Pemodelan *Quantile Regression* untuk Menentukan Faktor-Faktor Penyebab Penyakit Tuberkulosis Paru di Kabupaten Tasikmalaya

Hasni Khotimah*, Suwanda

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

* hasnikhotimah74@gmail.com, suwanda@unisba.ac.id

Abstract. In regression analysis there are classical assumptions that must be met, namely normality, homoscedasticity, and non-autocorrelation, while in this case there are outliers. The existence of outliers in the data will result in a large and inhomogeneous error variance, so that the estimator generated on data containing outliers will not be Best Linear Unbiased Estimator (BLUE). However, if there are outliers or other assumptions are violated, the good nature of this estimator will be disturbed. In this study, it is known that there are outliers in the tuberculosis data and the factors that influence it. To overcome this outlier problem, robust regression can be used, one of which is the Quantile Regression method. The data used in this study were secondary data obtained from the report of the Puskesmas to the Health Office and Population Control in Tasikmalaya Regency in 2019. The area used as the object of research was 39 sub-districts in Tasikmalaya Regency. In this study, modeling was carried out using 2 quantile levels, namely Quantile Regression (0.33th and 0.67). So as to provide information when cases of Tuberculosis in Tasikmalaya Regency are at risk (low, medium and high). In this study, it was found that at the level of the 0.33 quantile only the clean water variable (X_2) affected the pulmonary TB incident rate, while at the 0.67 quantile level there was no significant variable. Quantile Regression can be applied to determine the relationship between the number of cases of pulmonary tuberculosis with the percentage of proper sanitation (healthy latrines), the percentage of clean water availability and the percentage of non-smoking habits in Tasikmalaya Regency in 2019.

Keywords: *Quantile Regression, Outliers, Robust.*

Abstrak. Dalam analisis regresi terdapat asumsi klasik yang harus dipenuhi yaitu normalitas, homoskedastisitas, dan non autokorelasi, sedangkan dalam kasus ini terdapat data pencilan (outlier). Adanya pencilan/outlier pada data akan mengakibatkan varians error yang besar dan tidak homogen, sehingga estimator yang dihasilkan pada data yang mengandung outlier tidak akan bersifat Best Linear Unbiased Estimator (BLUE). Akan tetapi jika terdapat outlier atau ada asumsi lain yang dilanggar, sifat baik dari penaksir ini akan terganggu. Pada penelitian ini diketahui adanya outlier dalam data penyakit tuberkulosis dan faktor-faktor yang mempengaruhi nya Untuk mengatasi masalah outlier ini dapat menggunakan regresi yang robust, salah satunya metode regresi kuantil. Data yang digunakan dalam penelitian ini data sekunder yang diperoleh dari laporan Puskesmas kepada Dinkes dan Pengendalian Penduduk di Kabupaten Tasikmalaya tahun 2019. Wilayah yang dijadikan sebagai objek penelitian 39 kecamatan di Kab Tasikmalaya. Pada penelitian ini, pemodelan dilakukan dengan menggunakan 2 level kuantil, yaitu regresi kuantil (ke-0.33 dan ke-0.67). Sehingga memberikan informasi kapan kasus Tuberkulosis di Kabupaten Tasikmalaya beresiko (rendah, sedang dan tinggi). Pada penelitian ini diperoleh bahwa pada level kuantil ke-0.33 hanya variabel air bersih (X_2) yang mempengaruhi incident rate TB paru, sedangkan pada level kuantil ke 0.67 tidak ada variabel yang signifikan. Regresi kuantil dapat diaplikasikan untuk menentukan hubungan antara jumlah kasus tuberkulosis paru dengan persentase sanitasi yang layak (jamban sehat), persentase ketersediaan air bersih dan persentase kebiasaan tidak merokok di Kabupaten Tasikmalaya tahun 2019.

Kata Kunci: *Regresi Kuantil, Pencilan, Robust.*

Corresponding Author
Email: suwanda@unisba.ac.id

A. Pendahuluan

Tuberkulosis (TB) merupakan penyakit menular yang menyebabkan masalah kesehatan terbesar kedua di dunia setelah HIV. Penyakit ini disebabkan oleh basil dari *Mycrobacterium tuberculosis*. Indonesia sendiri termasuk lima besar negara dengan jumlah pengidap TB terbanyak di Asia tenggara pada tahun 2012. Gejala yang dialami pada pasien TBC paru adalah batuk berdahak selama 2 minggu atau lebih. Batuk ini disertai dengan gejala tambahan seperti dahak yang bercampar dengan darah, batuk darah, badan lemas, nafsu makan menurun, sesak nafas, berat badan menurun, malaise, malam hari berkeringat tanpa kegiatan fisik, demam meriang lebih dari satu bulan

Dinas Kesehatan dan Pengendalian Penduduk Kabupaten Tasikmalaya merupakan instansi pemerintahan di Kabupaten Tasikmalaya yang bertanggung jawab dalam bidang kesehatan. Dinas Kesehatan dan Pengendalian Penduduk Kabupaten Tasikmalaya memiliki beberapa tujuan, salah satunya yaitu meningkatkan pencegahan dan penanggulangan penyakit. Oleh karena itu, informasi mengenai penyakit khususnya tentang penyakit tuberkulosis paru (TB Paru) di Kabupaten Tasikmalaya diperoleh dari Dinas Kesehatan dan Pengendalian Penduduk Kabupaten Tasikmalaya.

