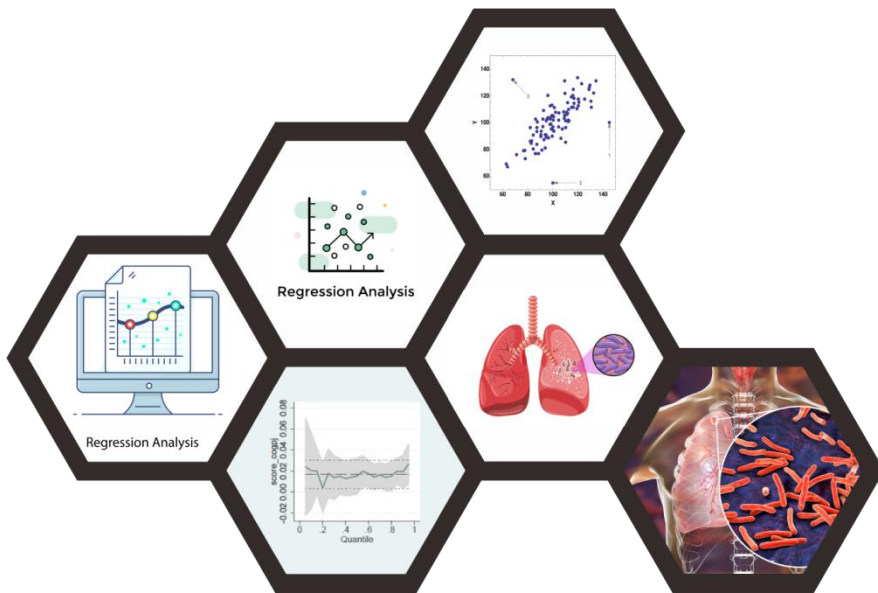


**REGRESI KUANTIL DENGAN PENALTI GROUP LASSO UNTUK
MEMODELKAN DATA JUMLAH KASUS TUBERKULOSIS DI PROVINSI
SULAWESI SELATAN**

**QUANTIL REGRESSION WITH GROUP LASSO PENALTY TO MODEL DATA
ON THE NUMBER OF TUBERCULOSIS CASES IN SOUTH SULAWESI
PROVINCE**



**IRWAN USMAN
H062221010**



**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

**REGRESI KUANTIL DENGAN PENALTI GROUP LASSO UNTUK
MEMODELKAN DATA JUMLAH KASUS TUBERKULOSIS DI PROVINSI
SULAWESI SELATAN**

**IRWAN USMAN
H062221010**



**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

REGRESI KUANTIL DENGAN PENALTI GROUP LASSO UNTUK
MEMODELKAN DATA JUMLAH KASUS TUBERKULOSIS DI PROVINSI
SULAWESI SELATAN

Tesis
sebagai salah satu syarat untuk mencapai gelar magister

Program Studi Magister Statistika

Disusun dan diajukan oleh

IRWAN USMAN
H062221010

Kepada

**PROGRAM STUDI MAGISTER STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2024**

TESIS

REGRESI KUANTIL DENGAN PENALTI GROUP LASSO UNTUK MEMODELKAN DATA JUMLAH KASUS TUBERKULOSIS DI PROVINSI SULAWESI SELATAN

IRWAN USMAN
H062221010

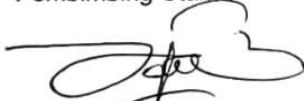
telah dipertahankan di depan Panitia Ujian Magister pada 16 Agustus
2024
dan dinyatakan telah memenuhi syarat kelulusan

pada

Program Studi Magister Statistika
Departemen Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Hasanuddin
Makassar

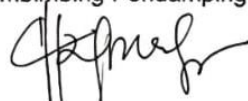
Mengesahkan:

Pembimbing Utama



Dr. Anna Islamiyati, S.Si., M.Si
NIP. 197708082005012002

Pembimbing Pendamping



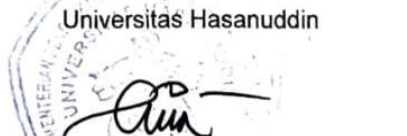
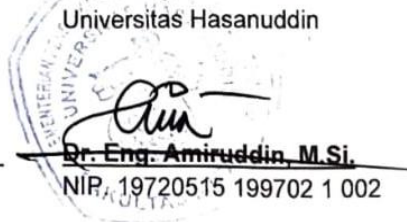
Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Ketua Program Studi
Magister Statistika




Dr. Erna Tri Herdiani, S.Si., M.Si.
NIP. 19750429 200003 2 001

Dekan Fakultas Matematika dan Magister
Ilmu Pengetahuan Alam
Universitas Hasanuddin

Dr. Eng. Amiruddin, M.Si.
NIP. 19720515 199702 1 002

PERNYATAAN KEASLIAN TESIS DAN PELIMPAHAN HAK CIPTA

Dengan ini saya menyatakan bahwa, tesis dengan judul "Regresi Kuantil dengan Penalti Group LASSO untuk Memodelkan Data Jumlah Kasus Tuberkulosis Di Provinsi Sulawesi Selatan" adalah benar karya saya dengan bantuan dan arahan dari pembimbing (Dr. Anna Islamiyati, S.Si.,M.Si. dan Dr. Erna Tri Herdiani, S.Si., M.Si.). Karya Ilmiah ini belum diajukan dan tidak sedang diajukan dalam bentuk apapun kepada perguruan tinggi manapun. Sumber informasi yang berasal atau dikutip dari karya yang diterbitkan maupun tidak diterbitkan dari penulis lain telah disebutkan dalam teks dan dicantumkan dalam Daftar Pustaka tesis ini. Sebagian dari tesis ini akan dipublikasikan di SCIK Publishing Corporation sebagai artikel dengan judul " *Quantile Regression Modeling with Least Absolute Shrinkage and Selection Operator Group Classification on Tuberculosis Data* ".

Dengan ini saya melimpahkan hak cipta dari karya tulis saya berupa tesis ini kepada Universitas Hasanuddin.

