SKRIPSI

PENGGUNAAN METODE REGRESI AKAR LATEN *ROBUST* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS DAN *OUTLIER*

Disusun dan diajukan oleh

INDRY ANGELIN

H051171002

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

PENGGUNAAN METODE REGRESI AKAR LATEN *ROBUST* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS DAN *OUTLIER*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains
pada Program Studi Statistika Departemen Statistika Fakultas
Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

INDRY ANGELIN

H051171002

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN
MAKASSAR
2021

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini :

Nama	: Indry Angelin
NIM	: H051171002
Program Studi	: Statistika
Jenjang	: Sarjana (S1)

Menyatakan dengan ini bahwa karya tulisan saya berjudul

PENGGUNAAN METODE REGRESI AKAR LATEN *ROBUST* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS DAN *OUTLIER*

Adalah karya tulisan saya sendiri dan bukan merupakan pengambil alihan tulisan orang lain bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 9 Maret 2021

METERAI TEMPEL
3B1BBADF471194235
6000
ENAM RIBU RUPIAH

INDRY ANGELIN

NIM. H051171002

PENGGUNAAN METODE REGRESI AKAR LATEN *ROBUST* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS DAN *OUTLIER*

Disetujui Oleh:

Pembimbing Utama,

Drs. Raupong, M.Si.

NIP. 19621015 198810 1 001

Pembimbing Pertama,

Andi Kresna Jaya, S.Si., M.Si.

NIP. 19731228 200003 1 001

Ketua Program Studi

Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2002

Pada Tanggal : 9 Maret 2021

LEMBAR PENGESAHAN (TUGAS AKHIR)

PENGGUNAAN METODE REGRESI AKAR LATEN *ROBUST* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS DAN *OUTLIER*

Disusun dan diajukan oleh

INDRY ANGELIN

H051171002

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin pada tanggal 9 Maret 2021 dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	**Pembimbing Pertama,**
Drs. Raupong, M.Si.	**Andi Kresna Jaya, S.Si., M.Si.**
NIP. 19621015 198810 1 001	**NIP. 19731228 200003 1 001**

Ketua Program Studi

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS HASANUDDIN DEPARTEMEN STATISTIKA FAKULTAS MIPA

Dr. Nurtiti Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2002

iv

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, atas berkat dan kasih karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir ini yang berjudul "Pengggunaan Metode Regresi Akar Laten *Robust* Pada Data Yang Mengandung Multikolinearitas dan *Outlier*" sebagai salah satu syarat akademik untuk meraih gelar sarjana pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Dalam menyelesaikan tugas akhir ini, ada banyak pihak yang turut membantu baik dalam memberikan dukungan maupun motivasi untuk dapat menyelesaikan tugas akhir ini.

Pada kesempatan ini, penulis mengucapkan terima kasih yang sebesarnya kepada Ayahanda **Simon Matta** dan Ibunda **Martha Limbong** atas pengorbanan, kesabaran, dukungan, semangat, dan doanya yang menjadi kekuatan bagi penulis sehingga penulis mampu berada dititik saat ini. Terima kasih pula penulis ucapkan kepada kakakku tersayang **Gabriel Marsi** atas segala bentuk motivasi, doa, dan semangat yang telah diberikan kepada penulis. Serta untuk keluarga besar penulis, terima kasih atas doa dan dukungan yang telah diberikan selama ini.

Penghargaan yang tulus serta ucapan terima kasih yang sebesarnya juga penulis ucapkan kepada :

1. Ibu **Prof. Dr. Dwia Aries Tina Pulubuhu, MA,** selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. Bapak **Dr. Eng. Amiruddin,** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin beserta seluruh jajarannya.
3. Ibu **Dr. Nurtiti Sunusi, S.Si., M.Si.,** selaku Ketua Departemen Statistika dan Bapak **Andi Kresna Jaya, S.Si., M.Si.,** selaku Sekretaris Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin dan **segenap dosen pengajar** serta **staf Departemen Statistika,** terima kasih telah

memberikan ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa.

4. Bapak **Drs. Raupong, M.Si.,** selaku pembimbing utama dan Bapak **Andi Kresna Jaya, S.Si., M.Si.,** selaku pembimbing pertama dan pembimbing akademik yang telah meluangkan waktunya serta dengan penuh kesabaran membimbing, memberikan motivasi dan semangat, serta saran sehingga tugas akhir ini dapat terselesaikan dengan baik.
5. Ibu **Dr. Erna Tri Herdiani, S. Si., M.Si.** dan Ibu **Dr. Dr. Georgina Maria Tinungki, M.Si.,** selaku penguji yang telah meluangkan waktunya serta memberikan saran dan kritik yang membangun sehingga membuat tugas akhir ini menjadi lebih baik.
6. Sahabat seperjuangan yang selalu menemani penulis selama masa perkuliahan dengan segala tingkah lakunya yang konyol dan rasa perhatian satu sama lain, **Chika, Novi, Salsa,** dan **Sri**. Terima kasih atas kebersamaan, kebahagiaan, kesedihan, dan kebaikannya selama masa perkuliahan.Untuk **Sahabat Tiktok Online (Eva, Haura,** dan **Puhe)** yang selalu memberikan semangat dan motivasi, serta menghibur penulis selama penyusunan tugas akhir ini.
7. Sahabat-sahabat **Sobat Libur (Andreas, Adolf, Arung, Ayu, Elitha, Ema, Gheo, Kevin,** dan **William)** yang selalu ada ketika penulis membutuhkan bantuan, serta selalu memberikan semangat dan doa untuk penulis dalam menyelesaikan tugas akhir ini.
8. Teman-teman seperjuangan **Statistika 2017,** terima kasih atas kebersamaan serta canda dan tawanya selama menjalani pendidika di Departemen Statistika.
9. Teman-teman **KKN Gelombang 104 Kecamatan Tamalanrea** yang selalu memberikan dukungan kepada penulis selama penyusunan tugas akhir ini.
10. Kepada semua pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun tidak langsung dalam penyusunan tugas akhir ini yang tidak sempat penulis sebutkan satu persatu.

