PEMODELAN *LINEARIZED RIDGE REGRESSION* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS

SKRIPSI

MUKRIMIN ADAM

H051171304

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

PEMODELAN *LINEARIZED RIDGE REGRESSION* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS

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

MUKRIMIN ADAM

H051171304

PROGRAM STUDI STATISTIKA DEPARTEMEN STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

2021

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama	: Mukrimin Adam
NIM	: H051171304
Program Studi	: Statistika
Jenjang	: Sarjana (S1)

Menyatakan dengan ini bahwa karya tulis saya yang berjudul

PEMODELAN *LINEARIZED RIDGE REGRESSION* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS

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

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, 6 September 2021

<u>MUKRIMIN ADAM</u>

NIM. H051171304

PEMODELAN *LINEARIZED RIDGE REGRESSION* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS

Disetujui Oleh:

Pembimbing Utama,	**Pembimbing Pertama,**
<u>Sitti Sahriman, S.Si., M.Si.</u>	**<u>Dra. Nasrah Sirajang, M.Si.</u>**
NIP. 19881018 201504 2002	**NIP. 19650519 199303 2002**

Ketua Departemen Statistika

<u>Dr. Nurtiti Sunusi, S.Si., M.Si.</u>

NIP. 19720117 199703 2002

Pada Tanggal: 6 September 2021

LEMBAR PENGESAHAN

PEMODELAN *LINEARIZED RIDGE REGRESSION* PADA DATA YANG MENGANDUNG MULTIKOLINEARITAS

Disusun dan diajukan oleh

MUKRIMIN ADAM

H051171304

Telah dipertahankan dihadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin

pada tanggal 6 September 2021

dan dinyatakan telah memenuhi syarat kelulusan.

Menyetujui,

Pembimbing Utama,	Pembimbing Pertama,
Sitti Sahriman, S.Si., M.Si.	**Dra. Nasrah Sirajang, M.Si.**
NIP. 19881018 201504 2002	**NIP. 19650519 199303 2002**

Ketua Departemen Statistika

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN UNIVERSITAS HASANUDDIN FAKULTAS MIPA

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

NIP. 19720117 199703 2002

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, atas berkat dan karunia-Nya sehingga penulsi dapat menyelesaikan tugas akhir ini yang berjudul "Pemodelan *Linearized Ridge Regression* Pada Data Yang Mengandung Multikolinearitas" 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 **Baharudin** dan Ibunda **Husni** atas pengorbanan, kesabaran, dukungan, semangat, dan doanya yang menjadi kekuatan bagi penulis sehingga penulis mampu berada dititik saat ini. 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 yang telah seperti orang tua sendiri. Segenap dosen pengajar dan staf **Departemen Statistika** yang telah membekali ilmu dan kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Statistika.
4. **Ibu Sitti Sahriman, S.Si., M.Si.** selaku Pembimbing Utama penulis yang telah ikhlas meluangkan waktu dan pemikirannya untuk memberikan arahan, pengetahuan, motivasi dan bimbingan ditengah kesibukan beliau serta menjadi tempat berkeluh kesah untuk penulis.

5. **Ibu Dra. Nasrah Sirajang, M.Si.** selaku Pembimbing Pertama sekaligus penasehat akademik penulis yang telah meluangkan waktunya ditengah kesibukan untuk memberikan arahan bagi penulis.
6. **Bapak Drs. Raupong, M.Si.** selaku tim penguji yang telah memberikan saran dan kritikan yang membangun dalam penyempurnaan penyusunan tugas akhir.
7. **Alm. Dr. La Podje Talangko, M.Si.** selaku dosen pemimbimbing akademik yang telah banyak memberikan saran atau masukan selama perkuliahan. Semoga beliau diberikan tempat yang terbaik disisi-Nya
8. Teman-teman **Statistika 2017**, terima kasih atas kebersamaan, suka, dan duka selama menjalani pendidikan di Departemen Statistika.
9. Keluarga besar **DISKRIT 2017**, terima kasih telah memberikan pelajaran yang berharga dan arti kebersamaan selama ini kepada penulis. Pengalaman yang berharga telah penulis dapatkan dari teman-teman selama berproses.
10. **Keluarga Mahasiswa FMIPA Unhas** terkhusus anggota keluarga **Himatika FMIPA Unhas** dan **Himastat FMIPA Unhas**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan di proses perkuliahan dan telah menjadi keluarga selama penulis kuliah di Universitas Hasanuddin.
11. Kepada seluruh pihak yang tidak dapat penulis sebutkan satu persatu, terima kasih setinggi-tingginya untuk segala dukungan dan partisipasi yang diberikan kepada penulis semoga bernilai ibadah di sisi Allah *Subhanahu Wa Ta'ala.*

