

**ESTIMASI PARAMETER REGRESI *RIDGE ROBUST*
PADA DATA PROFIL KESEHATAN SULAWESI
SELATAN**

SKRIPSI



HENDRIETE TIUR MAROWI WAIBUSI

H12116701

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

**ESTIMASI PARAMETER REGRESI *RIDGE ROBUST*
PADA DATA PROFIL KESEHATAN SULAWESI
SELATAN**

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**



HENDRIETE TIUR MAROWI WAIBUSI

H12116701

**PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

2023

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

**ESTIMASI PARAMETER REGRESI *RIDGE ROBUST*
PADA DATA PROFIL KESEHATAN SULAWESI SELATAN**

Adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun

Makassar, 11 Januari 2023

Hendriete Tiur Marowi Waibusi

NIM. H12116701

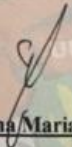
Universitas Hasanuddin

**ESTIMASI PARAMETER REGRESI *RIDGE ROBUST*
PADA DATA PROFIL KESEHATAN SULAWESI SELATAN**


Disetujui Oleh:

Pembimbing Utama

Pembimbing Pertama

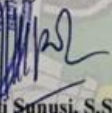

Dr. Dr. Georgina Maria Tinungki, M.Si.

NIP. 19620926 198702 2 001


Sitti Sahrinan, S.Si., M.Si.

NIP.19881018 201504 2 002

Ketua Program Studi


Dr. Nurfitri Sunusi, S.Si., M.Si.

NIP. 19720117 199703 2 002

Pada 11 Januari 2023

iv

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh:

Nama : Hendriete Tiur Marowi Waibus
NM : H12116701
Program Studi : Statistika
Judul Skripsi : Estimasi Parameter Regresi *Ridge Robust* Pada Data Profil Kesehatan Sulawesi Selatan

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratn yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

UNIVERSITAS HASANUDDIN

DEWAN PENGUJI

1. Ketua : Dr. Dr. Georgina Maria Tinungki, M.Si. (.....)
2. Sekretaris : Sitti Sahriman, S.Si., M.Si. (.....)
3. Anggota : Drs. Raupong, M.Si. (.....)
4. Anggota : Siswanto, S.Si., M.Si. (.....)

Ditetapkan di : Makassar
Tanggal : 11 Januari 2023

v

KATA PENGANTAR

Segala puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan kasihNya, yang telah menyertai penulis selama proses penyelesaian skripsi dengan judul “Estimasi Parameter Regresi *Ridge Robust* Pada Data Profil Kesehatan Sulawesi Selatan” sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

Dalam penyelesaian skripsi ini, penulis telah melewati perjuangan panjang dan pengorbanan. Namun, karena berkat dan kasih Tuhan serta dukungan dari berbagai pihak yang turut membantu sehingga penulisan tugas akhir ini dapat terselesaikan. Penulis menyampaikan ucapan terima kasih yang setinggi-tingginya kepada Ayahanda Marthen Kauki Waibusi, S.P., M.Si dan Ibunda tercinta Pdt. Kartini Sitinjak, S.Th., atas doanya yang tak pernah putus, serta kasih sayang yang melimpah dalam mendidik dan membesarkan penulis dengan begitu banyak pengorbanan yang tak pernah ternilai harganya. Juga untuk kakak penulis Kartika G. T. Waibusi, S.Ip., Tifani M. Waibusi, S.Si(Theo)., Ostina Waibusi, S.Si. dan adik penulis Marthen G. Waibusi, Paleria W. Waibusi yang telah memberikan motivasi, semangat serta dukungan kepada penulis yang begitu besar. Penghargaan yang tulus dan ucapan terima kasih yang sebesar-besarnya juga penulis ucapkan kepada:

1. Bapak Prof. Dr. Ir. Jamaluddin Jompa, M.Sc. selaku Rektor Universitas Hasanuddin beserta seluruh jajarannya.
2. Bapak Dr. Eng. Amiruddin, M.Si. 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 dan segenap dosen pengajar yang telah memberikan ilmu pengetahuan kepada penulis selama menjadi mahasiswa di Departemen Statistika.
4. Ibu Dr. Dr. Georgina Maria Tinungki, M.Si selaku pembimbing utama yang dengan tulus meluangkan waktunya dengan penuh kesabaran memberikan bimbingan dan membagi ilmunya kepada penulis.
5. Ibu Sitti Sahrman, S.Si., M.Si. selaku pembimbing pertama yang telah

bersedia memberikan bimbingan memotivasi penulis dalam penyusunan tugas akhir ini.