Penulis menyadari bahwa tugas akhir ini masih sangat jauh dari kesempurnaan dan tentunya masih banyak terdapat kekurangan. Oleh karena itu, segala kritik dan saran yang membangun sangat penulis terima. Akhir kata, semoga tugas akhir ini dapat memberikan manfaat bagi semua pihak.

Makassar, 5 Maret 2021

Penulis

ABSTRAK

Analisis regresi merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan antara variabel respon dengan variabel prediktor dalam suatu penelitian. Namun, apabila terjadi multikolinearitas pada variabel prediktor yang digunakan maka mengakibatkan model regresi yang diperoleh jauh dari akurat. Untuk mengatasi masalah tersebut maka digunakan metode regresi akar laten yang merupakan perluasan dari analisis komponen utama. Dalam mengatasi masalah multikolinearitas pada model regresi, regresi akar laten membentuk komponen utama yang diperoleh dari matriks korelasi gandengan dengan variabel respon dan variabel prediktor yang telah dibakukan. Selain masalah multikolinearitas, masalah lain pada model regresi adalah adanya pencilan pada data. Adanya pencilan pada data tidak dapat diatasi dengan menggunakan regresi akar laten, oleh sebab itu digunakan suatu metode estimasi yang disebut regresi *robust* dengan metode estimasi MM. Metode estimasi MM mampu mengatasi masalah pencilan baik dari variabel respon maupun variabel prediktor dan memiliki *high breakdown point* sebesar 50%. Untuk memperoleh estimator yang kekar terhadap pencilan perlu dilakukan iterasi hingga diperoleh hasil yang konvergen, sehingga digunakan *Iteratively Reweighted Least Square* (IRLS) yang pembobotnya berubah setiap pengiterasian dan dalam hal ini, pembobot yang digunakan adalah pembobot *Tukey Bisquare*.

Kata Kunci: Regresi Akar Laten, Multikolinearitas, Pencilan, Regresi Robust, *Tukey Bisquare*.

ABSTRACT

Regression analysis is one of statistical method used to analyze the relationship between the response variable and predictor variable in a research. However, if there is multicollinearity in the predictor variable used, the regression model obtained is far from accurate. To solve this problem, the latent root regression method is used which is an expansion of the principal component analysis. To solve this problem, the latent root regression method is used which is an expansion of the principal component analysis. In overcoming the multicollinearity problem in the regression model, latent root regression forms the principal component obtained from the augmented correlation matrix with standardized response variable and predictor variable. In addition to multicollinearity problem, another problem in regression model is the existence of outlier in data. The existence of outlier in data can not be solved by using latent root regression, therefore an estimation method called robust regression with the MM estimation method is used. The MM estimation method is able to solve the outlier problem of both the response variable and predictor variable and has a high breakdown point of 50%. To obtain a robust estimator for outlier it is necessary to iterate until convergent result are obtained, so that the Iteratively Reweighted Least Square (IRLS) is used, whose weight changes with each iteration and in this case, the weight used is the Tukey Bisquare weight.

Keywords: Latent Root Regression, Multicollinearity, Outlier, Robust Regression, Tukey Bisquare.

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1 Latar Belakang

Aspek kesehatan merupakan salah satu aspek kualitas sumber daya manusia yang penting untuk diperhatikan di seluruh dunia dan juga sebagai pencapaian komitmen internasional yang tertuang pada tujuan *Sustainable Development Goals* (SDGs). Salah satu indikator penting untuk mengetahui derajat kesehatan di suatu negara dan bahkan untuk mengukur tingkat kemajuan suatu bangsa adalah angka kematian bayi. Angka Kematian Bayi (AKB) adalah jumlah kematian bayi usia 0-11 bulan yang dinyatakan dalam 1000 kelahiran hidup pada tahun yang sama (Dinkes Bali, 2018). Berdasarkan hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017 yang menunjukkan bahwa Angka Kematian Bayi (AKB) sebesar 24 per 1000 kelahiran hidup telah mengalami penurunan dari tahun-tahun sebelumnya, seperti pada tahun 2007 yang masih mencapai 34 per 1000 kelahiran hidup dan tahun 2012 sebesar 32 per 1000 kelahiran hidup. Meski cenderung turun, tapi negara-negara yang tergabung dalam *Sustainable Development Goals* (SDGs) termasuk Indonesia masih menargetkan penurunan Angka Kematian Bayi (AKB) yang dapat dicegah pada bayi baru lahir dan balita, dimana setiap negara menargetkan untuk mengurangi kematian neonatal setidaknya menjadi kurang dari 12 per 1000 kelahiran hidup dan kematian balita menjadi 25 per 1000 kelahiran hidup pada tahun 2030.

Angka kematian bayi dapat dipengaruhi oleh beberapa faktor, antara lain pemberian ASI eksklusif pada bayi usia 0-6 bulan yang tidak terpenuhi, jumlah sarana kesehatan yang minim, tidak rutin memeriksakan kesehatan bayi, pemberian vitamin A pada bayi usia 6-11 bulan tidak terpenuhi, dan kurangnya jumlah tenaga medis. Untuk mengetahui pengaruh faktor-faktor tersebut, maka dapat digunakan analisis regresi. Analisis regresi merupakan salah satu metode statistika yang digunakan untuk menganalisis hubungan antara variabel respon dengan variabel prediktor dalam suatu penelitian. Di dalam analisis regresi, hubungan yang sebenarnya tidak dapat diketahui secara pasti tetapi model hubungan tersebut dapat di estimasi berdasarkan data pengamatan. Terdapat dua jenis regresi, yaitu regresi linier dan regresi non linier. Untuk regresi linier dibedakan menjadi dua, yaitu regresi linier sederhana yang hanya melibatkan satu variabel prediktor dan analisis regresi linier berganda yang melibatkan/dipengaruhi oleh lebih dari satu variabel prediktor (Myers & Milton, 1991). Pada analisis regresi linier berganda dengan banyak variabel prediktor, permasalahan yang sering timbul adalah adanya multikolinearitas.