Penulis berharap skripsi ini dapat memberikan tambahan pengetahuan baru bagi para pembelajar statistika. Penulis menyadari bahwa dalam penulisan tugas akhir ini masih banyak terdapat kekurangan. Oleh karena itu, dengan segala kerendahan hati penulis memohon maaf. Akhir kata, semoga dapat bermanfaat bagi pihak-pihak yang berkepentingan. *Aamiin Yaa Rabbal Alamin.*

Makassar, 6 September 2021

Mukrimin Adam

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Mukrimin Adam
NIM	: H051171304
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 Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

"Pemodelan *Linearized Ridge Regression* Pada Data yang Mengandung Multikolinearitas"

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 memublikasikan 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 6 September 2021

Yang menyatakan

Mukrimin Adam

ABSTRAK

Salah satu asumsi yang harus dipenuhi dalam model regresi linier berganda adalah tidak terjadinya masalah multikolinearitas diantara variabel bebas. Namun jika terjadi masalah multikolinearitas, maka estimasi parameter dapat dilakukan dengan menggunakan metode *linearized ridge regression* (LRR). Metode LRR memiliki kelebihan pemilihan konstanta optimal yang mudah ditentukan dan juga memiliki nilai PRESS yang minimum. Pada penelitian ini, angka kematian bayi di Provinsi Sulawesi Selatan akan dimodelkan dengan menggunakan metode LRR berdasarkan variabel jumlah pemberian vitamin A, jumlah pelayanan kesehatan, jumlah bayi lahir dengan berat badan rendah, jumlah ibu bersalin yang ditolong tenaga medis, dan jumlah bayi yang diberi ASI eksklusif. Salah satu ukuran untuk melihat kebaikan model regresi adalah *Prediction Error Sum of Squares* (PRESS). Berdasarkan uji t pada tingkat signifikansi 5% diperoleh jumlah cakupan pemberian vitamin A dan jumlah bayi lahir dengan berat badan rendah memberikan pengaruh yang signfikan terhadap angka kematian bayi dengan nilai PRESS sebesar 0.6846.

Kata Kunci: Angka Kematian Bayi, Multikolinearitas, *Linearized Ridge Regression*, *Prediction Error Sum of Squares.*

.

ABSTRACT

One of the assumptions that must be met in the multiple linear regression model is that there is no multicollinearity problem among the independent variables. However, if there is a multicollinearity problem, then parameter estimation can be done using the linearized ridge regression (LRR) method. The LRR method has the advantage of choosing an optimal constant that is easy to determine and also has a minimum PRESS value. In this study, the infant mortality rate in South Sulawesi Province will be modeled using the LRR method based on the variables of the amount of vitamin A given, the number of health services, the number of babies born with low weight, the number of mothers who give birth assisted by medical personnel, and the number of babies who are breastfed. exclusive. One measure to see the goodness of the regression model is the Prediction Error Sum of Squares (PRESS). Based on the t-test at a significance level of 5%, the total coverage of vitamin A administration and the number of babies born with low weight gave a significant effect on infant mortality with a PRESS value of 0.6846.

Keywords: Infant Mortality Rate, Multicollinearity, Linearized Ridge Regression, Prediction Error Sum of Squares.

DAFTAR ISI

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 suatu negara dan bahkan untuk mengukur tingkat kemajuan suatu bangsa adalah angka kematian bayi (AKB). Masih tingginya AKB menjadi salah satu masalah kesehatan utama di Indonesia. Berdasarkan hasil Survei Demografi dan Kesehatan Indonesia (SDK1) tahun 2019 yang menunjukkan bahwa AKB sebesar 6.02 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.