6. Bapak Dr. Raupong, M.Si. selaku pembimbing akademik dan anggota penguji yang telah memberikan banyak bantuan selama penulis menjalani pendidikan dan memberikan kritik dan saran yang membangun untuk penulisan tugas akhir ini.
7. Bapak Siswanto, S.Si., M.Si. selaku anggota penguji yang telah memberikan masukan kepada penulis dalam perbaikan skripsi ke arah yang lebih baik.
8. Segenap Staf Pegawai Departemen Statistika yang telah membantu proses administrasi selama penulis menyelesaikan tugas akhir.
9. Sahabat terbaik penulis Fransiskus Hasian Hutahaean, SE. yang telah memberikan semangat dan motivasi selama penyusunan skripsi.
10. Para pejuang skripsi mahasiswa Afirmasi angkatan 2016 Amelia Hisage, S.Kep. Ns, Fazrina Nuraini Letsoin, SE., Sofia I. Wanma, Jhosua H.J. Dimara, SE. yang telah memberikan semangat dan motivasi selama penyusunan skripsi.
11. Teman – teman penghuni Ramsis pejuang gelar Apoteker Lia, Nailil dan Amanda yang selalu menemani dan memberi semangat serta motivasi dalam penulisan skripsi.
12. Keluarga Besar Statistika UNHAS 2016 terima kasih atas motivasi, ilmu, dan bantuan yang diberikan kepada penulis baik sebelum dan selama penyusunan skripsi.
13. Senior Statistika 2015 terkhusus kepada Masjidil Aqsha, S.Si. terima kasih atas motivasi, ilmu, dan bantuan yang diberikan kepada penulis baik sebelum dan selama penyusunan skripsi.
14. Seluruh warga Statistika UNHAS, sebagai keluarga yang memberi nasehat-nasehat sehingga penulis dapat lebih mengerti arti pentingnya kebersamaan dan semua pihak yang telah banyak membantu penulis dan tak sempat penulis sebutkan satu per satu. Semoga segala bantuan dan partisipasinya dibalas oleh Tuhan Yang Maha Esa.
15. Semua pihak yang telah memberikan bantuan dan dukungan secara langsung maupun tidak langsung yang tidak sempat penulis sebutkan satu-persatu,

semoga kita semua selalu dalam lindungan dan berkat Tuhan.

Penulis menyadari bahwa masih banyak kekurangan dalam skripsi ini, untuk itu dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga tulisan ini memberikan manfaat bagi para pembaca.

Makassar, 11 Januari 2023

Hendriete Tiur Marowi Waibusi

**PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIK**

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Hendriete Tiur Marowi Waibusi
NM : H12116701
Program Studi : Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Non-esklusif (*Non-exclusive Royalty Free Right*)** atas tugas akhir saya yang berjudul :

**“Estimasi Parameter Regresi *Ridge Robust*
Pada Data Profil Kesehatan Sulawesi Selatan”**

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal, 11 Januari 2023

Yang menyatakan

Hendriete Tiur Marowi Waibusi

ABSTRAK

Multikolinieritas adalah salah satu pelanggaran asumsi dalam analisis regresi. Keberadaan multikolinieritas menyebabkan standar error meningkat. Regresi *ridge* adalah salah satu pendekatan analisis regresi yang dapat mengatasi masalah tersebut. Selain multikolinieritas, masalah lain yang sering terjadi adalah terdapat pencilan. Keberadaan pencilan menyebabkan data tidak berdistribusi normal. Regresi *robust* merupakan metode regresi yang digunakan untuk mengatasi data dari masalah pencilan. *Least trimmed square* adalah salah satu penduga dari *robust* yang dapat mengatasi masalah pencilan dengan meminimumkan jumlah kuadrat error. Data yang digunakan dalam penelitian ini adalah data Data Profil Kesehatan Sulawesi Selatan Tahun 2017 yang mengalami masalah multikolinieritas dan pencilan. Tujuan dari penelitian ini adalah untuk mendapatkan faktor-faktor yang mempengaruhi Angka Kematian Bayi (AKB) sebagai variabel respon (Y) dengan variabel prediktornya adalah kepadatan penduduk (X_1), persentase rumah tangga berperilaku hidup bersih dan sehat (X_2), jumlah bayi berat badan lahir rendah (X_3), jumlah bayi yang diberi asi eksklusif (X_4), jumlah bayi yang mendapat kunjungan neonatal 1 kali (X_5), jumlah bayi yang mendapat pelayanan kesehatan (X_6). Hasil analisis data menunjukkan 4 variabel prediktor yang signifikan mempengaruhi AKB diantaranya kepadatan penduduk (X_1), jumlah bayi yang diberi asi eksklusif (X_4), jumlah bayi yang mendapat kunjungan neonatal 1 kali (X_5), jumlah bayi yang mendapat pelayanan kesehatan (X_6). Nilai koefisien determinasi yang dihasilkan sebesar 88% artinya ada sekitar 88% keragaman variabel respon AKB (Y) dijelaskan oleh variabel prediktor (X) dan nilai *Mean Square Error* (MSE) sebesar 1.96. Dengan demikian model yang dihasilkan cukup baik berdasarkan nilai R^2 dan MSE yang dihasilkan.

Kata kunci: Regresi *Ridge Robust*, *Least Trimmed Square*, Multikolinieritas, Pencilan, Angka Kematian Bayi.

ABSTRACT

Multicollinearity is one of the assumption violations in regression analysis. The existence of multicollinearity causes the standard error to increase. Ridge regression is one of the regression analysis approaches that can overcome this problem. Besides multicollinearity, another problem that often occurs is outliers. The existence of outliers causes the data not to be normally distributed. Robust regression is a regression method used to overcome data from outlier problems. Trimmed least squares is one of the strongest estimators that can overcome the outlier problem by minimizing the total squared error. The data used in this study is Health Profile Data for South Sulawesi in 2017 which experiences multicollinearity problems and outliers problems. The purpose of this study was to obtain the factors that influence the Infant Mortality Rate (IMR) as the response variable (Y) with the predictor variable being population density (X_1), percentage of houses households with clean and healthy behavior (X_2), number of low birth weight babies (X_3), number of babies who receive exclusive breastfeeding (X_4), number of babies who get 1 neonatal visit (X_5), number of babies who receive health services (X_6). The results of data analysis showed that 4 predictor variables had a significant effect on IMR including population density (X_1), number of babies who were exclusively breastfed (X_4), number of babies who received 1 neonatal visit (X_5), number of babies who received health services (X_6). The resulting coefficient of determination is 88%, meaning that there is around 88% of the diversity of the IMR response variable (Y) explained by the predictor variable (X) and the Mean Square Error (MSE) value of 1.96. Thus the resulting model is quite good based on the resulting R^2 and MSE values.