Upaya penanggulangan kasus TB Paru dapat dilakukan dengan melakukan kontrol terhadap faktor-faktor yang diduga mempengaruhinya. Untuk itu, perlu dilakukan suatu pendekatan untuk memodelkan faktor-faktor yang diduga mempengaruhi kasus TB Paru di Kab. Tasikmalaya. Analisis statistik yang dapat digunakan untuk memodelkan hubungan antara kasus TB Paru dengan faktor yang diduga mempengaruhinya adalah analisis regresi.

Pemodelan penyakit TB Paru pada pemodelan ini akan melibatkan variabel prediktor. Pada saat identifikasi data terdeteksi adanya *outlier*, sehingga untuk mengatasinya diperlukan metode yang *robust*. Metode *robust* merupakan metode regresi yang digunakan ketika distribusi dari galat tidak normal dan atau adanya beberapa pencilan yang berpengaruh pada model. Selain itu, dalam penelitian ini akan membagi kejadian penyakit Tuberkulosis berdasarkan faktor-faktor yang mempengaruhinya tersebut kedalam beberapa klasifikasi. Adanya informasi tersebut, akan diketahui kapan risiko TB Paru di Kabupaten Tasikmalaya dikatakan rendah, sedang, ataupun tinggi.

Dalam analisis regresi terdapat asumsi klasik yang harus dipenuhi yaitu normalitas, homoskedastisitas, dan non autokorelasi. Penaksir Ordinary Least Square (OLS) atau kuadrat terkecil mempunyai sifat penduga parameter yang bersifat BLUE *(Best Linier Unbiased Estimator)* jika semua asumsi terpenuhi. Akan tetapi jika terdapat *outlier* atau ada asumsi lain yang dilanggar, sifat baik dari penaksir ini akan terganggu. Untuk mengatasi masalah *outlier* ini dapat menggunakan regresi yang *robust,* salah satunya metode regresi kuantil.

Regresi kuantil *(Quantile Regression)* merupakan salah satu metode yang *robust* terhadap adanya *outlier* dengan memperhatikan berbagai level kuantil, sehingga memberikan informasi kapan kasus Tuberkulosis di Kabupaten Tasikmalaya beresiko rendah, sedang dan tinggi. Regresi kuantil pertama kali dikenalkan oleh Koenker dan Basset pada tahun (1978). Dalam metode ini, akan diperoleh informasi lebih banyak mengenai hubungan antara variabel respon dan variabel prediktor melalui penggunaan nilai kuantil yang dinotasikan dengan $\tau \in [0,1]$.

Pada penelitian ini, penulis akan melakukan pemodelan untuk mengetahui hubungan penyakit tuberkulosis paru dengan faktor-faktor yang diduga mempengaruhinya yaitu Persentase fasilitas sanitasi yang layak (Jamban sehat), Persentase Kepemilikan Air Bersih dan Persentase Kebiasaan Tidak Merokok.

B. Metodologi Penelitian

Analisis Regresi

Analisis regresi adalah teknik dalam statistika yang digunakan untuk mengkaji dan memodelkan hubungan antar variabel, penerapan dari regresi linier dapat digunakan di berbagai bidang dan merupakan teknik statistika yang berkembang secara luas (Montgomery, dkk., 2012).

Apabila Y adalah variabel respon dan X adalah variabel prediktor yang membentuk sebuah model regresi secara umum sebagai berikut:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i \tag{1}$$

Keterangan :
Yi = nilai variabel dependen dalam pengamatan ke-*i*
β_0 = koefisien *intercept*
X_i = observasi ke-I untuk variabel bebas ke-*k*
ε_i = *error* dari model

Regresi *Mean*

Secara umum, metode regresi yang biasa digunakan oleh peneliti merupakan dasar dari teknik regresi, yaitu metode kuadrat terkecil (OLS). Prinsip kerjanya adalah meminimalkan jumlah deviasi kuadrat atau *error* dari observasi yang diamati ke nilai rata-rata yang didefinisikan sebagai berikut:

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - \boldsymbol{X}_i^{\boldsymbol{T}} \boldsymbol{\beta}\right)^2 \tag{2}$$

dimana :
$\boldsymbol{\beta} = \left(\beta_0, \beta_1, \ldots, \beta_p\right)$ dan $\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{T}} = (X_{1i}, X_{2i}, \ldots, X_{pi})$
Persamaan (2.9) akan minimum untuk:

$$\widehat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T y$$

Metode ini harus memenuhi asumsi-asumsi regresi klasik yang ada, seperti asumsi linieritas, asumsi normalitas, asumsi homogenitas, dan asumsi non-autokorelasi, namun dapat terjadi penyimpangan dari asumsi tersebut, sehingga jika salah satu asumsi regresi tersebut tidak terpenuhi, maka penduga metode kuadrat terkecil tidak lagi bisa digunakan.