Makassar, 16 Agustus 2024



IRWAN USMAN
NIM. H062221010

UCAPAN TERIMA KASIH

Segala puji hanya milik Allah *Subhanallahu Wa Ta'ala* atas limpahan rahmat dan hidayah-Nya kepada penulis. Shalawat dan salam tercurahkan kepada Rasulullah *Shallallahu 'Alaihi Wa sallam*, keluarganya, *tabi'in, tabi'ut tabi'in*, serta orang-orang sholeh yang haq hingga kadar Allah berlaku atas diri mereka. *Alhamdulillahirobbil'aalamiin*, berkat rahmat dan kemudahan dari Allah *Subhanallahu Wa Ta'ala*, penulis dapat menyelesaikan tesis berjudul Regresi Kuantil dengan Penalti Group LASSO untuk Memodelkan Data Jumlah Kasus Tuberkulosis Di Provinsi Sulawesi Selatan " sebagai salah satu syarat memperoleh gelar magister pada Program Studi Magister Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Terima Kasih yang tak terhingga kepada kedua orang tuaku A'ba (**Usman A**) dan Mama (**Ramlah**) yang tiada hentinya memberikan perhatian, serta doa yang tak pernah putus, juga kepada saudara-saudaraku, (**Abian Usman dan Irsyad Usman**) yang selalu memberikan dukungan, semangat dan dalam penyelesaian tesis ini . Ucapan rasa hormat dan juga terima kasih yang tulus kepada:

1. Yth. **Prof. Dr. Ir. Jamaluddin Jompa, M.Sc** selaku Rektor Universitas Hasanuddin
2. Yth. **Dr. Eng. Amiruddin, M.Si** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam beserta seluruhnya jajarannya.
3. Yth. **Dr. Anna Islamiyati, S.Si., M.Si** selaku Ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam dan sekaligus pembimbing utama yang senantiasa meluangkan waktu, tenaga, pemikiran dalam membimbing dan mengarahkan penulis dalam menyelesaikan tesis.
4. Yth. **Dr. Erna Tri Hardiani, S.Si., M.Si** selaku Ketua Program Studi Magister Statistika dan sekaligus sebagai pembimbing pendamping yang senantiasa meluangkan waktu, tenaga, pemikiran dalam membimbing dan mengarahkan penulis dalam menyelesaikan tesis.
5. Yth. **Prof. Dr. Dr Georgina Maria Tinungki, M.Si** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis.
6. Yth. **Prof.Dr. Nurtiti Sunusi, S.Si., M.Si** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis
7. Yth. **Dr. Nirwan, M.Si** selaku penguji yang telah bersedia menguji serta memberikan masukan-masukan dan arahan dalam penyusunan tesis

8. Sahabat-sahabat terbaik penulis yang bernama **Zainal Abidin, Zulkifli, Misbahuddin dan A. Riska Atika** yang telah dengan tulus membantu dan mendukung saya untuk terus berjuang menyelesaikan tesis ini.
9. Teman-teman seperjuangan **Mahrani, Amaliah Rafiq, Samsul, Arya Winanda Tahir** dan **Taufiqurrahman**.
10. Teman-teman seperjuangan ruang diskusi, **Muhammad Fadil, Dwi Auliyah, Haura** dan teman-teman yang lain yang tidak bisa disebutkan satu-persatu yang selalu memberikan semangat dan bantuan.
11. Semua pihak yang telah membantu penulis yang tidak bisa disebutkan satu per satu terimakasih atas doa serta dukungannya

Semoga Tuhan Yang Maha Esa memberikan balasan yang berlipat ganda, kasih dan hikmat-Nya atas segala kebaikan yang telah diberikan kepada penulis. Penulis menyadari bahwa masih banyak kekurangan dalam tesis ini, untuk itu dengan segala kerendahan hati, penulis memohon maaf.

Makassar, 16 Agustus 2024

IRWAN USMAN

ABSTRAK

IRWAN USMAN. **Regresi Kuantil dengan Penalti Group LASSO untuk Memodelkan Data Jumlah Kasus Tuberkulosis Di Provinsi Sulawesi Selatan** (dibimbing oleh Dr. Anna Islamiyati, S.Si.,M.Si. dan Dr. Erna Tri Herdiani, S.Si., M.Si.)

Latar Belakang. Tuberkulosis (TB) merupakan salah satu dari 10 penyebab kematian terbesar di dunia dan merupakan penyakit menular mematikan di Indonesia. Salah satu provinsi di Indonesia yang menyumbang kasus TBC terbanyak pada tahun 2018 adalah Sulawesi Selatan, dengan 84 kasus per 100.000 penduduk. **Tujuan.** Penelitian ini bertujuan untuk mengidentifikasi variabel-variabel yang dapat menjelaskan proporsi kasus TBC di Sulawesi Selatan. **Metode.** Data yang digunakan mempunyai variabel prediktor yang banyak dan terdapat outlier. Regresi kuantil dapat digunakan untuk mengatasi data outlier, namun tidak dapat mengatasi permasalahan multikolinearitas. Multikolinearitas menyebabkan varians parameter estimasi menjadi terlalu besar dan mengurangi keakuratan estimasi, sehingga memerlukan pendekatan analisis data yang berbeda. Terdapat berbagai macam metode untuk menangani analisis regresi pada data yang mengalami permasalahan multikolinearitas. Salah satu metode regresi berpenalti yang paling umum dikenal adalah Grup LASSO. Group LASSO dapat digunakan untuk menyeleksi variabel dan mengatasi multikolinearitas. **Hasil.** Dalam penelitian ini, terdapat enam variabel kelompok sektor yang terbentuk secara alami yang diduga mempengaruhi proporsi kasus TBC. Pemodelan regresi kuantil dengan penalti kelompok LASSO dilakukan dengan menggunakan 3 level kuantil yaitu (0,25, 0,5, dan 0,75). Hasil analisis regresi kuantil dengan penalti group LASSO diperoleh model yang berbeda untuk setiap kuantil. **Kesimpulan.** Model terbaik yang mampu menjelaskan proporsi kasus TBC diperoleh pada level kuantil 0,5 dengan nilai R^2 sebesar 0,99 mendekati 1 dibandingkan level model kuantil lainnya.

Kata Kunci : group LASSO; multikolinearitas; outliers; regresi kuantil; tuberkulosis.

ABSTRACT

IRWAN USMAN. **Regression With Group Lasso Penalty To Model Data On The Number Of Tuberculosis Cases In South Sulawesi Province** (guided by Dr. Anna Islamiyati, S.Si., M.Si. and Dr. Erna Tri Herdiani, S.Si., M.Si.)

Background. Tuberculosis (TB) is one of the 10 biggest causes of death in the world and is a deadly infectious disease in Indonesia. One of the provinces in Indonesia that contributed the most TB cases in 2018 was South Sulawesi, with 84 cases per 100,000 population. **Aim.** This study aims to identify variables that can explain the proportion of TB cases in South Sulawesi. **Method.** The data used have many predictor variables and outliers. Quantile regression can be used to overcome outlier data, but cannot overcome multicollinearity problems. Multicollinearity causes the variance of the estimated parameters to be too large and reduces the accuracy of the estimate, thus requiring a different data analysis approach. There are various methods for handling regression analysis on data that experiences multicollinearity problems. One of the most commonly known penalized regression methods is Group LASSO. Group LASSO can be used to select variables and overcome multicollinearity. **Results.** In this study, there are six naturally formed sector group variables that are suspected of influencing the proportion of TB cases. Quantile regression modeling with LASSO group penalty was carried out using 3 quantile levels, namely (0.25, 0.5, and 0.75). The results of the quantile regression analysis with LASSO group penalty obtained different models for each quantile. **Conclusion.** The best model that is able to explain the proportion of TB cases is obtained at the 0.5 quantile level with an R^2 value of 0.99 approaching 1 compared to other quantile model levels.