Keywords: *Robust Ridge Regression, Least Trimmed Square, Multicollinearity, Outliers, Infant Mortality Rate.*

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL	ii
LEMBAR PERNYATAAN KEOTENTIKAN	iii
HALAMAN PERSETUJUAN PEMBIMBINGError! Bookmark not defined.	
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	vi
HALAMAN PERNYATAAN PERSETUJUAN PUBLKASI	ix
ABSTRAK	x
ABSTRACT	xi
DAFTAR ISI	xii
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xv
DAFTAR LAMPIRAN	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
BAB II TINJAUAN PUSTAKA	4
2.1 Analisis Regresi Linear	4
2.2 Metode Kuadrat Terkecil	5
2.3 Uji Asumsi Klasik.....	6
2.3.1 Uji Normalitas	6
2.3.2 Uji Asumsi Multikolinieritas	6
2.4 Pencilan	7
2.5 Regresi <i>Robust</i>	8
2.5.1 <i>Robust Least Trimmed Square</i>	9
2.6 Regresi <i>Ridge</i>	9

2.7	Regresi <i>Ridge Robust</i>	10
2.8	Koefisien Determinasi	11
2.8.1	<i>Mean Square Error</i>	11
2.9	Uji Signifikansi Parameter	11
2.9.1	Pengujian Simultan	12
2.9.2	Pengujian Parsial	12
2.10	Angka Kematian Bayi	13
BAB III METODOLOGI PENELITIAN		15
3.1	Sumber Data	15
3.2	Variabel Penelitian	15
3.3	Metode Analisis	15
BAB IV HASIL DAN PEMBAHASAN.....		17
4.1	Estimasi Parameter Regresi <i>Ridge Robust Least Trimmed Square</i>	17
4.2	Statistik Deskriptif	18
4.3	Uji Normalitas	19
4.4.	Uji Multikolinearitas	20
4.5	Uji DfFITS	20
4.6	Estimasi Parameter Model Regresi <i>Robust</i>	21
4.7	Uji Signifikansi Parameter	23
4.8	Simulasi Pada Data Yang Berdistribusi Normal	24
4.8.1	Uji Normalitas	25
4.8.2	Estimasi Parameter Model Regresi <i>Robust</i>	26
4.8.3	Uji Signifikansi Parameter.....	28
4.8.4	Uji Kebaikan Model.....	31
BAB V KESIMPULAN DAN SARAN		32
5.1	Kesimpulan	32
5.2	Saran.....	32
DAFTAR PUSTAKA		33
LAMPIRAN.....		35

DAFTAR GAMBAR

Gambar 4.1 *Scatter Plot* Data Bangkitan.....26

DAFTAR TABEL

Tabel 4.1 Hasil Deskriptif	19
Tabel 4.2 Uji Normalitas	19
Tabel 4.3 Uji VIF	20
Tabel 4.4 Parameter β dari metode MKT	21
Tabel 4.5 Iterasi Parameter <i>Robust</i>	22
Tabel 4.6 Parameter Regresi <i>Ridge Robust Least Trimmed Square</i>	23
Tabel 4.7 Uji Simultan	23
Tabel 4.8 Uji Parsial Satu-Satu	24
Tabel 4.9 Uji Normalitas	25
Tabel 4.10 Parameter β dari metode MKT	26
Tabel 4.11 Iterasi Parameter <i>Robust</i>	27
Tabel 4.12 Parameter Regresi <i>Ridge Robust Least Trimmed Square</i>	28
Tabel 4.13 Uji Simultan	28
Tabel 4.14 Uji Parsial Satu-Satu	29
Tabel 4.15 Parameter <i>Ridge Robust Least Trimmed Square</i>	30
Tabel 4.16 Uji Kebaikan Model	31

DAFTAR LAMPIRAN

Lampiran 1 Data AKB Di Provinsi Sulawesi Selatan Tahun 2017.....	36
Lampiran 2 Hasil Uji DfFITS	37
Lampiran 3 Iterasi Parameter Regresi <i>Robust Least Trimmed Square</i>	38
Lampiran 4 Iterasi Parameter Regresi <i>Robust LTS</i> Data Bangkitan	39
Lampiran 5 Hasil Uji DfFITS Data Bangkitan	40
Lampiran 6 Variabel X Signifikan Terhadap Y	41

BAB I

PENDAHULUAN

1.1 Latar Belakang

Angka Kematian Bayi (AKB) merupakan indikator yang digunakan untuk melihat status kesehatan anak. Kematian bayi adalah kejadian kematian yang terjadi pada periode sejak bayi lahir sampai bayi belum berusia tepat satu tahun. Kematian bayi dipengaruhi oleh jumlah kematian neonatal. Neonatal adalah pelayanan sesuai standar yang diberikan tenaga kesehatan, sedikitnya tiga kali selama periode 0-28 hari, kunjungan neonatal pertama dilakukan 6 sampai 48 jam setelah bayi lahir. Penyebab kematian neonatal antara lain akibat infeksi (pneumonia, tetanus, diare). Provinsi Sulawesi Selatan adalah salah satu provinsi di Indonesia yang memiliki masalah kematian neonatal. Tingkat kematian bayi di Provinsi Sulawesi Selatan menjadi tolak ukur permasalahan kesehatan masyarakat yang berkaitan langsung dengan faktor-faktor yang berhubungan dalam pelayanan kesehatan bayi di Sulawesi Selatan. Setelah mengetahui faktor-faktor penyebab AKB maka dapat dilakukan pengendalian sehingga dapat ditentukan upaya yang dapat menurunkan AKB di Sulawesi Selatan (Dinkes Sulsel, 2018). Oleh karena itu, perlu dilakukan pemodelan AKB terhadap faktor-faktor yang mempengaruhi AKB tersebut sehingga tingkat kematian bayi dapat dikendalikan.