Model regresi *mean* juga dianggap kurang tepat, karena jika data berbentuk tidak simetris maka nilai *mean* menjadi sangat peka terhadap adanya data *outlier*, sehingga nilai *mean* tidak dapat mewakili bagian lain dari distribusi respon. Adanya *outlier* pada data akan mengakibatkan varians *error* yang besar dan tidak homogen, sehingga estimator yang dihasilkan pada data yang mengandung *outlier* tidak akan bersifat *Best Linear Unbiased Estimator* (BLUE). Oleh karena itu, diperlukan suatu metode *robust* untuk mengatasi permasalahan tersebut, salah satunya dengan menggunakan regresi kuantil.

Regresi Kuantil

Regresi kuantil pertama kali dikenalkan oleh Koenker dan Basset pada tahun (1978). Regresi kuantil yaitu teknik regresi yang menjelaskan hubungan antara variabel respon dan variabel prediktor pada berbagai kuantil, bukan hanya pada ukuran pemusatan (median) variabel respon. Penggunaan nilai kuantil, yang dinotasikan dengan $\tau \in (0{,}1)$ akan menyebabkan lebih banyak informasi yang diperoleh mengenai hubungan antara variabel prediktor dan variabel respon (Mulyani, 2017). Regresi kuantil merupakan suatu teknik dalam analisis regresi yang bersifat *robust*. Regresi kuantil sangat baik digunakan pada sebaran data yang berdistribusi asimetris atau padat pada ujung sebaran data *(truncated distribution)*, karena estimator yang dihasilkan akan lebih efisien (Andriyana, 2015).

Regresi kuantil dibagi menjadi regresi kuantil bersyarat dan regresi kantil tidak bersyarat. Regresi kuantil tidak bersyarat hanya melibatkan variabel respon *Y* sebagai variabel acak dalam bentuk distribusi kumulatif. Dalam regresi kuantil bersyarat variabel respon *Y* bergantung pada himpunan dari variabel *X*, dimana kelebihan dari kuantil bersyarat adalah fleksibilitas dalam menghasilkan model data.

Jika pada regresi linier $E(\boldsymbol{\varepsilon}|\boldsymbol{X}) = 0$ sehingga $\mathrm{E}(\boldsymbol{Y}|\boldsymbol{X}) = \boldsymbol{X}'\boldsymbol{\beta}$ dan untuk mengestimasi parameter pada regresi linier dengan meminimumkan jumlah kuadrat kekeliruan, seperti persamaan berikut:

$$\min_{\boldsymbol{\beta}} \frac{1}{n} \sum\nolimits_{i=1}^{n} (Y_i - \mathbf{X}_i' \boldsymbol{\beta}\ ) \tag{3}$$

Sementara dalam regresi kuantil diasumsikan bahwa $q_\tau(\boldsymbol{\varepsilon}|\mathbf{X}) = \boldsymbol{a}^{\boldsymbol{\tau}}$, dimana *error* dalam regresi kuantil tidak membutuhkan asumsi distribusi. Sehingga bentuk umum dari kondisional kuantil sebagai berikut :

$$q_\tau(Y|\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}^{\boldsymbol{\tau}} \quad (4)$$

Nilai $\boldsymbol{\beta}$ tersebut dapat diestimasi dengan meminimumkan fungsi objektif kuantil sebagai berikut:

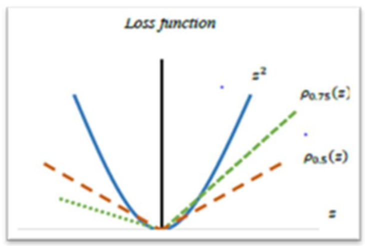
$$\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - \mathbf{X}_i'\boldsymbol{\beta}) \quad (5)$$

Fungsi $\rho_\tau(z)$ didefinisikan sebagai berikut:

$$\rho_\tau(z) = \begin{cases} \tau z & jika\ z > 0 \\ -(1-\tau)z & lainnya \end{cases} \quad (6)$$

Dengan *z* sebagai kekeliruan.

Untuk mengetahui perbedaan *loss function*, berikut ini menggambarkan kuantil *loss function* $\tau = 0.75$ dan $\tau = 0.25$



Gambar 1. Loss Function Quantile dan OLS

Regresi kuantil bersifat *non-differensiable* sehingga penaksir koefisien regresi kuantil tidak dapat ditunjukkan secara eksplisit. Maka, untuk menaksir koefisien regresi kuantil tersebut dapat melalui proses mengubah fungsi objektif kuantil kedalam penyelesaian masalah program linier untuk menemukan solusi yang optimal. (Koenker dan Basset, 1978)

LP-Problem merupakan suatu model umum yang digunakan dalam memecahkan masalah dengan meminimumkan maupun memaksimumkan suatu fungsi secara optimum dengan dibatasi oleh fungsi kendala linier yang dapat berupa persamaan atau pertidaksamaan (Andriyana, 2015).