Keywords: group LASSO, multicollinearity, outliers, quantile regression, tuberculosis

DAFTAR ISI

HALAMAN JUDUL	ii
PERNYATAAN PENGAJUAN	iii
HALAMAN PENGESAHAN	Error! Bookmark not defined.
PERNYATAAN KEASLIAN TESIS	v
UCAPAN TERIMA KASIH.....	v
ABSTRAK	viii
ABSTRACT	ix
DAFTAR ISI.....	x
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
BAB I.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Kajian Teori.....	3
1.6.1 Analisis Regresi.....	3
1.6.3 Estimasi Parameter Regresi Kuantil	4
1.6.4 Algoritma Simpleks Regresi Kuantil.....	5
1.6.5 Data Outlier	6
1.6.6 Multikolinieritas	6
1.6.7 Regresi Kuantil dengan LASSO	7
1.6.8 Group LASSO.....	8
1.6.9 Tuberculosis (TBC).....	9
1.7 Kerangka Konsep	10
BAB II.....	11
2.1 Sumber Data	11
2.2 Variabel Penelitian.....	11
2.3 Metode Analisis	14
BAB III.....	15

3.1	Estimasi Parameter model regresi kuantil group LASSO.....	15
3.2	Analisis Deskriptif.....	20
3.2.1	Deskripsi Jumlah Kasus Tuberkulosis (TBC)	21
3.2.2	Deskripsi Faktor-faktor yang Mempengaruhi Jumlah Kasus Tuberkulosis.....	22
3.3	Deteksi Outlier.....	25
3.4	Uji Multikolinearitas.....	25
3.5	Penerapan Regresi Kuantil Group LASSO	27
BAB IV	30
4.1	Kesimpulan	30
4.2	Saran.....	31
DAFTAR PUSTAKA	32

DAFTAR TABEL

Tabel 1. Variabel Penelitian	11
Tabel 2. Jumlah Kasus TBC Per-Kabupaten/Kota di Sulawesi Selatan.....	21
Tabel 3. Deskripsi Data Kota/kabupaten Berdasarkan Kelompok	22
Tabel 4. Kerangka Konsep.....	26
Tabel 5. Nilai Lambda Optimal untuk setiap kuantil	27
Tabel 6. Estimasi parameter regresi kuantil dengan penalti group LASSO	27
Tabel 7. Nilai R^2 Pada Tiap Kuantil	29

DAFTAR GAMBAR

Gambar 1. Kerangka Konsep.....	10
Gambar 2. Box plot dari Variabel Penelitian	25
Gambar 3. Korelasi Antar Variabel	26

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu statistika yang dapat digunakan untuk mengetahui pola hubungan antara variabel prediktor dan variabel respon, sehingga menghasilkan persamaan matematis yang menggambarkan hubungan fungsional antara variabel-variabel tersebut (Dani et al., 2022). Pada umumnya, dalam analisis regresi, variabel respon bersifat kontinu. Salah satu jenis regresi yang cocok untuk menganalisis variabel respon yang kontinu adalah Regresi Kuantil. Metode ini diperkenalkan oleh Koenker dan Bassett pada tahun 1978. Regresi kuantil berguna dalam mengatasi keterbatasan regresi linear klasik, terutama ketika asumsi seperti galat yang tidak berdistribusi normal, sensitivitas terhadap data pencilan (outlier), dan varians galat yang tidak konstan tidak terpenuhi. Namun, metode ini masih memiliki kelemahan, yaitu tidak mampu menangani masalah multikolinearitas.

Multikolinearitas menyebabkan varians dari parameter yang diestimasi menjadi sangat besar dan mengurangi ketepatan estimasi (Vatcheva et al., 2016). Padahal, salah satu asumsi penting dalam regresi adalah tidak adanya multikolinearitas antara variabel-variabel prediktor yang digunakan dalam model. Oleh karena itu, data yang mengalami masalah ini memerlukan pendekatan analisis data yang berbeda. Ada berbagai metode untuk menangani regresi pada data yang menghadapi masalah multikolinearitas. Beberapa metode regresi dengan penalti yang umum digunakan adalah regresi LASSO (Least Absolute Shrinkage and Selection Operator) dan regresi ridge. Perbedaan utama antara LASSO dan ridge terletak pada penalti yang diberikan, di mana pada LASSO, penalti dihitung dengan mengalikan nilai absolut dari koefisien regresi, sedangkan pada ridge, penalti dihitung dengan mengalikan kuadrat dari koefisien regresi. Penelitian oleh (Yanke et al., 2022) mengenai penanganan multikolinearitas dalam pemodelan regresi pertumbuhan ekonomi Indonesia berdasarkan teori pertumbuhan ekonomi endogen menunjukkan bahwa metode LASSO adalah yang terbaik dalam mengatasi multikolinearitas berdasarkan nilai mean square error (MSE). Namun, pada data berdimensi tinggi, metode LASSO kurang efektif dalam menjelaskan hubungan antara variabel prediktor dan variabel respon (Chen et al., 2020), sehingga diperlukan pengembangan lebih lanjut dari LASSO, yaitu metode group LASSO.

Group LASSO adalah pengembangan dari metode LASSO, di mana kendala kelompok ditambahkan ke dalam metode ini. Group LASSO sangat efektif untuk menangani data berdimensi tinggi, terutama ketika variabel-variabel prediktor dikelompokkan dalam grup-grup tertentu (Yuan & Lin, 2006).

Tuberkulosis (TBC) adalah penyakit menular yang disebabkan oleh bakteri *Mycobacterium tuberculosis*, yang umumnya menyerang paru-paru, namun juga bisa menyerang organ tubuh lainnya (Rosandali, 2016). Berdasarkan data Kemenkes RI, jumlah kasus TBC di Sulawesi Selatan mengalami penurunan pada tahun 2020. Meski demikian, penurunan ini harus tetap diwaspadai mengingat adanya kesenjangan dalam penyebaran TBC di berbagai daerah di Sulawesi Selatan. Misalnya, pada tahun 2018, jumlah kasus TBC di Kabupaten Barru dan Sidrap masing-masing sebanyak 182 dan 493 kasus, namun pada tahun 2020 meningkat menjadi 202 dan 591 kasus (BPS, 2020). Peningkatan ini diduga disebabkan oleh adanya pencilan (outlier) dalam data. Oleh karena itu, perlu dilakukan analisis untuk memahami faktor-faktor yang mempengaruhi penyebaran TBC. Untuk tujuan ini, penulis bermaksud menerapkan regresi kuantil dengan klasifikasi group LASSO pada data kasus TBC di Provinsi Sulawesi Selatan.