Multikolinearitas merupakan suatu kondisi adanya hubungan (korelasi) antara dua atau lebih variabel bebasnya. Adanya multikolinearitas yang tinggi pada variabel bebas mengakibatkan model regresi yang diperoleh jauh dari akurat, sehingga tidak memungkinkan melihat pengaruh variabel bebas terhadap variabel responnya (Gujarati, 1992). Terdapat beberapa metode untuk mengatasi adanya multikolinearitas dalam regresi linier berganda, salah satunya adalah regresi akar laten (*latent root regression*). Metode regresi akar laten (*latent root regression*) merupakan perluasan dari metode regresi komponen utama. Perluasan dari regresi komponen utama diajukan oleh J. T. Webster, R. F. Gunst, dan R. L. Mason, dalam "*Latent root regression analysis*", *Technometrics*, 16, 1974. Webster dan rekannya menggabungkan matriks data yang berasal dari variabel respon dengan variabel prediktor yang telah dibakukan. Perluasan ini dinamakan regresi akar laten (*latent root regression*) (Draper & Smith, 1992). Pada regresi komponen utama, variabel-variabel bebas yang saling berkorelasi di ubah ke dalam bentuk variabel-variabel baru (komponen utama) yang bersifat saling bebas dan selanjutnya diregresikan dengan variabel respon (Marcus, Wattimanela, & Lesnussa, 2012). Sementara komponen utama yang terbentuk pada regresi akar laten diperoleh dari matriks korelasi gandengan dengan variabel bebas dan variabel respon yang telah dibakukan, sehingga komponen utama pada regresi akar laten lebih banyak mengandung informasi dibandingkan regresi komponen utama (Vigneau & Qannari, 2002).

Beberapa penelitian terdahulu telah menggunakan *Latent Root Regression* (LRR), di antaranya Dwi, dkk (2014) menerapkan *Latent Root Regression* (LRR) pada model regresi linier berganda, Edi, dkk (2014) menggunakan *Latent Root Regression* (LRR) untuk memprediksi penjualan mobil, dan Desy dan Mathilda (2017) menerapkan *Latent Root Regression* (LRR) pada faktor-faktor yang mempengaruhi IHSG di Bursa Efek Indonesia. Akan tetapi, penggunaan metode *Latent Root Regression* (LRR) pada masalah multikolinearitas tidak efisien jika data tidak berdistribusi normal yang seringkali disebabkan karena adanya *outlier* (pencilan) pada data yang digunakan. Oleh karena itu, dibutuhkan estimator yang *robust* terhadap adanya pencilan. Notiragayu dan Khoirin (2008) menggunakan regresi komponen utama *robust* untuk data mengandung pencilan, Sony (2011) menggunakan pendekatan *Robust Principal Component Analysis* (ROBPCA) untuk mengatasi masalah multikolinearitas dan *outlier* pada data Angka Kematian Bayi (AKB) di Jawa Timur, dan Siska, dkk (2020) menggunakan regresi komponen utama *robust* dengan Metode *Minimum Covariance Determinant – Least Trimmed Square* (MCD-LTS). Berdasarkan uraian tersebut, penulis akan mengkaji tentang penggunaan metode regresi akar laten *robust* pada data yang mengandung multikolinearitas dan *outlier* (studi kasus: Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan tahun 2017).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana model *Robust Latent Root Regression* (RLRR) pada data yang mengandung multikolinearitas dan *outlier* (studi kasus: Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan tahun 2017)?

2. Bagaimana pengaruh variabel prediktor terhadap variabel respon pada kasus Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan tahun 2017?

1.3 Batasan Masalah

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

1. Penggunaan *robust* MM-*estimator* pada regresi akar laten.
2. Data yang digunakan adalah data Angka Kematian Bayi (AKB) Provinsi Sulawesi Selatan tahun 2017.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai pada penelitian ini adalah sebagai berikut:

1. Untuk memperoleh model *Robust Latent Root Regression* (RLRR) pada data yang mengandung multikolinearitas dan *outlier* (studi kasus: Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan tahun 2017).
2. Untuk memperoleh pengaruh variabel prediktor terhadap variabel respon pada kasus Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan tahun 2017.

1.5 Manfaat Penelitian

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah wawasan pengetahuan mengenai model *Robust Latent Root Regression* (RLRR) pada data yang mengandung multikolinearitas dan *outlier*.
2. Memberikan informasi mengenai faktor-faktor yang mempengaruhi angka kematian bayi di Provinsi Sulawesi Selatan dengan menggunakan model *Robust Latent Root Regression* (RLRR), sehingga dapat menjadi acuan bagi pemerintah dan masyarakat dalam upaya menurunkan angka kematian bayi di Indonesia khususnya di Provinsi Sulawesi Selatan.

BAB II

TINJAUAN PUSTAKA

2.1 Matriks

Matriks adalah suatu susunan bilangan-bilangan dalam bentuk baris dan kolom berbentuk persegi panjang atau persegi yang ditulis di antara dua tanda kurung, yaitu ( ) atau [ ] (Anton, 2001). Bilangan-bilangan dalam susunan itu disebut sebagai anggota dalam matriks atau elemen matriks. Elemen-elemen matriks yang mendatar membentuk baris dan elemen-elemen vertikal membentuk kolom. Banyaknya baris dan kolom suatu matriks menyatakan ukuran matriks tersebut. Apabila dalam suatu matriks terdapat *m* baris dan *n* kolom maka ukuran matriks tersebut adalah $m \times n$, matriks tersebut dapat dinyatakan dalam bentuk berikut :

$$\boldsymbol{A}_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

atau dapat ditulis $\boldsymbol{A}_{m\times n} = \left[a_{ij}\right]_{m\times n}$, dimana a_{ij} adalah elemen matriks dengan $i = 1,\ 2,\ \ldots,\ m$ dan $j = 1,\ 2,\ \ldots,\ n$.