Regresi Kuantil Tidak Bersyarat

Salah satu cara untuk mengetahui distribusi variabel acak adalah melalui fungsi distribusi kumulatif. Asumsikan bahwa *Y* adalah variabel acak, maka peluang bahwa *Y* lebih kecil atau sama dengan suatu nilai tertentu dilambangkan dengan *c*, sehingga fungsi distribusi kumulatif $F_Y(c)$ dari *c* dapat ditulis sebagai berikut:

$$F_Y(c) = P(Y \le c) \quad (7)$$

Kuantil ke-τ dari *Y* didasarkan pada *objective function L_1* atau *loss function* akan menunjukkan lokasi-lokasi khusus dari suatu distribusi. Kuantil ke-τ dari variabel acak *Y* didefinisikan sebagai berikut:

$$q_\tau(Y) = F_Y^{-1}(\tau) = inf\{c : F_Y(c) \ge \tau\} \quad (8)$$

Fungsi $F_Y(.)$ dapat diganti dengan $F_n(Y)$ dan ${F_y}^{-1}(\tau)$ yang merupakan penaksir dari ${F_y}^{-1}(\tau)$ dapat diperoleh dengan meminimumkan:

$$q_\tau(Y) = argmin_c E[\rho_\tau(Y - c)] \quad (9)$$

dimana *c* merupakan nilai dari kuantil ke-τ..

Kemudian, regresi kuantil ke τ dan F_Y dapat diperoleh di Persamaan berikut:

$$E[\rho_\tau(Y - c)] = (\tau - 1) \int_{-\infty}^{c} (Y - c)dF_Y(y) + \tau \int_{c}^{\infty} (Y - c)dF_Y(y) \quad (10)$$

Persamaan tersebut diminimumkan dan disamakan dengan nol sehingga diperoleh:

$$(1 - \tau) \int_{-\infty}^{c} dF_Y(y) - \tau \int_{c}^{\infty} dF_Y(y) = F_Y(c) - \tau = 0 \quad (11)$$

Fungsi $F_Y(.)$ merupakan fungsi monoton naik, sehingga setiap elemen dari $\{y: F_Y(y) = \tau\}$ meminimumkan fungsi (2.18). Dari persamaan (2.19) diperoleh bahwa $c = F_y{}^{-1}(\tau)$ yang merupakan solusi yang unik.

Regresi Kuantil Bersyarat

Misalkan Y adalah variabel respon yang bergantung pada $(X^{(1)}, \dots, X^{(p)})$ yang merupakan himpunan dari variabel-variabel prediktor, maka model regresi pada persamaan (2.1) dapat dibentuk menjadi sebagai berikut:

Dengan $\boldsymbol{Y} = (Y_1, \dots, Y_n)^T$, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^T$ dan $\boldsymbol{X} = (1, X^{(1)}, \dots, X^{(p)})^{\boldsymbol{T}}$, $\boldsymbol{X}^{(k)} = (X_1^{(k)}, \dots, X_n^{(k)})^T$ untuk k=0,1, ..., p dan $X^{(0)}$ adalah unit vektor dengan panjang n. Kemudian **ε** diasumsikan mempunyai suatu distribusi dengan notasi F. Kuantil ke –τ dari error **(ε)** adalah :

$$F^{-1}(\tau) = inf\{u: P\{\boldsymbol{\varepsilon} \leq u\} \geq \tau\} \quad (12)$$

dengan u adalah titik kuantil dari *error* model regresi pada persamaan (2.1).

Persamaan kurva kuantil untuk Y bersyarat $\boldsymbol{X}$ dapat ditulis sebagai berikut:

$$q_\tau(Y|\boldsymbol{X}) = [\beta_0 + F_\varepsilon^{-1}(\tau)] = \beta_1 X^{(1)} + \cdots + \beta_p X^{(p)} = \boldsymbol{X^T \beta^\tau} \quad (13)$$

dimana $\boldsymbol{\beta(\tau)} = ((\beta_0 + F^{-1}(\tau)), \dots, \beta_p)^T$. Penaksir parameter $\boldsymbol{\beta(\tau)}$ akan diperoleh dengan cara meminimumkan persamaan empiris dari *objective function* pada Persamaan (2.13) berikut :

$$\min_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau\left(y_i - \boldsymbol{X_i^T \beta}\right) \quad (14)$$

dimana $\boldsymbol{X_i} = (1, X_i{}^{(1)}, \dots, X_i{}^{(p)})^T$ merupakan observasi ke – i dari **X.**

Selang Kepercayaan pada Regresi Kuantil

Selang kepercayaan yaitu suatu interval antara dua angka, dimana nilai parameter dari populasi terletak di dalam interval tersebut. Metode ini memiliki kelebihan yang dapat menjelaskan penyebaran data dengan menggunakan ukuran kuantil sesuai sebaran data, menghasilkan hasil yang efisien dalam menaksir selang kepercayaan dibandingkan metode lainnya (Zhou, 1996).