1.2 Rumusan Masalah

Berdasarkan latar uraian belakang masalah, dibuat perumusan masalah sebagai berikut.

1. Bagaimana bentuk estimasi parameter model regresi kuantil dengan Group LASSO?
2. Bagaimana model regresi kuantil dengan Group LASSO pada data kasus Tuberkulosis di Provinsi Sulawesi Selatan tahun 2021?

1.3 Batasan masalah

Berdasarkan rumusan masalah, maka ruang lingkup dalam penelitian ini dibatasi pada beberapa hal yaitu sebagai berikut:

1. Nilai kuantil yang digunakan dalam penelitian ini adalah $\theta = 0.25; 0.50; \text{ dan } 0.75$
2. Data Kasus Tuberkulosis dibatasi hanya pada tahun 2021.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Untuk mendapatkan bentuk estimasi parameter model regresi kuantil dengan Group LASSO
2. Untuk mendapatkan model regresi kuantil dengan Group LASSO dalam memodelkan data kasus Tuberkulosis di Provinsi Sulawesi Selatan tahun 2021

1.5 Manfaat Penelitian

Manfaat yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengembangkan wawasan keilmuan dan pengetahuan mengenai regresi kuantil dan klasifikasi group lasso
2. Memberikan informasi mengenai faktor-faktor yang mempengaruhi jumlah kasus Tuberkulosis di Provinsi Sulawesi Selatan tahun 2021

1.6 Kajian Teori

1.6.1 Analisis Regresi

Analisis regresi adalah metode statistik yang banyak digunakan dalam ilmu terapan untuk memahami hubungan antara variabel dalam sebuah model dan untuk tujuan peramalan. Galton menemukan bahwa meskipun terdapat kecenderungan bahwa orang tua yang tinggi memiliki anak-anak yang tinggi dan orang tua yang pendek memiliki anak-anak yang pendek, distribusi tinggi dalam suatu populasi tidak berubah secara signifikan dari satu generasi ke generasi berikutnya. Galton menjelaskan bahwa rata-rata tinggi anak-anak dari orang tua dengan tinggi tertentu cenderung kembali (regresi) mendekati tinggi rata-rata populasi secara keseluruhan (Rory, 2016).

Analisis regresi digunakan untuk melihat hubungan antara variabel respon dengan satu atau lebih variabel prediktor. Misalkan terdapat data berpasangan (x_i, y_i) untuk n pengamatan, maka hubungan antara variabel x_i dan variabel y_i dapat dinyatakan sebagai berikut :

$$y_i = f(x_i) + \varepsilon_i; i = 1, 2, \dots, n, \quad (1)$$

Dimana $f(x_i)$ adalah kurva regresi dan ε_i adalah error random yang diasumsikan identik, independen, dan berdistribusi normal dengan rata-rata nol dan varians σ^2 (Eubank, 1999). Analisis regresi memiliki dua tujuan utama. Pertama, untuk mengeksplorasi hubungan antara variabel respon dan variabel prediktor, dan kedua, untuk membuat prediksi (Rory, 2016).

1.6.2 Regresi Kuantil

Regresi kuantil adalah metode yang berguna untuk mengestimasi parameter, metode ini tidak mudah terpengaruh oleh kehadiran pencilon sehingga pencilon menjauh dan tidak mengganggu kestabilan data yang diperoleh (Idris et al., 2018). Regresi kuantil sangat berguna dalam berbagai bidang, diantaranya ekonometrika, biomedik, keuangan, kesehatan, lingkungan dan sebagainya.

Regresi kuantil adalah metode analisis regresi yang diperkenalkan oleh Koenker dan Bassett pada tahun 1978. Metode ini memperkirakan berbagai fungsi kuantil dari distribusi y sebagai fungsi dari X . Regresi kuantil sangat berguna ketika data tidak homogen (misalnya, varian y berubah dengan perubahan X) dan tidak simetris, serta dalam kasus adanya ekor pada distribusi atau distribusi yang terpotong (Koenker, 2005). Metode ini diterapkan secara luas dalam berbagai bidang, termasuk ekonometrika, biomedik, keuangan, kesehatan, dan lingkungan.

Persamaan umum regresi kuantil linier khusus untuk kuantil bersyarat $Q(\theta|X_{1i}, X_{2i}, \dots, X_{ki})$ dari variabel dependen Y_i yaitu :

$$Y_i = \beta_0(\theta) + \beta_1(\theta)X_{1i} + \dots + \beta_k(\theta)X_{ki} + \varepsilon_i(\theta), \quad (2)$$

dengan $i = 1, 2, \dots, n$.

$$\begin{aligned}
y_1(\theta) &= \beta_0(\theta) + \beta_1(\theta)x_{11} + \cdots + \beta_k(\theta)x_{k1} + \varepsilon_1(\theta) \\
y_2(\theta) &= \beta_0(\theta) + \beta_1(\theta)x_{21} + \cdots + \beta_k(\theta)x_{k2} + \varepsilon_2(\theta) \\
&\vdots \\
y_n(\theta) &= \beta_0(\theta) + \beta_1(\theta)x_{n1} + \cdots + \beta_k(\theta)x_{kn} + \varepsilon_n(\theta)
\end{aligned} \tag{3}$$

Jika model regresi kuantil disajikan dalam bentuk matriks, persamaan diatas dapat ditulis

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0(\theta) \\ \beta_1(\theta) \\ \vdots \\ \beta_k(\theta) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(\theta) \\ \varepsilon_2(\theta) \\ \vdots \\ \varepsilon_n(\theta) \end{bmatrix}$$

Selanjutnya persamaan diatas, dapat ditulis dalam bentuk model linier berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}(\theta) + \boldsymbol{\varepsilon}(\theta), \tag{4}$$

1.6.3 Estimasi Parameter Regresi Kuantil

Regresi kuantil mengestimasi parameter dengan meminimalkan jumlah absolut residual, yang dikenal dengan istilah Least Absolute Deviation (LAD). Regresi kuantil dilakukan dengan meminimumkan jumlah nilai mutlak dari error yang merupakan minimum penjumlahan error positif dan error negatif. Hal tersebut memberikan perbedaan bobot θ untuk error positif dan pembobot $1 - \theta$ untuk error negatif. Pada regresi kuantil median dapat mendefinisikan median sebagai solusi untuk meminimumkan jumlah nilai mutlak dari error (Uthami, 2013). Perkalian antara residual dengan bobot yang diberikan disebut sebagai fungsi kerugian (loss function) (ρ_θ), yang dinyatakan sebagai:

$$\rho_\theta(\varepsilon) = \sum_{i=1, u \geq 0}^n \theta |\varepsilon_i| + \sum_{i=1, u < 0}^n (1 - \theta) |\varepsilon_i| \tag{5}$$