2.1.1 *Transpose* Matriks

Jika suatu matriks A berordo $m \times n$, maka *transpose* **A** dinyatakan oleh $\mathbf{A}^{\mathrm{T}}$ dan didefinisikan sebagai matriks $n \times m$ yang diperoleh dengan menukarkan baris – baris dan kolom – kolom dari **A** sehingga kolom pertama dari $\mathbf{A}^{\mathrm{T}}$ adalah baris pertama dari **A**, kolom kedua dari $\mathbf{A}^{\mathrm{T}}$ adalah baris kedua dari **A**, dan seterusnya (Anton & Rorres, 2004). Jika terdapat **A** berukuran $m \times n$ yang dinyatakan sebagai berikut :

$$\boldsymbol{A}_{m\times n}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

maka $\mathbf{A}^{\mathrm{T}}$ dapat dituliskan seperti bentuk berikut :

$$\boldsymbol{A}^{T}=\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

2.1.2 Matriks Varians-Kovarians dan Matriks Korelasi

Suatu matriks yang elemen-elemennya terdiri atas variansi dan kovariansi dari sekumpulan variabel disebut dengan matriks varians-kovarians. Matriks varians-kovarians dari sampel dinotasikan dengan $\mathbf{\Sigma}$ atau $\boldsymbol{S}$ dan dapat dinyatakan dalam bentuk berikut ini :

$$\mathbf{\Sigma}=\begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

Oleh karena $\sigma_{ij}=\sigma_{ji}$, maka untuk setiap $i=1,\ 2,\ \ldots\ ,n$ dan $j=1,\ 2,\ \ldots\ ,n$ dengan $i = j$ berlaku :

$$\mathbf{\Sigma}=\begin{bmatrix} \sigma_{11} & \sigma_{21} & \cdots & \sigma_{n1} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{bmatrix}$$

Korelasi atau biasa disebut dengan koefisien korelasi merupakan ukuran keeratan hubungan linier antara variabel acak A_i dan A_j yang dinyatakan dalam varians σ_{ii} dan kovarians σ_{ij} seperti pada persamaan berikut:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} \tag{2.1}$$

Adapun matriks koefisien korelasi adalah matriks simetri berukuran $n \times n$ yang dinyatakan dalam persamaan berikut :

$$\rho_{n\times n} = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{21}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{11}}} & \cdots & \frac{\sigma_{n1}}{\sqrt{\sigma_{nn}}\sqrt{\sigma_{11}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{n2}}{\sqrt{\sigma_{nn}}\sqrt{\sigma_{22}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1n}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{nn}}} & \frac{\sigma_{2n}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{nn}}} & \cdots & \frac{\sigma_{nn}}{\sqrt{\sigma_{nn}}\sqrt{\sigma_{nn}}} \end{bmatrix} \tag{2.2}$$

$$\rho_{n\times n} = \begin{bmatrix} 1 & \rho_{21} & \cdots & \rho_{n1} \\ \rho_{12} & 1 & \cdots & \rho_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} & \rho_{2n} & \cdots & 1 \end{bmatrix}$$

2.1.3 Nilai Eigen dan Vektor Eigen

Jika **A** merupakan matriks berukuran $n \times n$, maka terdapat nilai eigen atau nilai karakteristik (λ) dan vektor eigen atau vektor karakteristik (x) yang bersesuaian dengan λ sehingga dipenuhi $\boldsymbol{A}x = \lambda x$ dengan $x \neq 0$ (Sembiring, 1995).

Untuk mengetahui nilai eigen dari matriks **A** yang berukuran $n \times n$ dapat dihitung dengan persamaan berikut :

$$\left|\left(\boldsymbol{A} - \lambda \boldsymbol{I}\right)\right| = 0 \tag{2.3}$$

Persamaan (2.3) dinamakan persamaan karakteristik dari matriks **A**. Akar-akar atau skalar-skalar yang memenuhi persamaan inilah yang disebut nilai-nilai eigen dari matriks **A**. Nilai eigen disebut juga akar laten (*latent root*). Sementara untuk mengetahui vektor eigen (vektor laten) yang bersesuaian dengan suatu nilai eigen dapat dicari dengan melalui persamaan berikut ini :

$$\left(\boldsymbol{A}-\lambda\boldsymbol{I}\right)x=0 \tag{2.4}$$

2.2 Analisis Regresi Linier Berganda

Analisis regresi merupakan salah satu teknik analisis data dalam statistika yang seringkali digunakan untuk mengkaji hubungan antara beberapa variabel dan meramal suatu variabel (Kutner, Nachtsheim, & Neter, 2004). Menurut Drapper dan Smith (1992), hubungan antara satu variabel respon dengan satu atau lebih variabel prediktor dapat dinyatakan dalam regresi linier berganda. Hubungan tersebut dapat dinyatakan secara umum sebagai berikut :

$$\begin{aligned} Y_i &= \beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_p X_{ip}+\varepsilon_i \\ Y_i &= \beta_0+\sum_{j=1}^{p}\beta_j X_{ij}\ ;\ i=1,\ 2,\ \ldots\ ,\ n\ \ ;\ \ n>p \end{aligned} \tag{2.5}$$

dengan y_i adalah variabel respon pada pengamatan ke-*i*, β_0 adalah intersep, β_j adalah koefisien regresi ke-*j*, x_{ij} adalah variabel prediktor ke-*j* pada pengamatan ke-*i*, dan ε_i adalah *error* pada pengamatan ke-*i* yang diasumsikan independen berdistribusi normal dengan rataan nol dan varians σ^2 (Montgomery, Peck, & Vining, 2012).