Analisis regresi merupakan salah satu metode yang dapat digunakan untuk mengetahui faktor-faktor yang mempengaruhi AKB. Dalam analisis regresi, AKB digunakan sebagai respon (Y) dengan variabel prediktor (X) adalah faktor-faktor yang mempengaruhi kematian bayi. Metode kuadrat terkecil (MKT) adalah salah satu teknik estimasi parameter regresi yang umumnya digunakan yaitu untuk mengestimasi parameter model regresi. Bentuk model yang baik harus memenuhi asumsi-asumsi regresi. Adapun asumsi-asumsi regresi yang harus dipenuhi antara lain adalah data berdistribusi normal, homoskedastisitas, tidak terdapat autokorelasi dan tidak terjadi multikolinieritas (Supranto, 2005).

Multikolinieritas merupakan masalah yang timbul berkaitan dengan adanya hubungan linear diantara variabel-variabel prediktor dalam model regresi. Terdapatnya multikolinieritas menyebabkan pemakaian metode penaksir kuadrat

terkecil menjadi kurang tepat, karena taksiran koefisien regresinya tidak stabil dan variansi koefisien regresinya sangat besar. Oleh karena itu, masalah multikolinearitas harus dianggap sebagai suatu kelemahan yang mengurangi keyakinan dalam uji signifikansi konvensional (Sumodiningrat, 2002). Multikolinearitas dapat diatasi dengan menggunakan metode analisis regresi *ridge*. Regresi *ridge* diajukan sebagai suatu cara untuk mengatasi multikolinearitas, regresi *ridge* mengurangi dampak multikolinearitas dengan menentukan penduga yang bias tetapi mempunyai variansi minimum (Pratiwi, 2016).

Masalah lain yang muncul dalam analisis regresi adalah terdapat pencilan pada data. Pencilan adalah salah satu masalah dalam analisis regresi linier. Adanya pencilan dalam data menyebabkan garis regresi berdasarkan MKT tertarik ke arah pencilan dan menyebabkan eror yang dihasilkan menjadi besar sehingga mengganggu kenormalan data, untuk mengatasi masalah tersebut, digunakan metode *robust*. Regresi *robust* adalah salah satu metode regresi yang digunakan untuk mengatasi pencilan. (Cahyawati, 2009).

Pada penelitian sebelumnya, telah dilakukan oleh Ainur (2018), tentang perbandingan metode LTS dan metode S. Hasil dari penelitian tersebut menunjukkan bahwa pada regresi *robust* yang digunakan, metode LTS merupakan metode yang lebih baik dibandingkan metode S. Hal ini dapat dilihat dari nilai *MSE* metode LTS yang lebih kecil dibandingkan metode S. Penelitian sebelumnya yang telah dilakukan oleh Putri (2014) menunjukkan bahwa metode LTS lebih efektif daripada metode MM karena nilai salah baku penduga yang dihasilkan metode LTS lebih kecil dibandingkan metode MM.

Berdasarkan uraian yang telah dibahas sebelumnya, penulis tertarik untuk mengkaji penelitian tentang masalah kesehatan, dalam hal ini kematian bayi. Pada penelitian ini penulis akan menggunakan metode regresi *ridge robust* estimasi *least trimmed square* untuk mengkaji kasus mengenai Angka Kematian Bayi di Provinsi Sulawesi Selatan Tahun 2017 beserta faktor-faktor yang mempengaruhinya. Oleh karena itu, penulis memilih judul untuk penelitian ini yaitu **“Estimasi Parameter Regresi Ridge Robust Pada Data Profil Kesehatan Sulawesi Selatan”**.

1.2 Rumusan Masalah

Rumusan masalah dari penelitian ini adalah:

1. Bagaimana estimasi parameter model regresi *ridge robust* dengan penduga *Least Trimmed Square* pada data profil kesehatan Sulawesi Selatan?
2. Bagaimana mengetahui faktor-faktor yang mempengaruhi AKB di Sulawesi Selatan berdasarkan regresi *ridge robust* dengan penduga *least trimmed square*?

1.3 Batasan Masalah

Untuk membatasi masalah dalam penelitian ini agar sesuai dengan yang dimaksud dan tidak terjadi penyimpangan maka peneliti perlu memberikan batasan masalah yaitu data yang digunakan dalam penelitian ini adalah data profil Kesehatan Provinsi Sulawesi Selatan pada Tahun 2017. Selain itu metode penduga pada regresi *robust* yang digunakan adalah *least trimmed square*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mendapatkan hasil estimasi parameter model regresi *ridge robust* dengan penduga *least trimmed square* pada data profil kesehatan Sulawesi Selatan Tahun 2017.
2. Mendapatkan faktor-faktor yang mempengaruhi AKB di Sulawesi Selatan berdasarkan regresi *ridge robust* dengan penduga *least trimmed square*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan penulis dari hasil penyusunan tugas akhir ini adalah dapat menambah wawasan dalam bidang statistika khususnya regresi linear dengan metode *ridge robust least trimmed square* sehingga dapat digunakan sebagai referensi untuk penulisan selanjutnya, dan dapat menjadi sumber informasi profil kesehatan perkembangan ibu dan bayi di Provinsi Sulawesi Selatan.