Dengan demikian, dalam regresi kuantil terdapat fungsi kuantil ke $-\theta$ dari variabel Y dengan syarat X yang mempertimbangkan penduga $\beta(\theta)$, Hal ini menghasilkan solusi untuk permasalahan tersebut yang dinyatakan sebagai berikut:

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\theta(y_i - Q_\theta(Y|X)); \theta \in (0,1) \tag{6}$$

dengan :

$\rho_\theta(\cdot)$: Loss function (fungsi kerugian)

θ : Indeks kuantil dengan $\theta \in (0,1)$

$Q_\theta(Y|X)$: fungsi kuantil ke $-\theta$ dari variabel Y dengan syarat X

Dalam regresi kuantil, pada kuantil ke $-\theta$ dari F_y , meminimumkan *loss function* (7) adalah

$$\begin{aligned}\beta(\theta) &= \min_{\beta} \sum_{i=1}^n \rho_{\theta}(\varepsilon_i) \\ &= \min_{\beta} \sum_{i=1}^n \rho_{\theta}(\mathbf{y} - \mathbf{x}_i^T \boldsymbol{\beta})\end{aligned}\quad (7)$$

dimana $\theta \in (0,1)$ dan $\rho_{\theta}(\varepsilon_i)$ pada persamaan (2.9) didefinisikan

$$\rho_{\theta}(\varepsilon_i) = \begin{cases} (1 - \theta)\varepsilon_i, & \text{dengan } \varepsilon_i < 0, \\ \theta\varepsilon_i, & \text{dengan } \varepsilon_i \geq 0 \end{cases}$$

Jika Y merupakan fungsi X yang diketahui dan memiliki fungsi probabilitas $F_{Y|X}(y)$, maka kuantil ke $-\theta$ dari fungsi tersebut dapat dituliskan seperti pada persamaan berikut :

$$\min_q \theta \int_{y>q} |y - q| dF_Y(y) + (1 - \theta) \int_{y<q} |y - q| dF_Y(y). \quad (8)$$

dengan mempertimbangkan $\beta(\theta)$, maka diperoleh solusi untuk permasalahan tersebut sebagai berikut:

$$\hat{\beta}(\theta) = \min_{\beta} \left\{ \theta \sum_{i=1, y>q}^n |y_i - \mathbf{X}^T \boldsymbol{\beta}(\theta)| + (1 - \theta) \sum_{i=1, y<q}^n |y_i - \mathbf{X}^T \boldsymbol{\beta}(\theta)| \right\} \quad (9)$$

1.6.4 Algoritma Simpleks Regresi Kuantil

Algoritma simpleks adalah metode numerik untuk pendugaan parameter dalam regresi kuantil yang dikembangkan oleh Barrodam dan Robert pada tahun 1974. Metode ini digunakan untuk menentukan kombinasi optimal dari tiga variabel atau lebih dan dapat memberikan solusi untuk permasalahan program linier yang melibatkan banyak variabel. Bentuk umum formulasi metode primal dari permasalahan regresi kuantil linier adalah sebagai berikut (Balami, 2017).

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\theta}(\mathbf{y}_i - \mathbf{x}_i^T \boldsymbol{\beta}) \quad (10)$$

dan dapat dijabarkan

$$\min \{ \theta e_n^T u + (1 - \theta) e_n^T v \mid y - X\beta = u - v, \beta \in \mathbb{R}^p, (u, v) \in \mathbb{R}^{2n} \}, \quad (11)$$

Akan dicari bentuk kanonik dengan mengidentifikasi :

$$c = (0_p^T, \theta e_n^T, (1 - \theta) e_n^T)^T, \quad x = (\beta^T, u^T, v^T)^T, \quad A = [X: I - I], \quad \beta = y, \quad T = 0_n$$

Seperti yang dijelaskan di atas, sifat himpunan polyhedral dan fungsi objektif menunjukkan bahwa himpunan tersebut dapat diarahkan pada kendalanya. Dengan demikian, himpunan tersebut dapat diindeks sebagai $\binom{n}{p}$ pada elemen $h \in H$ yang diperoleh dari

$$\begin{aligned}\beta(h) &= X(h)^{-1} y(h), \\ u(h) &= v(h) = 0, \\ u(\bar{h}) &= (y - X\beta(h))^+, \\ v(\bar{h}) &= (y - X\beta(h))^-,\end{aligned}\tag{12}$$

Setiap titik memiliki batas slackness $u_i v_i = 0$, dan setidaknya terdapat indeks p , dengan $i \in h$, dimana $u_i = v_i = 0$. Titik tersebut mencerminkan kesesuaian terhadap pengamatan p dalam subset h dan mengatur hubungan antara vektor u dan v dari solusi yang sama terhadap bagian positif dan negatif, sehingga menghasilkan vektor residual. Permasalahan regresi kuantil primal memiliki masalah dual yang sesuai, yaitu:

$$\max\{y^T z \mid X^T z = 0, z \in [\theta - 1, \theta]^n\}.$$

dengan kendala $c - A^T y \in S^*$ menjadi

$$\begin{pmatrix} 0_p \\ \theta e_n \\ (1 - \theta)e_n \end{pmatrix} - \begin{pmatrix} X^T \\ I_n \\ -I_n \end{pmatrix} d \in 0_p \times \mathbb{R}_+^{2n},\tag{13}$$

dimana kondisi $yy \in T^*$ tidak mempengaruhi $z \in \mathbb{R}^n$. Dengan cara yang sama, permasalahan pada parameter dual semula untuk memecahkan solusi berikut:

$$\max\{y^T a \mid X^T e_n, a \in [0, 1]^n\}, a = d + (1 - \theta)e_n\tag{14}$$

1.6.5 Data Outlier

Data outlier atau pencilan dapat didefinisikan sebagai pengamatan yang menyimpang secara signifikan dari pengamatan lainnya. Keberadaan data outlier ini dapat memiliki dampak besar dalam pengambilan kesimpulan atau keputusan dalam penelitian. Ada beberapa cara untuk mengidentifikasi apakah suatu data merupakan outlier atau tidak. Salah satu metode pendeteksian adalah menggunakan metode grafik seperti box plot. Suatu observasi data dianggap outlier jika data tersebut terletak jauh dari whisker dalam box plot (Sihombing dkk., 2022).