Dalam regresi kuantil terdapat tiga metode yang digunakan untuk menghitung selang kepercayaan $\hat{\beta}$ salah satunya adalah direct fungsi sparsity. Setiap τ yang telah ditetapkan, taksiran interval dengan formula sebagai berikut (Davino, Furno, dan Vistocco, 2014: 132-134):

$$P\left(\hat{\beta}_j(\tau) - t_{\left(\frac{\alpha}{2},\mathrm{df}\right)} se\left(\hat{\beta}_j(\tau)\right) \leq \beta_j(\tau) \leq \hat{\beta}_j(\tau) + t_{\left(\frac{\alpha}{2},\mathrm{df}\right)} se\left(\hat{\beta}_j(\tau)\right)\right) = 1 - \alpha \quad (2.15)$$

Dari perhitungan matriks $\widehat{\omega}^2(\tau)\boldsymbol{D}^{-1}$ diperoleh nilai $se\left(\hat{\beta}_j(\tau)\right)$ nilai diagonal, dimana $\boldsymbol{D^{-1}} = (\boldsymbol{X^T X})^{\boldsymbol{-1}}$ dan $\widehat{\omega}^2(\tau) = \tau(1 - \tau)/\mathrm{f}\,(\mathrm{F}^{-1}(\tau)\,)^2$. Untuk F adalah fungsi distribusi kumulatif dan $f = F'$ adalah fungsi densitas.

Pengujian Hipotesis

Parameter pada kuantil ke-τ dimungkinkan sangat banyak pada τ ϵ (0,1), sehingga uji signifikansi pada parameter pengujian digunakan untuk mengetahui pengaruh antara variabel dependen dengan variabel independen. Dengan model regresi kuantil dari persanmaan (2.1) dengan hipotesis sebagai berikut:

Hipotesis :

$H_0 : \beta_j(\tau) = 0$ (pengaruh peubah ke-*j* pada level kuantil ke-τ tidak signifikan)

$H_1 : \beta_j(\tau) \neq 0$ (pengaruh peubah ke-*j* pada level kuantil ke-τ signifikan)

dengan *j* = 1,2,...,*p*

Statistik uji yang digunakan yaitu uji Wald :

$$W = \frac{\hat{\beta}_j(\tau)}{se\left(\hat{\beta}_j(\tau)\right)} \quad (16)$$

H_0 ditolak apabila $W > Z_{(\propto/2)}$. *Standar error* pada uji ini menerangkan salah satu komponen untuk membentuk konfidensi interval. Apabila terdapat penolakan terhadap H_0 berarti parameter yang di uji signifikan atau memberikan pengaruh yang nyata terhadap variabel *Y*.

Data
Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari laporan Puskesmas kepada Dinas Kesehatan dan Pengendalian Penduduk di Kabupaten Tasikmalaya tahun 2019. Wilayah yang dijadikan sebagai objek penelitian adalah 39 kecamatan di Kabupaten Tasikmalaya. Variabel yang digunakan dalam penelitian ini ialah satu variabel respon, tiga variabel prediktor .Variabel respon yang digunakan ialah IR Tuberkulosis Paru di Kabupaten Tasikmalaya tahun 2019, Variabel prediktor yang digunakan ialah Faktor persentase fasilitas sanitasi yang layak/ jamban sehat (X_1), Faktor Persentase ketersediaan air bersih dan Faktor persentase kebiasaan tidak merokok (X_3).

Metode
Untuk mendapatkan model regresi kuantil yang sesuai serta memperoleh selang kepercayaan untuk kurva regresi kuantil dalam penelitian ini maka tahapan-tahapan regresi kuantil dilakukan sebagai berikut:

1. Menentukan variabel respon dan prediktor yang akan digunakan, dalam penelitian ini menggunakan jumlah persentasi penyakit tuberkulosis sebagai variabel respon, dengan faktor yang mempengaruhinya yaitu faktor sebagai variabel prediktor.
2. Menentukan prespesifikasi model berdasarkan pola data yang diperoleh pada scatter plot antara variabel respon dengan masing-masing variabel independennya. Dari scatter plot dapat dijelaskan hubungan antara variabel respon dengan variabel prediktornya.
3. Mendeteksi pencilan dari hasil scatterplot. Langkah ini untuk melihat perilaku error atau distribusi error dengan rata-rata sebagai ukuran pemusatan data pada regresi kuantil. Untuk penelitian ini model yang dipakai menggunakan analisis regresi kuantil karena adanya pencilan (outlier) tersebut, maka penggunaan regresi kuantil akan bersifat robust terhadap pencilan.
4. Menaksir nilai estimasi parameter model yang optimal pada beberapa titik kuantil. Pada penelitian ini ingin mendapat pola atau pattern dari persentasi penyakit tuberkulosis terhadap faktor-faktor yang mempengaruhinya ke dalam tiga kelompok (rendah, sedang, dan tinggi) maka pada penelitian ini digunakan 2 level kuantil yaitu level kuantil ke-0.33 dan level kuantil ke-0.67 ($\tau = 0.33$; $\tau = 0.67$).
5. Menaksir selang kepercayaan untuk kurva regresi kuantil dimana akan menghasilkan penaksir parameter yang lebih representatif dibandingkan sistem penaksir titik pada langkah ke-5 dengan menggunakan metode langsung (direct). Package R yang digunakan dalam pemodelan adalah package quantreg.