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Regresi Linear

Analisis regresi adalah suatu metode yang berguna untuk menentukan hubungan suatu variabel respon dengan satu atau lebih variabel prediktor. Salah satu tujuan analisis regresi adalah menentukan model regresi yang baik, sehingga model dapat digunakan untuk menerangkan dan memprediksi hal-hal yang berhubungan dengan variabel-variabel yang terlibat dalam model regresi. Model regresi merupakan model yang memberikan gambaran mengenai hubungan antara variabel prediktor dan variabel respon (Sembiring, 1995). Bentuk umum model regresi linier adalah sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_p x_{ip} + e_i, \quad i = 1, 2, \dots, n \quad (2.1)$$

dengan:

y_i = variabel respon pada pengamatan ke- i ,

x_{ij} = variabel prediktor ke- j pada pengamatan ke- i , ($j = 1, 2, \dots, p$)

β_j = parameter regresi variabel prediktor ke- j , ($j = 1, 2, \dots, p$)

e_i = eror pengamatan ke- i

Persamaan (2.1) dapat dinyatakan dalam bentuk matriks seperti pada persamaan (2.2) berikut:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (2.2)$$

dengan;

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Menurut Harlan (2018), asumsi-asumsi yang harus dipenuhi pada model regresi linier, yaitu:

1. Asumsi linieritas, menyatakan bahwa hubungan antara variabel respon dan variabel prediktor adalah linier.
2. Asumsi normalitas, menyatakan bahwa suku eror berdistribusi normal dengan mean nol, yaitu $E(e_i) = 0$ dan $\text{var}(e_i) = \sigma^2$

3. Asumsi homoskedastisitas, menyatakan bahwa suku respon memiliki variansi yang konstan, yaitu $Var(e_i) = \sigma^2 \forall i$
4. Asumsi independensi dan non-autokorelasi, menyatakan bahwa suku eror saling bebas dan tak saling berkorelasi, yaitu $cov(e_i, e_{i*}) = 0$.
5. Asumsi non-multikolinieritas, menyatakan bahwa tidak boleh ada korelasi antar variabel prediktor, yaitu $cov(x_j, x_{j*}) = 0$.

2.2 Metode Kuadrat Terkecil

Metode kuadrat terkecil (MKT) merupakan salah satu metode untuk mengestimasi parameter regresi dengan cara meminimumkan jumlah kuadrat eror. Untuk mengestimasi parameter dapat dilakukan dengan pendekatan matriks sebagai berikut (Montgomery, 1992):

$$e = Y - X\beta \quad (2.3)$$

dan jumlah kuadrat eror:

$$\begin{aligned} S &= e'e \\ &= (Y - X\beta)'(Y - X\beta) \\ &= Y'Y - \beta'X'Y - Y'X\beta + \beta'X'X\beta \\ &= Y'Y - 2\beta'X'Y + \beta'X'X\beta \end{aligned} \quad (2.4)$$

dengan adanya sifat transpose matriks yaitu $(X\beta)' = \beta'X'$ maka jika $\beta'X'Y$ dibalik akan menghasilkan nilai yang sama yaitu $Y'X\beta$. Untuk mendapatkan β yang meminimumkan S , maka persamaan (2.4) diturunkan terhadap β dan disama dengarkan nol yaitu:

$$\begin{aligned} \frac{\partial(S)}{\partial\beta} \Big|_{\beta=\hat{\beta}} &= 0 \\ \frac{\partial(Y'Y - 2\beta'X'Y + \beta'X'X\beta)}{\partial\hat{\beta}} &= 0 \\ -2X'Y + 2X'X\hat{\beta} &= 0 \\ 2X'X\hat{\beta} &= 2X'Y \\ X'X\hat{\beta} &= X'Y \end{aligned} \quad (2.5)$$

Selanjutnya Persamaan (2.5) dikalikan dengan $(X'X)^{-1}$ akan menghasilkan:

$$(X'X)^{-1}(X'X)\hat{\beta} = (X'X)^{-1}X'Y \quad (2.6)$$

karena $(X'X)^{-1}(X'X) = I$, I adalah matriks identitas maka diperoleh

$$\hat{\beta} = (X'X)^{-1}X'Y \quad (2.7)$$

2.3 Uji Asumsi Klasik

Uji asumsi klasik adalah persyaratan statistik yang harus dipenuhi pada suatu analisis regresi linier berganda yang menggunakan estimasi MKT. Uji asumsi klasik tersebut antara lain, terpenuhinya distribusi normal, tidak terdapat multikolinieritas, varian konstan, dan homoskedastisitas serta tidak terdapat autokorelasi. Asumsi multikolinieritas adalah asumsi klasik yang dilanggar oleh data.

2.3.1 Uji Normalitas

Menurut (Ghozali,2013) uji normalitas bertujuan untuk menguji apakah dalam model regresi variabel respon dan variabel prediktor keduanya memiliki distribusi normal atau tidak. Model regresi yang baik adalah memiliki distribusi data normal atau mendekati normal. Normalitas data dapat dilihat dari penyebaran (titik) pada sumbu diagonal pada grafik normal P-Plot.