Angka kematian bayi dapat dipengaruhi oleh beberapa variabel, antara lain pemberian vitamin A yang tidak terpenuhi, tidak rutin memeriksakan kesehatan bayi, bayi lahir dengan berat badan yang rendah, ibu bersalin yang tidak ditolong tenaga kesehatan dan pemberian ASI eksklusif pada bayi yang tidak terpenuhi. Salah satu metode analisis yang dapat digunakan untuk mengetauhi pengaruh variabel tersebut terhadap AKB adalah analisis regresi. Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antara variabel terikat dan variabel bebas. Model yang menyatakan hubungan antara variabel terikat dan variabel bebas disebut model regresi. Model regresi yang biasa digunakan adalah model regresi linier sederhana dan model regresi linier berganda. Pada model regresi tersebut terdapat parameter regresi yang nilainya perlu diestimasi. Salah satu metode estimasi parameter yang sering digunakan dalam model regresi adalah metode kuadrat terkecil. Menurut Gauss-Markov, estimasi parameter yang dihasilkan oleh metode kuadrat terkecil memiliki sifat yang tidak bias dengan variansi yang minimum jika asumsi erornya terpenuhi. Metode kuadrat terkecil akan dapat mengestimasi parameter dengan baik apabila antara variabel bebasnya tidak terjadi masalah multikolinearitas.

Multikolinearitas adalah suatu keadaan yang mana variabel-variabel bebas dalam model regresi linier berganda saling bergantung linier. Ketika variabel bebas saling bergantung linier, estimasi parameter dengan menggunakan metode kuadrat terkecil menjadi kurang tepat karena akan menghasilkan hasil estimasi dengan variansi yang besar sehingga estimasi dengan metode kuadrat terkecil menjadi tidak stabil. Hasil estimasi yang tidak stabil ini memungkinkan nilai estimasi metode kuadrat terkecil jauh dari parameter yang sesungguhnya, sehingga dapat menghasilkan model regresi yang kurang tepat. Beberapa metode diusulkan untuk mengatasi masalah multikolinearitas yang sering terjadi pada data. Rosyadi (2018) menggunakan regresi *ridge* untuk mengatasi multikolinearitas pada kasus indeks pembangunan manusia di Jawa Tengah. Pada penelitian tersebut diperoleh metode *ridge* lebih baik dari pada metode kuadrat terkecil berdasarkan nilai *Mean Squared Error* (MSE) yang dihasilkan. Walidya (2013) menggunakan estimator *generalized ridge* sebagai estimasi parameter pada model regresi linier berganda untuk kasus multikolinearitas. Pada penelitian tersebut diperoleh metode *generalized ridge* lebih baik dari pada metode kuadrat terkecil berdasarkan nilai MSE yang dihasilkan. Pitrianingsih (2013) menggunakan estimator Liu sebagai estimasi parameter model regresi linier berganda pada kasus multikolinearitas. Pada penelitian tersebut diperoleh metode Liu lebih baik dari pada metode kuadrat terkecil berdasarkan nilai MSE yang dihasilkan.

Pada tahun 2011, Liu dan Gao memperkenalkan metode *Linearized Ridge Regression* (LRR) yang merupakan perbaikan dari metode-metode sebelumnya yang dapat mengatasi multikolinearitas. Dalam membentuk estimator LRR, Liu dan Gao menggabungkan estimator *generalized ridge* dengan estimator liu sehingga menghasilkan estimator yang memiliki kelebihan yaitu pemilihan konstanta optimal yang mudah ditentukan dan juga nilai *Prediction Error Sum of Squares* (PRESS) yang minimum. Nilai PRESS merupakan salah satu ukuran yang dapat digunakan untuk melihat kebaikan dari model regresi. Nilai PRESS yang besar dapat mengidentifikasikan bahwa model tidak merepresentasikan data sehingga model akan menghasilkan estimasi yang buruk. Konstanta optimal pada estimator LRR diperoleh dengan meminimumkan PRESS pada estimator LRR, sehingga dengan menggunakan konstanta yang optimal akan diperoleh nilai PRESS yang minimum dan nilai estimasi parameter LRR yang lebih optimal.

Dalam penelitian ini, angka kematian bayi akan dimodelkan dengan menggunakan metode LRR. Sebelumnya, Apriani (2014) juga telah melakukan penelitian dengan menggunakan metode LRR pada data hasil percobaan pengukuran panas yang terlibat selama pengerasan semen selama 180 hari. Pada penelitian tersebut diperoleh nilai PRESS dari estimator LRR yang minimum dibandingkan dengan estimator kuadrat terkecil, Liu dan *generalized ridge*. Berdasarkan uraian tersebut, penulis akan melakukan penelitian dengan judul "***Pemodelan Linearized Ridge Regression Pada Data yang Mengandung Multikolinearitas***".