Dalam notasi matriks, persamaan (2.5) dapat ditulis menjadi sebagai berikut :

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon} \tag{2.6}$$

dimana $\boldsymbol{Y}=\begin{bmatrix} y_1 & y_2 & \cdots & y_n\end{bmatrix}^T$ adalah vektor variabel respon berukuran $n\times 1$, $\boldsymbol{X}$ adalah matriks variabel prediktor berukuran $n\times(p+1)$, $\boldsymbol{\beta}=\begin{bmatrix}\beta_0 & \beta_1 & \cdots & \beta_p\end{bmatrix}^T$ adalah vektor dari parameter regresi berukuran $(p+1)\times 1$, dan $\boldsymbol{\varepsilon}=\begin{bmatrix}\varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n\end{bmatrix}^T$ adalah vektor *error* berukuran $n\times 1$.

Dengan metode *Ordinary Least Square* (OLS) yang bertujuan untuk meminimumkan jumlah kuadrat *error*, diperoleh penaksir (*estimator*) OLS untuk $\boldsymbol{\beta}$ sebagai berikut (Kutner, Nachtsheim, & Neter, 2004) :

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X} \right)^{-1} \boldsymbol{X}^T \boldsymbol{Y} \tag{2.7}$$

2.3 Multikolinearitas

Pada tahun 1934 istilah multikolinearitas pertama kali diperkenalkan oleh Ragnar Frisch, yang berarti adanya hubungan (korelasi) diantara variabel-variabel prediktor dari model regresi. Multikolinearitas merupakan adanya hubungan linier antara variabel prediktor dalam suatu model regresi linier berganda (Gujarati D. N., 2004). Adanya multikolinearitas pada model regresi tentu dapat memberikan dampak terhadap model tersebut, yaitu sebagai berikut (Neter, 1997):

1. Mengakibatkan variansi penduga kuadrat terkecil menjadi besar yang tentunya akan menghasilkan galat baku yang lebih besar sehingga berakibat terhadap selang kepercayaan untuk parameter model regresi menjadi lebih besar.
2. Adanya variabel prediktor yang saling berkolerasi berakibat pada penjelasan yang diberikan terhadap variabel respon menjadi tidak akurat (memberikan pengaruh yang sama terhadap variabel respon).
3. Hasil pengujian hipotesis parameter berdasarkan metode kuadrat terkecil menjadi tidak valid.

Beberapa cara yang dapat dilakukan untuk mendeteksi adanya multikolinearitas pada data pengamatan antara lain sebagai berikut (Naes, Isaksson, Fearn, & Davies, 2002) :

1. Menghitung koefisien korelasi antara variabel prediktor. Jika nilai koefisien korelasi lebih besar dari 0,8 maka hal tersebut menunjukkan adanya multikolinearitas antar variabel prediktor.
2. Menghitung *Variance Inflation Factor* (VIF). Jika nilai VIF lebih besar dari 10, berarti terdapat masalah multikolinearitas di dalam regresi. Adapun cara untuk menghitung nilai VIF adalah sebagai berikut (Neter, 1997) :

$$VIF(j) = \frac{1}{1-R_j^2}\ ;\ R_j^2 \text{ adalah nilai koefisien determinasi} \tag{2.8}$$

3. Menghitung nilai TOL, yang merupakan suatu ukuran *tolerance* untuk deteksi multikolinearitas dengan cara sebagai berikut:

$$TOL_j = \frac{1}{VIF_j} = 1 - R_j^2 = \begin{cases} 1, \text{ jika } X_j \text{ tidak berkorelasi dengan variabel prediktor lainnya} \\ 0, \text{ jika } X_j \text{ berkorelasi dengan variabel prediktor lainnya} \end{cases} \tag{2.9}$$

2.4 Regresi Akar Laten

Regresi akar laten (*Latent Root Regression*) merupakan salah satu metode yang digunakan dalam mengatasi adanya multikolinearitas pada model regresi linier berganda. Regresi akar laten merupakan metode perluasan dari regresi komponen utama. Letak perbedaan antara kedua metode ini yaitu pada nilai akar laten yang dihasilkan dari matriks korelasi. Matriks korelasi pada regresi akar laten diperoleh dari penggabungan variabel respon dan prediktor yang telah dibakukan seperti pada persamaan berikut (Draper & Smith, 1992) :

$$\boldsymbol{Z}^* = [\boldsymbol{Z_Y}, \boldsymbol{Z_X}] \tag{2.10}$$

dengan $\boldsymbol{Z}^*$ adalah penggabungan matriks **Y** dan **X** yang telah dibakukan, $\boldsymbol{Z_Y}$ adalah matriks **Y** yang telah dibakukan, dan $\boldsymbol{Z_X}$ adalah matriks **X** yang telah dibakukan.

Pembakuan data pada variabel respon diperoleh melalui rumus berikut :

$$Z_Y = \frac{(\mathbf{Y} - \bar{y})}{\sqrt{U_Y}} \text{ dengan } \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},\ \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n},\ \ U_Y = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2 \tag{2.11}$$

Sementara pembakuan data pada variabel prediktor diperoleh melalui rumus berikut :

$$Z_X = \frac{\left(\boldsymbol{X}_{ij} - \bar{x}_j\right)}{\sqrt{U_X}} \quad ; \quad i = 1,\ 2,\ \ldots,\ n\ ;\ j = 1,\ 2,\ \ldots,\ p \ \text{ dengan}$$

$$\boldsymbol{X}_{ij} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1,p} \\ x_{21} & x_{22} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n,p} \end{bmatrix},\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},\ U_X = \sum_{i=1}^{n}\left(X_{ij} - \bar{x}_j\right)^2 \tag{2.12}$$

Maka bentuk persamaan (2.10) setelah matriks **Y** dan **X** dibakukan adalah sebagai berikut (Draper & Smith, 1992) :

$$\boldsymbol{Z}^* = \begin{bmatrix} Z_{1y} & Z_{11} & \cdots & Z_{1,p} \\ Z_{2y} & Z_{21} & \cdots & Z_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{ny} & Z_{n1} & \cdots & Z_{n,p} \end{bmatrix} \tag{2.13}$$