Menurut (Jarque & Bera, 1987) banyak jenis uji statistik normalitas yang dapat digunakan diantaranya *Kolmogorov Smirnov*, *Lilliefors*, *Shapiro Wilk*, dan *Jarque Bera*. Pada penelitian ini akan digunakan uji normalitas dengan *Kolmogorov Smirnov*. Uji *Kolmogorov smirnov* adalah metode statistik yang digunakan untuk menguji hipotesis komparatif dari dua sampel prediktor dengan bentuk data ordinal yang disusun pada tabel distribusi frekuensi kumulatif dengan sistem interval kelas. Hipotesis *Kolmogorov Smirnov*:

H_0 : data berdistribusi normal

H_1 :data tidak berdistribusi normal

Pada uji ini, data berdistribusi tidak normal jika $p < 0.05$ (H_0 ditolak) dan data berdistribusi normal jika $p > 0.05$ (H_0 diterima).

2.3.2 Uji Asumsi Multikolinieritas

Multikolinieritas adalah terjadinya hubungan linier antara variabel prediktor dalam suatu model regresi linier berganda. Hubungan linier antara variabel prediktor dapat terjadi dalam bentuk hubungan linier yang sempurna (*perfect*) dan hubungan linier yang kurang sempurna (*imperfect*). Adapun

dampak adanya multikolinieritas dalam model regresi linier berganda adalah (Gujarati, 2003 dan Widarjono, 2007):

1. Penaksir MKT masih bersifat BLUE (*Best Linear Unbiased Estimator*), tetapi mempunyai variansi dan kovariansi yang besar sehingga sulit mendapatkan taksiran (estimasi) yang tepat.
2. Akibat penaksir MKT mempunyai variansi dan kovariansi yang yang besar, menyebabkan interval estimasi akan cenderung lebih lebar dan nilai hitung statistik uji T akan kecil, sehingga membuat variabel prediktor secara statistik tidak signifikan mempengaruhi variabel respon.
3. Walaupun hasil uji parsial satu-satu secara individu variabel prediktor tidak berpengaruh terhadap variabel respon melalui uji T , tetapi nilai koefisien determinasi (R^2) masih bias relatif tinggi.

Menurut Montgomery, salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinieritas pada regresi linier berganda adalah *Variance Inflation Factors* (VIF). Adanya multikolinieritas dinilai dari nilai VIF yang dihasilkan. *Variance Inflation Factors* (VIF) dapat dihitung dengan menggunakan rumus sebagai berikut:

$$VIF = \frac{1}{(1-R_j^2)} \quad (2.8)$$

dengan : R_j^2 merupakan koefisien determinasi ke- j , $j = 1, 2, \dots, p$.

Besarnya nilai VIF ini bergantung pada nilai koefisien determinasi (R_j^2) yang dihasilkan. Nilai VIF juga dapat membantu kita untuk mengidentifikasi variabel - variabel prediktor mana saja yang mengalami masalah multikolinieritas. Jika nilai VIF > 10 maka menunjukkan multikolinieritas. (Gujarati, 2003).

2.4 Pencilan

Menurut R.K Sembiring (1950), pencilan adalah pengamatan yang jauh dari pusat data yang mungkin berpengaruh besar terhadap koefisien regresi. Umumnya pengamatan yang dicurigai sebagai pencilan dikategorikan ke dalam pelanggaran asumsi.

Pencilan berpengaruh terhadap proses analisis data, misalnya terhadap nilai rata-rata dan standar deviasi. Oleh karena itu, keberadaan pencilan dalam

suatu data harus dihindari karena pencilan dapat menyebabkan kesalahan dalam pengambilan keputusan dan kesimpulan. Dampak keberadaan pencilan ini dapat mengganggu proses analisa data. Dalam kaitannya dengan analisis regresi, pencilan dapat menyebabkan hal-hal berikut:

1. Error yang besar dari model yang terbentuk.
2. Variansi pada data menjadi lebih besar.
3. Estimasi interval memiliki rentang yang lebar.

Salah satu metode untuk mengetahui adanya pencilan dalam suatu analisis adalah Metode DfFITS (*Difference in Fit Standardized*). Rumus DfFITS didefinisikan sebagai berikut (Soemartini, 2007):

$$(DfFITS) = \frac{\hat{Y}_i - \hat{Y}_{i-1}}{s_{i-1}^2 - \sqrt{h_{ii}}} \quad (2.9)$$

dengan:

\hat{Y}_i : nilai taksiran Y_i

\hat{Y}_{i-1} : nilai taksiran Y_i tanpa pengamatan ke- i

S_{i-1}^2 : jumlah kuadrat error tanpa pengamatan ke- i

h_{ii} : elemen diagonal ke- i

DfFITS merupakan suatu ukuran berpengaruh yang ditimbulkan oleh pengamatan ke- i terhadap nilai taksiran \hat{y}_i . Pendeteksian dengan DfFITS dengan melalui nilai \hat{y}_i dan gabungan nilai *leverage* (h_{ii}) dengan standar error (Soemartini, 2007). Untuk data berukuran kecil sampai sedang, jika nilai $DfFITS > 1$ maka data tersebut adalah data pencilan. Untuk data berukuran besar $DfFITS > 2\sqrt{\frac{p}{n}}$ dengan p adalah jumlah variabel prediktor dan n adalah banyaknya observasi maka data tersebut adalah data pencilan (Soemartini, 2007).

2.5 Regresi Robust

Regresi *robust* adalah salah satu metode regresi yang digunakan ketika terdapat beberapa pencilan pada data yang mempengaruhi kenormalan data. Metode ini digunakan untuk menganalisis data yang mengandung pencilan agar model yang dihasilkan stabil terhadap pencilan. Terdapat beberapa estimasi dalam regresi *robust* diantaranya estimasi-M, estimasi-S, estimasi-LMS, estimasi-LTS dan estimasi-MM. (Chen, 2002).

