikan yaitu variabel Tingkat Pengangguran Terbuka (X_1) dan Rasio Jenis Kelamin (X_3) , pada kuantil 0,25 dan 0,50 tidak terdapat variabel yang berpengaruh secara signifikan, pada kuantil 0,75 Tingkat Pengangguran Terbuka (X_1), Indeks Pembangunan Manusia (X_2),dan Rasio Jenis Kelamin (X_3), pada kuantil 0,95 variabel yang berpengaruh secara signifikan adalah Tingkat Pengangguran Terbuka (X_1), Indeks Pembangunan Manusia (X_2), Rasio Jenis Kelamin (X_3), dan Rata-rata Lama Sekolah (X_4). Dari hasil analisis yang telah dilakukan diperoleh kebaikan model pada nilai koefisien determinasi (R^2) masing-masing kuantil, model regresi kuantil ke-0,95 menunjukkan model baik dan stabil dan dapat disimpulkan bahwa variabel yang memiliki pengaruh yang signifikan terhadap tingkat kejahatan pada kuantil 0,05; 0,75 ; dan 0,95. Sedangkan pada kuantil 0,25 dan 0,50 tidak terdapat variabel yang berpengaruh secara signifikan terjadi karena adanya kemungkinan pengaruh faktor lain yang tidak terdapat dalam model. Pola perubahan antar variabel pada model yang menunjukkan adanya hubungan linear ketika data dibawah kuantil 0,05; 0,75; dan 0,95.

Pada penelitian ini, peneliti menggunakan model linear dari regresi kuantil. Selain model linear adapun model nonlinear yang dapat digunakan pada regresi kuantil. Sehingga, peneliti menyarankan untuk penelitian selanjutnya dapat menggunakan metode *nonlinear quantile regression*.

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94

ISSN(*Print*) : 3025-5511
ISSN(*Online*): 2985-475X

DAFTAR PUSTAKA

Andriani. Siska, (2017), "Uji Park Dan Uji Breusch Pagan Godfrey Dalam Pendeteksian Heteroskedastisitas Pada Analisis Regresi", *Al-Jabar:Jurnal Pendidikan Matematika*. Vol.8, No.1, 63-72.

Badan Pusat Statistik, (2023). *Provinsi Banten Dalam Angka* (*Banten Province in Figures).* Banten: Badan Pusat Statistik Provinsi Banten.

Badan Pusat Statistik, (2023). *Provinsi DKI Jakarta Dalam Angka* (*DKI Jakarta Province in Figures),* DKI Jakarta: Badan Pusat Statistik Provinsi DKI Jakarta.

Badan Pusat Statistik, (2023). *Provinsi Jawa Barat Dalam Angka* (*Jawa Barat Province in Figures),* Jawa Barat: Badan Pusat Statistik Provinsi Jawa Barat.

Badan Pusat Statistik, (2022). *Statistik Kriminal 2022,* Badan Pusat Statistik.

Balami, A.M, (2017), "Parameters Estimation of Quantille Regression in Dengue Fever Case in Surabaya", *Thesis*, Surabaya: Institut Teknologi Sepuluh Nopember.

Buhai, I. S, (2005), "Quantille Regression: Overview and Selected Applications", *Ad Astra Journal*, Vol 4, 1-17.

Chen, C & Wei, Y, (2005), "Computational Issues for Quantille Regression", *The Indian Journal of Statistics*. Vol 67, No.2, pp. 99-417.

Ghozali, Imam. (2009). *Ekonometrika Teori, Konsep, dan Aplikasi dengan SPSS 17*. Semarang: Badan Penerbit Universitas Diponegoro.

Gujarati, Damodar N, (2006). *Ekonometrika Dasar*. Jakarta: Erlangga.

Irevanie, R.S, (2017). "Perbandingan Metode Quantille Regression (QR) dan Geographically Weighted Regression (GWR) Pada Data Angka Harapan Hidup di Indonesia", *Thesis*. Surabaya: Institut Teknologi Sepuluh Nopember.

Janu, M. N. P, (2020). "Analisis Hasil Ujian Nasional Tahun 2016-2019 Tingkat Sekolah Menengah Pertama Dengan Pendekatan Sains Data"*, Thesis*. Yogyakarta: Universitas Sanata Dharma.

Koenker, R, (2005). *Quantille Regression, Series Econometric Society Monographs.* Cambridge University Press.

Koenker, R. & Bassett, G. (1978), "Regression Quantiles", *Journal Econometrica*, Vol. 46, No. 1, pp. 33-50.

Koenker, R & K.Hallock, (2001), "Quantile Regression", *Journal of Economic Perspective.* Vol. 15, pp. 143-156.

Koenker, R & Machado, J.A.F, (1999), "Goodness of Fit and Related Inference Processes for Quantille Regression", *Journal of the American Statistical Association,* Vol.94, No. 448, pp. 1296 -1310.

Priyatno, D, (2008). *Mandiri Belajar SPSS – Bagi Mahasiswa dan Umum.* Yogyakarta: Media Kom.

Soemartini, (2007). *Pencilan (Outlier)*. Bandung: Universitas Padjajaran.

DOI: https://doi.org/10.24036/ujsds/vol1-iss5/94