C. Hasil Penelitian dan Pembahasan

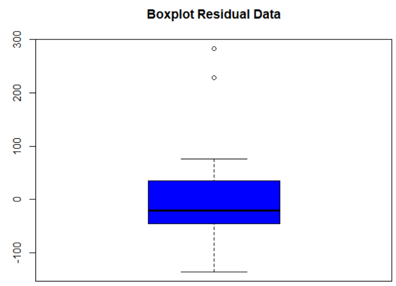
Deskripsi Data
Data yang digunakan adalah data IR TB Paru di 39 kecamatan yang ada di Kabupaten Tasikmalaya pada tahun 2019 beserta faktor-faktor yang diduga mempengaruhinya yang diperoleh dari Dinas Kesehatan dan Pengendalian Penduduk Kabupaten Tasikmalaya. Berikut merupakan hasil analisis deksriptif variabel penelitian ini.

Tabel 1. Analisis Deskriptif Variabel Penelitian

Variabel	Minimum	Median	Rata-rata	Maksimum
IR Tuberkulosis Paru (Y)	42,37	151,52	169,40	496,21
Jamban Sehat (X_1)	20,09	71,10	69,50	100,00
Air Bersih (X_2)	43,21	90,19	86,01	100,00
Kebiasaan Tidak Merokok (X_3)	46,25	54,33	57,37	94,65

Spesifikasi Model

Sebelum dilakukan pemodelan mengenai hubungan antara variabel dependen dengan variabel independent, terlebih dahulu akan dibuat plot kekeliruan dari permodelan dalam regresi *mean* untuk melihat perspesifikasi model terhadap sebaran data yaitu sebagai berikut:



Gambar 2. Kekeliruan dari Pemodelan dengan Menggunakan Regresi Mean

Berdasarkan **Gambar 2** di atas, dapat dijelaskan bahwa sebaran kekeliruan dari pemodelan regresi *mean* pada Boxplot tersebut terdapat adanya *outlier* atau data pencilan. Adanya *outlier* akan mengganggu proses analisis data khususnya dalam estimasi parameter. Oleh karena itu, perlu adanya metode regresi yang *robust* terhadap *outlier*, salah satunya yaitu dengan menggunakan regresi kuantil untuk mendapatkan penaksir parameter yang lebih efisien dari model. Model regresi kuantil yang akan digunakan adalah kuantil bersyarat karena melibatkan informasi variabel independen dalam modelnya.

Estimasi Parameter Model Regresi Kuantil

Regresi kuantil dapat digunakan dalam berbagai level kuantil sesuai dengan kebutuhan penelitian. Pada penelitian ini, akan dibuat pola atau *pattern* dari persentase penyakit tuberkulosis dengan faktor-faktor yang mempengaruhi kedalam tiga kelompok yaitu rendah, sedang dan tinggi maka akan digunakan dua level kuantil yaitu kuantil ke- 0.33 dan kuantil ke-0.67. Berikut merupakan hasil estimasi parameter pada kuantil ke-0.33 dan ke 0.67 dengan menggunakan *software* R Studio yang ditujukan pada tabel sebagai berikut:

Tabel 2. Estimasi Parameter Regresi Kuantil pada Level Kuantil 33% dan Level Kuantil 67%

Variabel	Kuantil (τ)			
	33%		67%	
	Nilai	*p-value*	Nilai	*p-value*
Intercept	30,097484	0,76015	97,28015	0,6288
X1	-0,14363	0,8323	0,42972	0,74804
X2	2,18344	0,00384	1,09175	0,43555
X3	-1,3609	0,36784	-0,62229	0,84585

Hipotesis :

$H_0 : \beta_j(\tau) = 0$ (pengaruh peubah ke-*j* pada level kuantil ke-τ tidak signifikan)

$H_1 : \beta_j(\tau) \neq 0$ (pengaruh peubah ke-*j* pada level kuantil ke-τ signifikan)

dengan $j = 1,2,\ldots,p$

Statistik uji :

Signifikansi jika p-value < 0.05

Kesimpulan :

Dari hasil Tabel 2 dapat dilihat bahwa pada level kuantil ke-0.33, terdapat variabel yang signifikan yaitu persentase ketersediaan air bersih dengan *p-value* sebesar 0,00384 sedangkan pada kuantil ke-0,67 tidak ada variabel yang signifikan.

Adapun model regresi kuantil yang terbentuk yaitu sebagai berikut:

$$\boldsymbol{q_{0,33}}(\mathbf{Y}|\mathbf{X}) = 30{,}097484 - 0{,}14363\, X_1 + 2{,}18344\, X_2 - 1{,}3609\, X_3$$

$$\boldsymbol{q_{0,67}}(\mathbf{Y}|\mathbf{X}) = 97{,}28015 + 0{,}42972\, X_1 + 1{,}09175\, X_2 - 0{,}62229\, X_3$$

Setelah diperoleh model regresi kuantil tersebut, selanjutnya dapat dilakukan estimasi nilai *incident rate* TB Paru berdasarkan faktor-faktor yang mempengaruhinya.