1.6.6 Multikolinieritas

Multikolinieritas adalah kondisi di mana terjadi korelasi antara variabel prediktor, sehingga variabel-variabel tersebut tidak saling independen. Untuk mendeteksi adanya multikolinieritas, salah satu ukuran yang digunakan adalah Faktor Inflasi Ragam (Variance Inflation Factor / VIF). VIF berfungsi sebagai kriteria

untuk mendeteksi multikolinearitas dalam regresi linier yang melibatkan lebih dari dua variabel prediktor. Nilai VIF yang lebih besar dari 10 menunjukkan adanya masalah multikolinearitas yang serius. VIF untuk koefisien regresi j diidentifikasi sebagai berikut (Sriningsih dkk., 2018):

$$VIF_j = \frac{1}{1-R_j^2} \quad (15)$$

dengan

j : Jumlah variabel prediktor ; $j = 1, 2, \dots, k$

R_j^2 : Koefisien determinasi dari variabel prediktor X_j dengan variabel prediktor lain

Kriteria pengambilan keputusan untuk uji multikolinearitas adalah sebagai berikut (Ghozali, 2016) :

1. Jika nilai VIF < 10 atau nilai Tolerance > 0,01, maka dinyatakan tidak terjadi multikolinearitas.
2. Jika nilai VIF > 10 atau nilai Tolerance < 0,01, maka dinyatakan terjadi multikolinearitas.

Untuk mengatasi multikolinearitas dalam regresi, beberapa metode yang dapat digunakan adalah (Sunyoto, 2007):

1. Menghilangkan salah satu atau lebih variabel prediktor yang memiliki korelasi tinggi.
2. Mempertahankan variabel prediktor dengan korelasi tinggi hanya untuk tujuan prediksi, tanpa menginterpretasikan koefisiennya.
3. Mengurangi hubungan linier antar variabel prediktor dengan menggunakan logaritma natural (ln).
4. Menggunakan metode analisis alternatif, seperti Group LASSO

1.6.7 Regresi Kuantil dengan LASSO

LASSO adalah metode yang digunakan untuk memilih variabel dalam data berdimensi tinggi (Tibshirani, 1996). Teknik ini beroperasi dengan menambahkan batasan pada metode kuadrat terkecil. Model regresi dengan LASSO dapat dinyatakan sebagai berikut:

$$\hat{\beta}^{(LASSO)} = \arg \min_{\beta} \|y - \sum_{j=1}^J X_j \beta_j\|^2 + \lambda \sum_{j=1}^J |\beta_j|, \quad (16)$$

dimana λ adalah parameter penalti yang bernilai positif. Model pada persamaan diatas dikenal sebagai penalti l_1 . Semakin besar nilai λ , semakin banyak koefisien yang disusutkan menuju atau menjadi nol. Fan dan Li (2001) menunjukkan bahwa LASSO dapat menghasilkan estimasi yang bias untuk model dengan banyak variabel. Oleh karena itu, properti oracle tidak selalu terpenuhi dalam estimasi menggunakan LASSO.

Metode LASSO selain dapat digunakan dalam regresi linier biasa, juga dapat diterapkan dalam regresi lain seperti regresi logistik, regresi kuantil dan sebagainya. Secara umum, untuk mendapatkan penaksir parameter metode

LASSO dapat disederhanakan dengan meminimumkan fungsi *loss* dan menambahkan fungsi kendala LASSO sebagai berikut:

$$f(t) = L(t) + (P(t)) \quad (17)$$

Dengan t menunjukkan parameter model, $f(t)$ sebagai fungsi tujuan dari metode yang terpenalti, $L(t)$ menunjukkan fungsi dari model yang digunakan, sedangkan $P(t)$ menunjukkan fungsi kendala atau penalti yang ditambahkan dalam hal ini fungsi kendala LASSO.

Li dan Zhu (2008) menggunakan penalti LASSO pada regresi kuantil. Parameter diestimasi dengan meminimumkan fungsi:

$$\min_{\|\beta\|=1} \sum_{i=1}^n \rho_{\theta}(y - x_i^T \beta) + \lambda \sum_{j=1}^J |\beta_j| \quad (18)$$

1.6.8 Group LASSO

Group LASSO sering digunakan untuk pemilihan variabel dalam data dimana variabel prediktor membentuk kelompok (Chen & Xiang, 2017). Metode ini tidak hanya berguna untuk pemilihan variabel dan menangani multikolinearitas, tetapi juga dapat diterapkan pada data kategorik (El Sheikh dkk., 2021). Sebagai pengembangan dari LASSO, Group LASSO menambahkan kendala kelompok pada metode LASSO. Penelitian oleh Yuan & Lin (2006) menunjukkan bahwa Group LASSO, yang termasuk dalam keluarga LASSO, cukup efektif untuk data berdimensi tinggi dengan variabel prediktor yang terkelompok. Beberapa jenis analisis LASSO meliputi LASSO, Group LASSO, dan Fused LASSO. Dalam kasus data kasus tuberkulosis di Jawa Barat, di mana variabel penjelas banyak dan terkelompok, analisis yang digunakan adalah LASSO dan Group LASSO. Sebaliknya, Fused LASSO lebih cocok untuk data dengan struktur urutan, seperti data genetika dalam bidang biologi (Hastie dkk., 2015).

Untuk Vektor $\eta \in R^d$, $d \geq 1$ dan K adalah matriks simetris yang definit positif yang berdimensi $d \times d$, dapat dituliskan sebagai berikut:

$$\|\eta\|_K = (\eta' K \eta)^{1/2}. \quad (19)$$

Dituliskan sebagai $\|\eta\| = \|\eta\|_{I_d}$ untuk lebih singkatnya. Diberikan matriks definit positif K_1, \dots, K_j , estimasi group LASSO didefinisikan sebagai berikut:

$$\hat{\beta}^{(g.LASSO)} = \frac{1}{2} \left\| Y - \sum_{j=1}^J X_j \beta_j \right\|^2 + \lambda \sum_{j=1}^J \|\beta_j\| \quad (20)$$

Dimana J adalah banyak kelompok, X_j adalah variabel bebas ke- j , dan β_j adalah koefisien regresi ke- j . Sedangkan $\lambda \geq 0$ adalah parameter penyetelan. Bakin (1999) mengusulkan persamaan di atas sebagai pengembangan dari LASSO untuk memilih kelompok variabel dan mengusulkan algoritma komputasi.