1.2 Rumusan Masalah

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

1. Bagaimana estimasi parameter model regresi linier berganda dengan menggunakan metode *Linearized Ridge Regression*?
2. Bagaimana dugaan model regresi linier berganda dengan menggunakan metode *Linearized Ridge Regression* pada data angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019 ?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah yang telah diuraikan, maka tujuan dari penelitian ini adalah:

1. Memperoleh hasil estimasi parameter model regresi linier berganda dengan menggunakan metode *Linearized Ridge Regression*.
2. Memperoleh dugaan model regresi linier berganda dengan menggunakan metode *Linearized Ridge Regression* pada data angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019.

1.4 Batasan Masalah

Dalam penelitian ini, penulis memberikan batasan masalah sebagai berikut:

1. Data yang digunakan adalah data angka kematian bayi di Provinsi Sulawesi Selatan tahun 2019 yang terdiri dari 24 kabupaten/kota.
2. Penentuan konstanta bias d_j yang optimal pada estimator *Linearized Ridge Regression* didasarkan pada nilai PRESS yang minimum.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan wawasan keilmuan yang lebih khusus kepada penulis tentang pemodelan *Linearized Ridge Regression* untuk data yang mengalami masalah multikolinearitas.
2. Memberikan informasi kepada pemerintah untuk menetapkan kebijakan dalam rangka mengatasi masalah kematian bayi di Provinsi Sulawesi Selatan.

BAB II

TINJAUAN PUSTAKA

2.1 Matriks

Matriks adalah susunan segi empat siku-siku dari bilangan-bilangan (Anton, 2000). Bilangan-bilangan dalam susunan matriks dinamakan entri. Matriks terdiri dari entri-entri yang disusun menurut baris dan kolom sehingga berbentuk persegi panjang dengan panjang dan lebar menunjukkan banyak baris dan banyak kolom. Matriks yang memiliki n baris dan p kolom disebut matriks berukuran $n \times p$. Bentuk umum dari matriks adalah sebagai berikut:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} \text{ atau } \left[a_{ij}\right]_{n\times p}$$

Entri yang menempati baris ke-i dan kolom ke-j disebut entri (i, j) dan ditulis sebagai $[A]_{ij} = [a_{ij}]$ dengan a_{ij} sebagai skalar. Matriks yang terdiri dari 1 baris dan n kolom ditulis $1 \times n$ disebut dengan matriks baris dan yang terdiri atas n baris dan 1 kolom disebut matriks kolom.

Misalkan $\boldsymbol{A}$ adalah matriks $p \times p$, maka vektor $\boldsymbol{v}$ disebut vektor eigen dari $\boldsymbol{A}$ jika $\boldsymbol{Av}$ adalah kelipatan skalar dari $\boldsymbol{v}$, yaitu $\boldsymbol{Av} = \lambda\boldsymbol{v}$ untuk suatu skalar λ. Skalar λ dinamakan nilai eigen dari $\boldsymbol{A}$. Untuk memperoleh nilai eigen dari sebuah matriks $\boldsymbol{A}$ berukuran $p \times p$, persamaan $\boldsymbol{Av} = \lambda\boldsymbol{v}$ dapat dituliskan kembali menjadi,

$$\boldsymbol{Av} = \lambda\boldsymbol{Iv}$$

$$\boldsymbol{Av} - \lambda\boldsymbol{Iv} = \boldsymbol{0}$$

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$$

Agar λ dapat menjadi nilai eigen, harus terdapat satu solusi tak nol dari persamaan ini. Persamaan ini memiliki solusi tak nol jika dan hanya jika