2.5.1 Robust Least Trimmed Square

Metode *Least Trimmed Squares* (LTS) pertama kali diperkenalkan oleh Rousseeuw pada tahun 1984 sebagai metode alternatif untuk mengatasi kelemahan MKT. Rousseeuw dan Hubert menjelaskan bahwa metode LTS mempunyai prinsip yang sama dengan MKT dalam mengestimasi parameter regresi yaitu meminimumkan jumlah kuadrat error. Namun metode LTS tidak menggunakan seluruh pengamatan dalam perhitungannya, melainkan hanya meminimumkan jumlah kuadrat error dari himpunan bagian data berukuran pengamatan dengan kuadrat terkecil. Fungsi tujuan dari metode LTS dapat dituliskan sebagai berikut (Rosseeuw, 1987):

$$\min \sum_{i=1}^h e_{(i)}^2 = \min \sum_{i=1}^h (Y_{(i)} - \hat{Y}_{(i)})^2 \quad (2.10)$$

Persamaan metode ini dapat dilihat sebagai berikut:

$$\hat{\beta}_{LTS} = \min_{\beta} \sum_i^h e_i^2 \quad (2.11)$$

dengan $h = \frac{n+p+2}{2}$, $\left(\frac{n}{2} < h < n\right)$

e_i^2 : kuadrat error dari pengamatan ke- i .

2.6 Regresi Ridge

Regresi *ridge* adalah salah satu metode yang diusulkan untuk mengatasi masalah multikolinieritas. Berdasarkan fungsi regresi dilakukan pendugaan parameter dengan menggunakan MKT. Apabila terdapat multikolinieritas pada variabel prediktor maka akan menyebabkan nilai $(X'X)^{-1}$ error. Oleh karena itu digunakan regresi *ridge* untuk mengatasi masalah multikolinieritas dengan menambahkan faktor koreksi yaitu konstanta (c) (Kutner M. H., 2004).

Penduga regresi *ridge* diperoleh dari penduga parameter MKT ditambahkan dengan faktor koreksi, yaitu konstanta (c) yang dikalikan dengan matriks identitas sehingga $(X'X)^{-1}$ menjadi $(X'X + cI)^{-1}$. Penambahan faktor koreksi tersebut bertujuan agar determinan matriks $X'X$ bisa dihitung. Persamaan regresi *ridge* dapat ditulis sebagai berikut:

$$\hat{\beta}_{Ridge} = (X'X + cI)^{-1} X'Y \quad (2.12)$$

Nilai c didapat dari rumus sebagai berikut (Montgomery dan Peck, 1992) :

$$c = \frac{(p+1)\hat{\sigma}_{LTS}^2}{\hat{\beta}'_{LTS}\hat{\beta}_{LTS}} \quad (2.13)$$

dengan:

p = banyak variabel prediktor

$\hat{\sigma}_{LTS}^2$ = kuadrat tengah galat dari penduga LTS

$\hat{\beta}_{LTS}$ = penduga parameter LTS

2.7 Regresi *Ridge Robust*

Regresi *ridge* adalah suatu teknik yang dikembangkan untuk menstabilkan nilai koefisien regresi karena adanya multikolinieritas tetapi jika terdapat pencilan pada data yang mengandung multikolinieritas dan pengamatannya berpengaruh besar, maka regresi *ridge* yang biasa tidak dapat digunakan untuk mengatasi kedua masalah tersebut. Oleh karena itu, metode regresi *ridge* tidak dapat menangani masalah pencilan dan multikolinieritas secara bersamaan maka diberikan rumus untuk menangani masalah tersebut dengan mengkombinasikan regresi *ridge* dan *robust* dengan cara menggabungkan kedua sifat. Penggabungan kedua metode tersebut disebut regresi *ridge robust*. Penduga regresi *ridge robust* yang dihasilkan akan stabil terhadap pencilan (Samkar, 2010).

Rumus hasil penggabungan regresi *ridge* dan regresi *robust* menurut Samkar (2010) adalah sebagai berikut:

$$\hat{\beta}_{RR} = (X'X + cI)^{-1}X'X\hat{\beta}_{LTS} \quad (2.14)$$

dengan:

$\hat{\beta}_{RR}$ = penduga parameter regresi *ridge robust*

$c = \frac{(p+1)\hat{\sigma}_{LTS}^2}{\hat{\beta}'_{LTS}\hat{\beta}_{LTS}}$

$\hat{\sigma}_{LTS}^2 = \frac{(Y-X\hat{\beta}_{LTS})'(Y-X\hat{\beta}_{LTS})}{n-(p+1)}$

p = banyak variabel prediktor

$\hat{\beta}_{LTS}$ = penduga parameter dari penduga regresi *robust-LTS*

2.8 Koefisien Determinasi

Koefisien determinasi adalah salah satu ukuran tentang kecocokan data dengan model. Rumus koefisien determinasi adalah sebagai berikut:

$$R^2 = 1 - \frac{JKE}{JKT} = 1 - \frac{\sum_{i=1}^n (\hat{y} - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y})^2} \quad (2.15)$$

dengan:

R^2 = koefisien determinasi

JKE = jumlah kuadrat error

JKT = jumlah kuadrat total

Semakin dekat nilai R^2 dengan 1 maka semakin baik kecocokan data dengan model, dan sebaliknya.