Berdasarkan penelitian dari Li dan Zhu (2008) menggunakan penalti LASSO pada regresi kuantil maka diperoleh bentuk penalti group LASSO adalah sebagai berikut:

$$\lambda \sum_{j=1}^J \|\beta_j\| \quad (21)$$

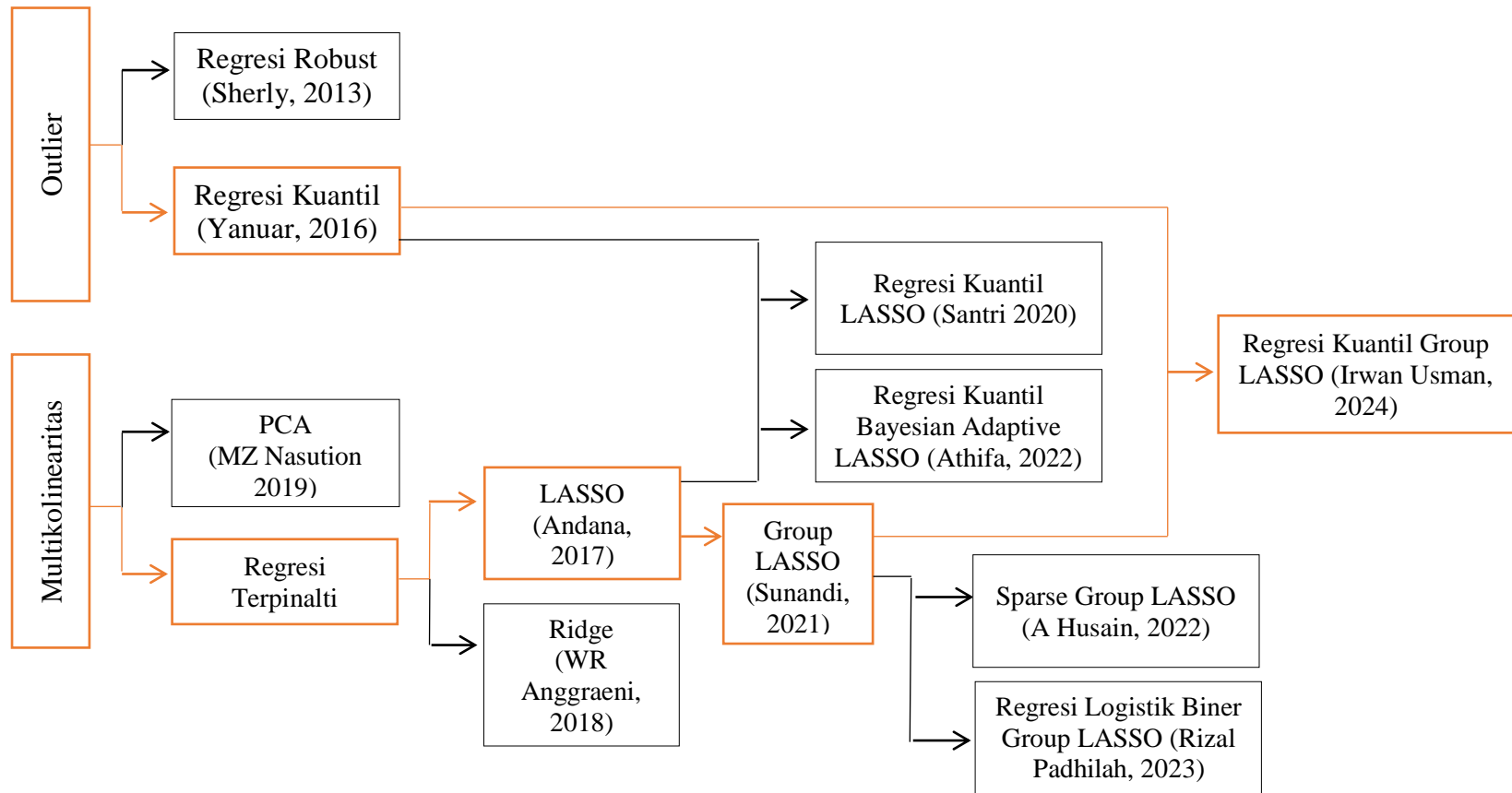
1.6.9 Tuberculosis (TBC)

Perhatian global terfokus pada deklarasi "kedaruratan global" tuberkulosis oleh WHO pada tahun 1993, mengingat banyak negara di seluruh dunia belum berhasil mengendalikan penyakit ini. Situasi ini disebabkan oleh tingginya angka penularan akibat rendahnya tingkat kesembuhan penderita. Tuberkulosis telah ada sejak ribuan tahun sebelum Masehi, dengan bukti penemuan penyakit ini pada mumi dari zaman Mesir kuno. Pada tahun 1882, ilmuwan Robert Koch berhasil mengidentifikasi kuman penyebab tuberkulosis, yaitu bakteri berbentuk batang yang dikenal sebagai *Mycobacterium tuberculosis* (Sejati, A & Sofiana, L., 2015).

Tuberkulosis merupakan penyakit kronik dan menular. Penyakit ini dapat menular dengan cepat kepada individu yang rentan dan memiliki daya tahan tubuh yang lemah, dengan satu penderita dapat menularkan penyakit ini kepada hingga 1 dari 10 orang di sekitarnya. Tuberkulosis berdampak signifikan pada sumber daya manusia dan umumnya lebih sering terjadi pada kelompok masyarakat dengan golongan sosial ekonomi rendah. (Sejati, A & Sofiana, L., 2015).

Di seluruh dunia, sekitar 10 juta orang menderita tuberkulosis (TBC) setiap tahun. TBC adalah salah satu dari 10 penyebab kematian teratas dan merupakan penyebab utama infeksi tunggal, melampaui HIV/AIDS. Penyakit ini dapat menyerang siapa saja, di mana saja, tetapi sebagian besar penderita TBC (sekitar 90%) adalah orang dewasa, dengan rasio pria dan wanita adalah 2:1. Tingkat kejadian TBC bervariasi secara signifikan di tingkat nasional, mulai dari kurang dari 50 hingga lebih dari 5000 kasus per 1 juta penduduk per tahun. Hampir 90% dari kasus ini terjadi di 30 negara dengan beban TBC yang tinggi (WHO, 2018).

1.7 Kerangka Konsep



Gambar 1. Kerangka Konsep

BAB II

METODOLOGI PENELITIAN

Bagian ini menjelaskan metode penelitian yang digunakan untuk mencapai tujuan penelitian, meliputi sumber data, variabel penelitian, dan langkah-langkah penelitian. Secara singkat, metodologi penelitian dalam tesis ini dimulai dengan studi literatur mengenai konsep dasar regresi kuantil dan Group LASSO. Setelah itu, metode tersebut diterapkan pada data kasus tuberkulosis di Provinsi Sulawesi Selatan tahun 2021.

2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder, berupa jumlah kasus TBC tahun 2021 yang diperoleh dari situs web Badan Pusat Statistik Sulawesi Selatan dan Open Data SulSel. Data tersebut mencakup informasi mengenai faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis di Provinsi Sulawesi Selatan.