Dari persamaan (2.13), maka selanjutnya dapat dihitung matriks korelasi gandengan (*augmented correlation matrix*) yang dapat dituliskan seperti berikut :

$$\boldsymbol{Z}^{*T}\boldsymbol{Z}^* = \begin{bmatrix} 1 & \gamma_{2Y} & \cdots & \gamma_{rY} \\ \gamma_{2Y} & 1 & \cdots & \gamma_{r1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{rY} & \gamma_{r1} & \cdots & 1 \end{bmatrix} \tag{2.14}$$

Seperti halnya pada analisis komponen utama, akar laten dan vektor laten diperoleh melalui matriks korelasi gandengan melalui persamaan berikut (Webster, Gunst, & Mason, 1974) :

$$\left|\boldsymbol{Z}^{*T}\boldsymbol{Z}^* - \lambda \boldsymbol{I}\right| = 0 \tag{2.15}$$

dan

$$\left(\boldsymbol{Z}^{*T}\boldsymbol{Z}^* - \lambda_j \boldsymbol{I}\right)\gamma_j = 0 \ ;\ j = 0,\ 1,\ 2,\ \ldots\ , p \tag{2.16}$$

Misalkan $\gamma_j^T = \left(\gamma_{0j},\ \gamma_{1j},\ \gamma_{2j},\ \ldots\ , \gamma_{rj}\right)$ merupakan vektor laten ke-*j* dari matriks korelasi $\boldsymbol{Z}^{*T}\boldsymbol{Z}^*$ dan $\gamma_j^0 = \left(\gamma_{1j},\ \gamma_{2j},\ \ldots\ , \gamma_{rj}\right)$ merupakan vektor laten yang terbentuk dari elemen yang sama dengan γ_j^T kecuali elemen pertama yang telah dibuang, maka komponen utama dari matriks $\boldsymbol{Z}^*$ adalah sebagai berikut :

$$PC_j = \boldsymbol{Z}_y \gamma_{0j} + \boldsymbol{Z}_x \gamma_j^0 \tag{2.17}$$

Untuk menentukan komponen utama yang akan digunakan pada tahap analisis, dilakukan dengan membuang komponen utama yang bersesuaian dengan nilai akar laten $\lambda_j \leq 0.05$ dan elemen pertama vektor laten $\left|\gamma_{0j}\right| < 0.10$ (Webster, Gunst, & Mason, 1974).

Kemudian, untuk menghitung vektor koefisien kuadrat terkecil termodifikasi pada regresi akar laten dilakukan dengan menggunakan rumus berikut (Webster, Gunst, & Mason, 1974) :

$$\boldsymbol{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_r^* \end{bmatrix} = c\sum_{j=0}^{p} \gamma_{oj} \lambda_j^{-1} \begin{bmatrix} \gamma_{1j} \\ \gamma_{2j} \\ \vdots \\ \gamma_{rj} \end{bmatrix} ; \quad c = -\left\{\sum_{j=0}^{p} \gamma_{oj}^2 \lambda_j^{-1}\right\}^{-1} \left\{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2\right\}^{1/2} \tag{2.18}$$

dengan λ_j adalah akar laten ke-*j* dari matriks $\boldsymbol{Z}^{*T}\boldsymbol{Z}^*$, γ_j adalah elemen vektor laten ke-*j*, dan γ_{0j} adalah elemen pertama dari vektor laten ke-*j*. Sementara penduga koefisien regresi pada variabel asal diperoleh dengan membagi penduga koefisien regresi pada peubah yang telah dibakukan dengan S_j sehingga diperoleh (Draper & Smith, 1992) :

$$\beta_j = \frac{\beta_j^*}{S_j} \; ; \quad S_j = \sqrt{\sum_{j=1}^{p} \left(x_j - \overline{x}_j\right)^2} \tag{2.19}$$

Sedangkan perhitungan intersep β_0 diperoleh berdasarkan rumus berikut (Draper & Smith, 1992) :

$$\beta_0 = \overline{y} - \beta_1 \overline{x}_1 - \beta_2 \overline{x}_2 - \ldots - \beta_p \overline{x}_p \tag{2.20}$$

Setelah persamaan kuadrat terkecil termodifikasinya diperoleh, Webster dan rekan-rekannya menyarankan untuk melakukan eliminasi langkah mundur (*backward*) untuk mengeluarkan peubah peramal dari persamaan itu (Webster, Gunst, & Mason, 1974).

2.5 Pencilan

Menurut Weisberg (2005), pencilan (*outlier*) merupakan suatu data yang tidak biasa, tidak cocok dari data lainnya atau data yang tidak mengikuti pola umum dari keseluruhan. Keberadaan pencilan akan mengganggu dalam proses analisis data dan harus dihindari dalam banyak hal. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal-hal berikut (Soemartini, 2007) :

1. *Error* yang besar dari model yang terbentuk atau $E(e) \neq 0$.
2. Varians pada data tersebut menjadi lebih besar.
3. Taksiran interval memiliki rentang yang lebar.

Pencilan dapat muncul karena kesalahan dalam memasukkan data, kesalahan pengukuran, analisis, atau kesalahan-kesalahan lain. Adakalanya pencilan menjadi sumber informasi yang tidak diberikan oleh titik data lainnya. Sehingga apabila menghapus data yang merupakan pencilan untuk memperbaiki persamaan yang cocok juga dapat berbahaya, karena tindakan tersebut dapat menimbulkan kesalahan ketelitian dalam mengestimasi atau memprediksi.

Terdapat tiga tipe pencilan mempengaruhi mempengaruhi estimasi kuadrat terkecil di antaranya (Soemartini, 2007) :

1. Pencilan vertikal (*vertical outlier*), merupakan pengamatan yang terpencil pada variabel respon (Y), namun tidak terpencil pada variabel prediktor (X) dan sangat berpengaruh pada estimasi intersep.
2. *Good leverage point*, merupakan pengamatan pencilan terhadap variabel prediktor (X) namun terletak dekat dengan garis regresi, artinya bahwa pengamatan x_i menjauh tetapi y_i cocok dengan garis regresi. Tipe ini berpengaruh terhadap inferensi statistic karena dapat meningkatkan estimasi standar eror.