Estimasi Koefisien Model Regresi Kuantil

Dalam regresi kuantil bersyarat akan melibatkan variabel respon yang dipengaruhi oleh variabel prediktor dengan menggunakan beberapa level kuantil yang nantinya akan menghasilkan garis regresi sebanyak yang dibutuhkan. Pada penelitian ini ingin membuat pola atau *pattern* dari incident rate TB Paru terhadap faktor-faktor yang mempengaruhinya ke dalam tiga kelompok, yaitu rendah, sedang dan tinggi. Oleh karena itu digunakan dua level kuantil yaitu level 0,33 dan level 0,67. Berikut merupakan nilai estimasi dari incident rate TB Paru terhadap masing-masing faktor yang mempengaruhinya untuk setiap level kuantilnya yang ditujukan pada Tabel 4.3 berikut:

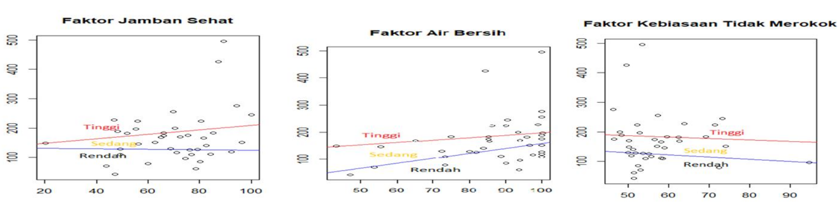
Tabel 3. Nilai Estimasi Kuantil dari Incident Rate TB Paru Bersyarat Faktor-Faktor pada Level Kuantil Ke-0.33 dan Level Kuantil Ke-0.67

No	Jamban Sehat		Air Bersih		Kebiasaan Tidak Merokok	
	Kuantil (τ)					
	33%	67%	33%	67%	33%	67%
1	126,34	193,59	139,88	188,35	128,09002	186,4338
2	126,09	196,53	137,12	186,97	123,65636	183,4769
3	125,90	198,78	129,18	183,00	130,19103	187,835
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
37	126,33	193,74	139,78	188,30	113,82737	176,9216
38	127,91	175,29	77,00	156,96	123,06131	183,08
39	127,26	182,86	112,25	174,56	122,72806	182,8577

Dari hasil estimasi *Incident Rate* TB Paru pada Tabel 3.4 tersebut dapat dibentuk kurva regresi kuantil masing-masing faktor yang mempengaruhinya yaitu Persentase Jamban Sehat, Persentase Air Bersih dan Persentase Kebiasaan Tidak Merokok.

Pengklasifikasian *Incident Rate* TB Paru beserta faktor-faktor yang diduga mempengaruhinya merupakan hal yang harus dilakukan. Tujuannya untuk mengetahui pada angka berapa TB Paru dengan faktor-faktor yang diduga mempengaruhinya tersebut berada pada tingkatan yang tinggi agar memperkuat pengambilan kebijakan pemerintah dalam mendeteksi dini penyakit TB Paru selanjutnya.

Adapun pola atau *pattern* yang terbentuk dari hasil karakteristik penyakit TB Paru terhadap faktor-faktor yang diduga mempengaruhinya dapat dilihat pada Gambar 4.2 berikut ini:



Gambar 3. Kurva Estimasi Kuantil untuk Penyakit TB Paru dengan Faktor-Faktor yang diduga mempengaruhinya

Fungsi dari kurva di atas yaitu untuk mengidentifikasi posisi *incident rate* TB Paru apakah angka tersebut berada pada tingkatan rendah, sedang atau tinggi dengan melihat pola dari pergerakan titik-titiknya.

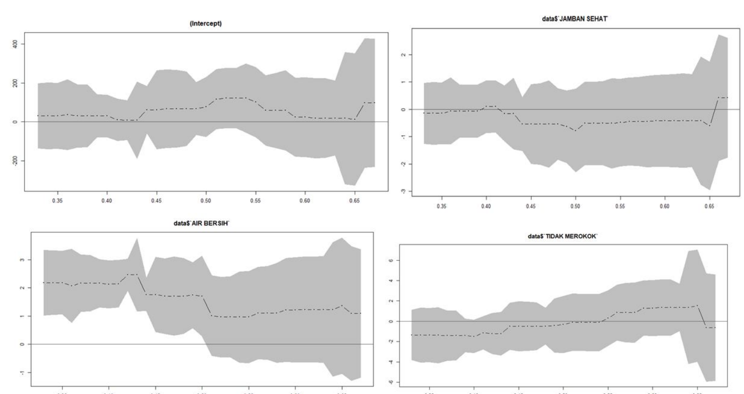
Taksiran Interval Regresi Kuantil pada Incident Rate TB Paru dan Faktor-faktor yang diduga mempengaruhinya

Berdasarkan penjelasan sebelumnya yang menyatakan bahwa sistem penaksir parameter populasi berdasarkan sampel adalah selang kepercayaan yang akan menghasilkan penaksir parameter yang lebih representatif dibandingkan sistem penaksir titik, dimana nilai dari estimasi penaksir populasinya terletak diantara dua angka dalam interval tersebut. Dengan bantuan *software* R Studio, selang kepercayaan dengan melibatkan 3 variabel independen pada setiap τ yang ditentukan adalah sebagai berikut.