2.2 Variabel Penelitian

Variabel penelitian yang digunakan terdiri dari satu variabel respon (y) dan 21 variabel prediktor (x). Variabel respon dalam penelitian ini adalah jumlah kasus tuberkulosis, sementara variabel prediktor mencakup faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis. Detail variabel-variabel ini dapat dilihat pada Tabel 3.1 berikut ini:

Tabel 1. Variabel Penelitian

KELOMPOK VARIABEL	VARIABEL	NAMA VARIABEL	SATUAN
Lingkungan	x_1	Luas Wilayah	Km ²
	x_2	Luas Kawasan Kumuh	HA
	x_3	Jumlah Rumah Tidak Layak Huni	Unit
	x_4	Jumlah Bank Sampah	Unit
Kependudukan	x_5	Jumlah Penduduk	Orang
	x_6	Kepadatan Penduduk	Jiwa/Km ²
	x_7	Laju Pertumbuhan Penduduk	Persen/Tahun
	x_8	Jumlah Penduduk Miskin	Orang
Kesehatan	x_9	Jumlah Rumah Sakit	Unit
	x_{10}	Jumlah Puskesmas	Unit

KELOMPOK VARIABEL	VARIABEL	NAMA VARIABEL	SATUAN
Ekonomi	x_{11}	Jumlah Perawat	Orang
	x_{12}	Jumlah Bidan	Orang
	x_{13}	Jumlah Kunjungan Wisatawan	Orang
	x_{14}	Jumlah Usaha Mikro Dan Kecil	Unit
	x_{15}	Jumlah Tenaga Kerja Industri Kecil Menengah	Orang
SDM	x_{16}	Jumlah Petani	KK
	x_{17}	Jumlah Nelayan	Orang
	x_{18}	Jumlah PKK	Kelompok
Pendidikan	x_{19}	Jumlah Sekolah	Unit
	x_{20}	Jumlah Murid	Orang
	x_{21}	Jumlah Pengunjung Perpustakaan	Orang

Identifikasi untuk setiap variabel adalah sebagai berikut.

- Jumlah Kasus Tuberkulosis di tiap kabupaten/kota di Provinsi Sulawesi Selatan tahun 2021 (Y)
Kasus Tuberkulosis adalah kasus responden yang pernah didiagnosis menderita Tuberkulosis oleh dokter.
- Luas Wilayah (X_1)
Luas wilayah merujuk pada area yang termasuk dalam batas kekuasaan teritorial suatu wilayah.
- Luas Kawasan Kumuh (X_2)
Kawasan kumuh adalah area dengan tingkat kepadatan populasi yang tinggi di sebuah kota, biasanya dihuni oleh masyarakat miskin dan memiliki kondisi hidup yang buruk.
- Jumlah Rumah Tidak Layak Huni (X_3)
Rumah tidak layak huni adalah rumah yang tidak memenuhi standar minimum untuk dihuni, dimana konstruksi bangunan tidak memadai, luasnya tidak sesuai dengan standar per orang, serta tidak memenuhi syarat kesehatan atau bahkan membahayakan penghuninya.
- Jumlah Bank Sampah (X_4)
Bank sampah adalah fasilitas yang digunakan untuk mengumpulkan dan mengelola sampah, dengan tujuan untuk mendaur ulang dan mengurangi jumlah limbah yang dibuang.
- Jumlah Penduduk (X_5)
Penduduk adalah orang yang berdomisili disuatu wilayah.

7. Kepadatan penduduk (X_6)
Kepadatan penduduk adalah rasio banyaknya penduduk per kilometer persegi.
8. Laju Pertumbuhan Penduduk (X_7)
Laju pertumbuhan penduduk adalah tingkat pertumbuhan penduduk per tahun dalam jangka waktu tertentu.
9. Jumlah Penduduk Miskin (X_8)
Penduduk miskin adalah individu yang berada di bawah garis kemiskinan, yang mencerminkan nilai pengeluaran minimum yang diperlukan untuk memenuhi kebutuhan pokok hidup selama sebulan, termasuk kebutuhan makanan dan non-makanan.
10. Jumlah Rumah Sakit (X_9)
Jumlah rumah sakit adalah total unit rumah sakit yang terdapat di suatu wilayah tertentu.
11. Jumlah Puskesmas (X_{10})
Jumlah Puskesmas adalah total unit Puskesmas yang terdapat di suatu wilayah tertentu.
12. Jumlah Perawat (X_{11})
Jumlah Perawat adalah total tenaga kesehatan yang berprofesi sebagai perawat dalam pelayanan kesehatan.
13. Jumlah Bidan (X_{12})
Jumlah Bidan adalah total tenaga kesehatan yang berprofesi sebagai bidan dalam pelayanan kesehatan.
14. Jumlah Kunjungan Wisatawan (X_{13})
Jumlah Kunjungan Wisatawan adalah total wisatawan, baik dari dalam negeri maupun luar negeri, yang berkunjung ke suatu daerah.
15. Jumlah Usaha Mikro dan Menengah (X_{14})
Jumlah Usaha Mikro dan Menengah adalah total aktivitas usaha yang dilakukan oleh perorangan atau badan usaha milik perorangan dalam kategori mikro dan menengah.
16. Jumlah Tenaga Kerja Industri Kecil Menengah (X_{15})
Jumlah Tenaga Kerja Industri Kecil Menengah adalah total orang yang bekerja di industri kecil dan menengah.
17. Jumlah Petani (X_{16})
Petani adalah seseorang yang bergerak dibidang pertanian. Pada penelitian ini digunakan semua jenis petani.
18. Jumlah Nelayan (X_{17})
Nelayan adalah seseorang yang bekerja menangkap ikan. Pada penelitian ini nelayan yang dimaksud adalah nelayan yang menangkap ikan baik di air asin maupun air tawar
19. Jumlah PKK (X_{18})
Jumlah PKK (Pembinaan Kesejahteraan Keluarga) adalah total organisasi kemasyarakatan yang berfokus pada pemberdayaan wanita dan peningkatan kesejahteraan keluarga.

20. Jumlah Sekolah (X_{19})
Jumlah Sekolah adalah total unit sekolah jenjang SD, SMP, SMK, dan SMA baik sekolah negeri maupun swasta.
21. Jumlah Murid (X_{20})
Jumlah Murid adalah total murid yang ada di jenjang SD, SMP, SMK, dan SMA baik sekolah negeri maupun swasta.
22. Jumlah Pengunjung Perpustakaan (X_{21})
Jumlah pengunjung perpustakaan adalah totalss pengunjung yang datang ke perpustakaan pada wilayah tertentu.

2.3 Metode Analisis

Pemodelan regresi kuantil dengan klasifikasi group LASSO terhadap faktor-faktor yang mempengaruhi jumlah kasus TBC di Sulawesi selatan.

- a. Melakukan eksplorasi data awal atau Exploratory Data Analysis (EDA) dengan menggunakan analisis statistika deskriptif.
- b. Mendeteksi adanya outlier atau pencilan dengan menggunakan box plot.
- c. Melakukan uji asumsi, yaitu dengan menguji multikolinearitas melalui analisis korelasi antarvariabel dan perhitungan Variance Inflation Factor (VIF).
- d. Melakukan komputasi regresi kuantil dengan penalti Group LASSO menggunakan paket grpreg dalam pemrograman R.
- e. Menganalisis dan menginterpretasikan hasil yang diperoleh dari koefisien variabel prediktor.
- f. Menarik kesimpulan dari hasil analisis.