3. *Bad leverage point*, merupakan pengamatan pencilan terhadap variabel prediktor (X) yang terletak jauh dari garis regresi dan berpengaruh signifikan terhadap estimasi kuadrat terkecil, baik terhadap intersep maupun *slope* dari persamaan regresi.

Adapun metode yang dapat digunakan untuk mengidentifikasi adanya pencilan (*outlier*) yang berpengaruh dalam koefisien regresi antara lain :

1. Metode Grafis

   Salah satu keuntungan dalam menggunakan metode ini ialah mudah dipahami karena menampilkan data secara grafis (gambar) dan tanpa melibatkan perhitungan yang rumit. Sementara kelemahan dari metode ini adalah keputusan bahwa suatu data merupakan pencilan sangat bergantung pada kebijakan (*judgement*) peneliti, karena hanya mengandalkan visualisasi grafis, untuk itu dibutuhkan seseorang yang ahli dan berpengalaman dalam menginterpretasikan plot tersebut. Contoh metode grafis adalah diagram pencar (*scatter plot*) dan *box plot*.

2. Metode *DfFITS* (*Difference fitted value FITS*) atau *Standardized DfFITS*

   Metode ini merupakan suatu ukuran yang berpengaruh yang timbul oleh pengamatan ke-*i* terhadap nilai taksiran $\hat{y}_i$. Nilai untuk $DfFITS_i$ dapat diperoleh melalui persamaan berikut :

$$\left(DfFITS\right)_i = t_i\sqrt{\frac{h_{ii}}{1-h_{ii}}} \quad , \; t_i : \textit{ externally studentized residual} \qquad (2.21)$$

   dengan h_{ii} adalah elemen diagonal ke-*i* dari matriks H dengan persamaan seperti berikut :

$$H = X\left(X^T X\right)^{-1} X^T \qquad (2.22)$$

   Suatu pengamatan ke-*i* pada data diidentifikasi sebagai pencilan jika nilai :

$$\left|DfFITS_i\right| > 1 \qquad\qquad untuk\ n \le 30$$

$$\left|DfFITS_i\right| > 2\left(\frac{p}{n}\right)^{1/2} \qquad\qquad untuk\ n > 30$$

dengan *p* adalah banyaknya parameter dan *n* adalah banyaknya pengamatan.

2.6 Fungsi Obyektif

Fungsi obyektif adalah fungsi yang digunakan untuk mencari fungsi pembobot pada regresi *robust*. Adapun fungsi pembobot yang digunakan adalah fungsi pembobot *Tukey Bisquare* dengan fungsi obyektif sebagai berikut :

$$\rho\left(e_i^*\right) = \begin{cases} \dfrac{c^2}{6}\left[1-\left[1-\left(\dfrac{e_i^*}{c}\right)^2\right]^3\right], & \left|e_i^*\right| < c \\ \dfrac{c^2}{6}, & \left|e_i^*\right| \ge c \end{cases} \tag{2.23}$$

dan fungsi *influence* yaitu :

$$\psi\left(e_i^*\right) = \rho'\left(e_i^*\right) = \frac{\partial\left(\rho\left(e_i^*\right)\right)}{\partial e_i^*} = \begin{cases} e_i^*\left[1-\left(\dfrac{e_i^*}{c}\right)^2\right]^2, & \left|e_i^*\right| < c \\ 0, & \left|e_i^*\right| \ge c \end{cases} \tag{2.24}$$

Sehingga diperoleh fungsi pembobot

$$w_i = w\left(e_i^*\right) = \frac{\psi\left(e_i^*\right)}{e_i^*} = \begin{cases} \left[1-\left(\dfrac{e_i^*}{c}\right)^2\right]^2, & \left|e_i^*\right| < c \\ 0, & \left|e_i^*\right| \ge c \end{cases} \tag{2.25}$$

dimana e_i^* adalah galat yang distandarisasi dan parameter *c* adalah *tunning constant* yang telah ditetapkan masing – masing pembobot untuk menentukan tingkat yang *robust* suatu pembobot. Adapun nilai *c* untuk pembobot *Tukey* adalah 4.685 (Nurmiati, Raupong, & Anna, 2014).

2.7 Regresi *Robust*

Regresi *robust* adalah salah satu metode regresi yang digunakan ketika distribusi dari galat (*error*) tidak normal dan atau adanya beberapa pencilan (*outlier*) yang berpengaruh pada model (Ryan, 1997). Regresi *robust* merupakan metode penting untuk menganalisis data yang dipengaruhi oleh pencilan sehingga dihasilkan model *robust* atau kekar terhadap pencilan. Suatu estimasi yang *robust* relatif tidak terpengaruh oleh perubahan kecil pada bagian besar data (Ryan, 1997). Pada regresi *robust* tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang kecil untuk data *outlier* (Soemartini, 2007).

Resistensi dan efisiensi merupakan dua hal penting yang sangat diperlukan dalam estimasi regresi *robust*. Suatu estimasi dikatakan resisten terhadap *outlier* apabila sebagian kecil dari data tidak dapat memberikan efek yang terlalu besar terhadap estimator, sedangkan estimasi memiliki efisiensi yang cukup baik pada berbagai sebaran jika ragamnya mendekati ragam minimum untuk setiap sebaran (Montgomery, Peck, & Vining, 1982). Adapun metode estimasi yang dapat digunakan dalam regresi *robust* antara lain (Chen, 2002).

1. Estimasi LTS (*Least Trimmed Squares*), merupakan metode yang menggunakan konsep pengepasan metode kuadrat terkecil untuk meminimumkan jumlah kuadrat eror (Akbar & Maftukhah, 2007) dan memiliki *high breakdown point*.
2. Estimasi S (*Scale*), merupakan salah satu metode penduga regresi *robust* yang memiliki *high breakdown point* yang diperlukan dalam mengidentifikasi *bad observation* untuk membedakan *good leverage point* dan *bad leverage point*. Meskipun memiliki *high breakdown point* yang sama dengan estimasi LTS, tetapi estimasi S memiliki efisiensi yang lebih tinggi.
3. Estimasi M (*Maximum likelihood type*), merupakan metode yang menganalisis data dengan mengasumsikan bahwa sebagian besar yang terdeteksi pencilan pada variabel independen.