Tabel 4. Taksiran Interval untuk Setiap Faktor yang Mempengaruhi Incident Rate TB Paru pada Level Kuantil ke-0.33 dan ke-0.67

Faktor-Faktor	**τ=0.33**			**τ=0.67**		
	Batas Bawah	Koefisien	Batas Atas	Batas Bawah	Koefisien	Batas Atas
Intercept	-84,22803	30,97483	213,6591	-220,69389	97,28015	217,82905
Jamban Sehat	-2,12185	-0,1436	0,66861	-0,91936	0,42972	1,80913
Air Bersih	-0,84308	2,18344	2,77306	0,48726	1,09175	3,349
Kebiasaan Tidak Merokok	-1,72003	-1,36089	3,27419	-3,03572	-0,62229	3,16021

Dari Tabel 4 di atas, dapat dilihat bahwa nilai koefisien berada diantara dua nilai interval dalam setiap faktor yang diduga mempengaruhinya. Hal ini menyatakan bahwa penaksir parameter tersebut sudah cukup merepresentatifkan populasinya. Adapun gambar taksiran interval untuk setiap faktor yang diduga mempengaruhi *incident rate* TB Paru pada level kuantil ke-0.33 dan ke-0.67 yaitu sebagai berikut:



Gambar 4. Taksiran Interval untuk Setiap Faktor yang mempengaruhi Incident Rate TB Paru pada Setiap Level Kuantil

Gambar 4 menunjukkan nilai-nilai selang kepercayaan 95% dari incident rate TB Paru berdasarkan faktor-faktor yang mempengaruhinya. Pada gambar tersebut terdapat garis horizontal di titik nol untuk menunjukkan jarak antar estimator terhadap nilai titik nol.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Pada penelitian ini data yang dimiliki terdapat outlier pada kekeliruan dengan pemodelan menggunakan regresi mean. Oleh karena itu perlu adanya metode yang robust terhadap outlier, sehingga pada penelitian ini menggunakan regresi kuantil.
2. Pada regresi kuantil akan dibentuk sebuah pattern dari variabel respon terhadap variabel prediktor dengan menggunakan dua level kuantil yaitu pada kuantil kuantil $\tau = 0.33$ dan $\tau = 0.67$, sehingga dari model yang dibangun dapat dikelompokkan menjadi 3 tingkat keparahan dari incident rate TB Paru, yaitu rendah, sedang dan tinggi sehingga model dapat digunakan untuk mengestimasi incident rate TB Paru menurut faktor-faktor yang diduga mempengaruhinya.
3. Adapun model regresi yang dihasilkan yaitu:
$$q_{0,33}(\mathrm{Y}|\mathrm{X}) = 30{,}097484 - 0{,}14363\, X_1 + 2{,}18344\, X_2 - 1{,}3609\, X_3$$
$$q_{0,67}(\mathrm{Y}|\mathrm{X}) = 97{,}28015 + 0{,}42972\, X_1 + 1{,}09175\, X_2 - 0{,}62229\, X_3$$
4. Pada level kuantil ke-0.33, hanya variabel air bersih (X2) yang mempengaruhi incident rate TB paru, sedangkan pada level kuantil ke 0.67, tidak ada variabel yang signifikan.

Acknowledge

Penulis mengucapkan banyak terima kasih kepada pihak-pihak yang terlibat dalam membantu, bimbingan, motivasi dan do'a hingga penelitian ini dapat terlaksana hingga selesai.

Daftar Pustaka

[1] Andriyana Y. (2015). P-Spline *Quantile Regression* in Varying Coefficient Models. Disertasi. Belgium : KU Leuven.

[2] Basset, K. d. (1978). Regression Quantile Econometrica. 46.33-50.

[3] Davino, Furno dan Vistocco (2014). *Quantile Regression* (Theory and application). Stanford Weisberg.

[4] Dinkes Kabupaten Tasikmalaya. (2020). Profil Kesehatan Kab.Tasikmalaya: Dinkes dan Pengendalian Penduduk Kabupaten Tasikmalaya.

[5] Dinkes Provinsi Jawa Barat (2017). Profil Kesehatan Jawa Barat. Dinkes Provinsi Jawa Barat.

[6] Gujarati, D. (2004). Basic Econometrics (Ekonometrika Dasar). Terjemahan oleh Sumarno Zain. Jakarta: Erlangga

[7] Karmakar, N. (1984). A New Polynomial Time Algorithm for Linear Programming. Combiatoria.

[8] Rousseeuw, P. J., Debryune, M., Engelen, S., dan Hubert, M. 2006, Robustness and outlier Detection in Chemometrics.Taylor & Francis Group, LLC

[9] Johnson R. A., Wichern D. W., (2002). Applied Multivariate Statistical Analysis. Fifth Edition, Prentice Hall, New Jersey.

[10] Kenneth Q. Zhou. Stephen L. Portnoy. "Direct use of regression quantiles to construct confidence sets in linear models." Ann. Statist. 24 (1) 287 - 306, February 1996. https://doi.org/10.1214/aos/1033066210

[11] Budiharti, Luhung Mustika. (2021). *Pemodelan dan Pemetaan Jumlah Penderita Kusta di Jawa Barat dengan Regresi Binomial Negatif dan Flexibly Shaped Spatial Scan Statistic*, Jurnal Riset Statistika, 1(2), 99-106.