2.8.1 Mean Square Error

Mean Square Error (MSE) adalah kesalahan rata-rata kuadrat diantara nilai aktual dan nilai peramalan. MSE secara umum digunakan untuk mengukur ketepatan nilai dugaan model yang dinyatakan dalam rata-rata kuadrat dari kesalahan dan juga dapat digunakan untuk membandingkan ketetapan ramalan antara metode peramalan yang berbeda. MSE memberikan ketelitian yang lebih baik dari pada *Mean Absolut Deviation* (MAD) sehingga MSE lebih banyak dipakai untuk mengevaluasi metode pengukuran dengan model regresi. Secara matematis dapat ditulis pada 2.16 (Montgomery, 1992):

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n-p-1} \quad (2.16)$$

dengan:

n = jumlah data

Y_i = nilai aktual

\hat{Y}_i = nilai hasil peramalan

2.9 Uji Signifikansi Parameter

Uji signifikansi parameter digunakan untuk mengetahui pengaruh dari variabel-variabel prediktor terhadap variabel respon baik secara simultan maupun secara parsial.

2.9.1 Pengujian Simultan

Pengujian simultan parameter dimaksudkan untuk mengetahui pengaruh secara bersama-sama dari variabel prediktor terhadap variabel respon. Pengujian hipotesis berikut (Gujarati, 2003).

Hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_j = \dots = \beta_p = 0,$$

$$H_1 : \text{minimal ada satu } \beta_j \neq 0, j = 1, 2, \dots, p$$

Maka statistik uji F yaitu:

$$F = \frac{MSR}{MSE} = \frac{\sum_{i=1}^n (\hat{y} - \bar{y})^2 / k}{\sum_{i=1}^n (y_i - \hat{y})^2 / n - p - 1} \quad (2.17)$$

dengan:

MSR = Mean Square Regression

MSE = Mean Square Error

R^2 = koefisien determinasi

n = jumlah data

p = banyak variabel

Kriteria pengambilan keputusan:

Jika $F_{hitung} > F_{tabel}$, maka H_0 ditolak

Jika $F_{hitung} \leq F_{tabel}$, maka H_0 diterima

2.9.2 Pengujian Parsial

Uji parsial digunakan untuk melakukan uji hipotesis pada setiap parameter. Adapun hipotesis dalam uji ini sebagai berikut:

$H_0: \beta_j = 0$, (tidak ada hubungan yang signifikan) , $j = 1, 2, \dots, p$.

$H_1: \beta_j \neq 0$, (paling tidak terdapat satu hubungan yang signifikan)

Dengan statistik uji yang digunakan adalah

$$t_{hit(j)} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.18)$$

dengan:

$t_{hit(j)}$ = nilai statistik t pada variabel prediktor ke- j

$\hat{\beta}_j$ = taksiran untuk β_j

$SE(\hat{\beta}_j)$ = simpangan baku dari β_j

Pengambilan keputusan didasarkan dengan perbandingan t_{hitung} dengan t_{tabel} . (Mukhlisin, 2019). Jika t_{hitung} lebih besar dari t_{tabel} , maka H_0 ditolak dan jika t_{hitung} lebih kecil dari t_{tabel} , maka H_0 diterima. t_{tabel} didapatkan dari tabel t , dengan tingkat signifikansi α dan $dk = n - 2$ (n adalah jumlah observasi).

2.10 Angka Kematian Bayi

Angka Kematian Bayi (AKB) merupakan salah satu indikator penting dalam menentukan tingkat kesehatan masyarakat karena dapat menggambarkan kesehatan penduduk secara umum (Dinkes Sulsel, 2013). Angka ini sangat sensitif terhadap perubahan tingkat kesehatan dan kesejahteraan. AKB tersebut dapat didefinisikan sebagai kematian yang terjadi saat setelah bayi lahir sampai bayi belum belum berusia tepat satu tahun, per 1000 kelahiran hidup pada satu tahun tertentu (BPS, 2013). Secara garis besar penyebab dari kematian bayi ada dua macam yaitu endogen dan eksogen. Kematian bayi endogen atau disebut juga dengan kematian neonatal adalah kematian bayi yang terjadi pada bulan pertama setelah dilahirkan, dan umumnya disebabkan oleh faktor-faktor yang dibawa anak sejak lahir yang diperoleh dari orangtuanya pada pada saat konsepsi atau didapat selama kehamilan. Sedangkan kematian bayi eksogen atau kematian post neonatal adalah kematian bayi yang terjadi setelah usia satu bulan sampai menjelang usia satu tahun yang disebabkan oleh faktor-faktor yang berkaitan dengan pengaruh lingkungan luar. AKB dapat dihitung dengan menggunakan rumus sebagai berikut:

$$AKB = \frac{D_{0-<1th}}{\text{lahir hidup}} \times 1000 \quad (2.19)$$

dengan:

AKB : Angka Kematian Bayi

$D_{0-<1th}$: Jumlah kematian bayi berumur kurang dari 1 tahun pada daerah tertentu

lahir hidup: Jumlah kelahiran bayi hidup selama 1 tahun pada daerah tertentu

Ada beberapa faktor yang mempengaruhi AKB di Sulawesi Selatan antara lain tersedianya berbagai fasilitas dan pelayanan kesehatan, serta tersedianya tenaga kesehatan, norma kehidupan masyarakat, dibidang kesehatan, gizi, dan pendapatan masyarakat (Dinkes, 2009). Oleh karena itu, pada penelitian mengenai AKB di Provinsi Sulawesi Selatan ini menggunakan enam variabel yang diduga berpengaruh diantaranya adalah kepadatan penduduk, persentase rumah tangga berperilaku hidup bersih dan sehat, jumlah bayi berat badan lahir rendah, jumlah bayi yang diberi asi eksklusif, jumlah bayi yang mendapat kunjungan neonatal 1 kali, jumlah bayi yang mendapat pelayan kesehatan.