4. Estimasi MM (*Method of Moment*), merupakan metode yang menggabungkan antara estimasi S yang memiliki *high breakdown point* dan estimasi M.

2.7.1 Estimasi M

Metode estimasi M merupakan salah satu metode estimasi dalam regresi *robust* yang paling sederhana dan paling sering digunakan. Estimasi M baik digunakan dalam mengatasi pencilan vertikal. Pada prinsipnya, estimasi M merupakan estimasi yang meminimumkan fungsi obyektif (ρ) dari *error*-nya (Montgomery, Peck, & Vining, 1982).

$$\min_{\beta} \rho\left(e_i\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(y_i - \sum_{j=0}^{p} x_{ij}\beta_j \right) \tag{2.26}$$

Dalam mengestimasi parameter regresi pada regresi *robust* menggunakan estimasi M, diperlukan suatu metode iterasi karena *error* tidak dapat dihitung sampai diperoleh model yang cocok dan parameter regresi juga tidak dapat dihitung tanpa mengetahui nilai *error*. Adapun metode iterasi yang banyak digunakan adalah *Iteratively Reweighted Least Square* (IRLS).

2.7.2 Estimasi S

Metode estimasi S merupakan metode *high breakdown point* yang pertama kali diperkenalkan oleh Rousseeuw dan Yohai pada tahun 1984. Metode ini dapat mengidentifikasi *bad observation* untuk membedakan *good leverage point* dan *bad leverage point* yang tidak dapat diatasi pada estimasi M. Estimasi S didefinisikan sebagai berikut :

$$\hat{\beta}_s = \arg\min_{\beta} \hat{\sigma}_s\left(e_1,\ e_2,\ \ldots,\ e_n\right) \tag{2.27}$$

Estimator pada metode regresi *robust* estimasi S diperoleh dengan cara melakukan iterasi hingga diperoleh hasil yang konvergen.

2.7.3 Estimasi MM

Metode estimasi MM merupakan metode yang pertama kali diperkenalkan oleh Yohai pada tahun 1987 yang menggabungkan suatu *high breakdown point* (50%) dengan efisiensi tinggi (mencapai 95%). Langkah pertama dalam estimasi MM adalah dengan mencari estimator S yang sangat *robust* dan resisten yang meminimumkan suatu skala *error*. Kemudian skala *error* tetap konstan dan diakhiri dengan menetapkan parameter – parameter regresi menggunakan estimasi M. Estimasi MM mempunyai *breakdown point* yang tinggi sama dengan estimasi S yaitu sebesar 50%, sehingga estimasi MM dapat menjelaskan bahwa banyak pencilan hingga separuh data pengamatan tidak berpengaruh terhadap estimasi MM. Estimasi MM didefinisikan sebagai berikut :

$$\hat{\beta}_{MM} = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{p} x_{ij}\beta_j}{\hat{\sigma}}\right) \tag{2.28}$$

2.8 *Breakdown Point*

Breakdown point adalah persentase nilai terkecil dari banyaknya data yang terkontaminasi atau banyaknya pencilan yang menyebabkan nilai dari taksiran menjadi besar (Rousseuw & Leroy, 1987). *Breakdown point* merupakan salah satu cara untuk menjelaskan ukuran ke-*robust*-an (kekerasan) dari suatu estimator. Kemungkinan tertinggi *breakdown point* untuk sebuah estimator adalah 50% yang dimiliki oleh MM-*estimator*. Apabila *breakdown poin*t lebih dari 50%, berarti estimasi model regresi tidak dapat menggambarkan informasi dari kebanyakan data.

2.9 Angka Kematian Bayi

Angka Kematian Bayi (AKB) adalah jumlah kematian bayi usia 0-11 bulan yang dinyatakan dalam 1000 kelahiran hidup pada tahun yang sama (Dinkes Bali, 2018) atau secara matematis dapat dituliskan menjadi sebagai berikut :

$$AKB = \frac{Jumlah\ Kematian\ Bayi}{Jumlah\ Kelahiran\ Bayi} \times 1000 \tag{2.29}$$

Angka kematian bayi merupakan salah satu indikator penting dalam menentukan tingkat kesehatan masyarakat karena dapat menggambarkan kesehatan penduduk secara umum dan termasuk di dalam salah satu target *Sustainable Development Goals* (SDGs). Namun, masih tingginya angka kematian bayi menjadi salah satu masalah kesehatan utama di Indonesia. Berdasarkan hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2017 yang menunjukkan bahwa Angka Kematian Bayi (AKB) sebesar 24 per 1000 kelahiran hidup telah mengalami penurunan dari tahun-tahun sebelumnya, seperti pada tahun 2007 yang masih mencapai 34 per 1000 kelahiran hidup dan tahun 2012 sebesar 32 per 1000 kelahiran hidup. Walaupun AKB berdasarkan hasil SDKI beberapa kurun waktu telah menunjukkan penurunan, namun bila dibandingkan dengan dengan negara lain, AKB di Indonesia masih tergolong tinggi.

Dari sisi penyebabnya, kematian bayi ada dua macam yaitu kematian bayi endogen dan kematian bayi eksogen. Kematian bayi endogen atau biasa disebut dengan kematian neonatal adalah kematian bayi yang disebabkan oleh faktor-faktor yang dibawa anak sejak lahir, yang diperoleh dari orang tuanya selama dalam kandungan dan terjadi pada bulan pertama setelah dilahirkan. Sedangkan kematian bayi eksogen atau kematian post-neonatal adalah kematian bayi yang disebabkan oleh faktor-faktor yang berhubungan dengan pengaruh lingkungan sekitar dan terjadi setelah satu bulan sampai menjelang usia satu tahun.