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0$$

2.2 Dekomposisi Spektral

Jika $\boldsymbol{A}$ matriks simetris, $\boldsymbol{Q}$ matriks yang kolom-kolomnya dibentuk oleh vektor eigen dari $\boldsymbol{A}$ dan $\boldsymbol{\Lambda}$ adalah matriks diagonal yang mana entri ke-j pada diagonal utamanya adalah nilai eigen dari $\boldsymbol{A}$ yang bersesuaian dengan vektor eigen kolom ke-j. Matriks $\boldsymbol{Q}$ dan $\boldsymbol{\Lambda}$ didefenisikan dengan (Rencher, 2002):

$$\boldsymbol{Q} = (\boldsymbol{q_1}, \boldsymbol{q_2}, \dots, \boldsymbol{q_p})$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & \lambda_p \end{pmatrix}$$

dengan $\boldsymbol{q_1}, \boldsymbol{q_2}, \dots, \boldsymbol{q_p}$ adalah vektor eigen dari $\boldsymbol{A}$. Dekomposisi spektral dari matriks $\boldsymbol{A}$ dinyatakan dengan:

$$\boldsymbol{A} = \boldsymbol{Q\Lambda Q'}$$

dengan $\boldsymbol{Q}$ adalah matriks ortogonal.

2.3 Produk Hadamard

Misalkan $\boldsymbol{A}$ dan $\boldsymbol{B}$ merupakan matriks berukuran $n \times p$ yang membentuk matriks $\boldsymbol{C}$, maka produk hadamard dari $\boldsymbol{A}$ dan $\boldsymbol{B}$ didefenisikan dengan:

$$(\boldsymbol{A} \circ \boldsymbol{B})_{ij} = (\boldsymbol{A})_{ij}(\boldsymbol{B})_{ij} = (\boldsymbol{C})_{ij}$$

Untuk setiap $1 \leq i \leq n, 1 \leq j \leq p$, dan $\boldsymbol{C}$ merupakan matriks berukuran $n \times p$.

Teorema 2.1

Misalkan $\boldsymbol{A}$ dan $\boldsymbol{B}$ merupakan matriks berukuran $n \times p$ yang membentuk matriks $\boldsymbol{C}$, maka $(\boldsymbol{A} \circ \boldsymbol{B}) = (\boldsymbol{B} \circ \boldsymbol{A})$.

2.4 Regresi Linier Berganda

Regresi linier berganda merupakan model persamaan yang menjelaskan hubungan satu variabel terikat (y) dengan dua atau lebih variabel bebas (X). Tujuan dari regresi linier berganda adalah untuk memprediksi nilai variabel terikat apabila nilai-nilai variabel bebasnya diketahui. Selain itu, juga untuk mengetahui bagaimana arah hubungan antara variabel terikat dengan variabel-variabel bebasnya.

Model regresi linier berganda dari variabel terikat (y) dengan variabel bebas (X) dapat ditulis sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (2.1)$$

dengan,

y_i = Variabel terikat untuk pengamatan ke-i ($i = 1,2,\dots,n$)

x_{ij} = Variabel bebas untuk pengamatan ke-i variabel ke-j ($j = 0,1,2,\dots,p$)

β_j = Parameter regresi ke-j

ε_i = Sisaan pengamatan ke-i

Persamaan (2.1) dapat diubah ke dalam bentuk notasi matriks, sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2.2)$$

dengan,

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$\boldsymbol{y}$ = Vektor pada variabel terikat berukuran ($n \times 1$)

$\boldsymbol{X}$ = Matriks dari variabel bebas berukuran ($n \times (p+1)$)

$\boldsymbol{\beta}$ = Vektor parameter regresi yang berukuran ($(p+1) \times 1$)

$\boldsymbol{\varepsilon}$ = Vektor sisaan pengamatan yang berukuran ($n \times 1$)

2.5 Multikolinearitas

Multikolinearitas adalah suatu keadaan yang mana variabel-variabel bebas dalam model regresi linier berganda saling bergantung linier. Salah satu ukuran yang dapat digunakan untuk menguji adanya multikolinearitas pada regresi linier berganda adalah *Variance Inflation Factors* (VIF). Adanya multikolinearitas dinilai dari nilai VIF yang dihasilkan. Nilai VIF dapat dihitung dengan menggunakan rumus sebagai berikut:

$$VIF_j = \frac{1}{1 - R_j^2} \qquad (2.3)$$

Pada umumnya, nilai $\mathrm{VIF_j}$ yang sering digunakan adalah lebih besar dari 10 ($\mathrm{VIF_j} > 10$). Karena nilai $\mathrm{VIF_j} > 10$, maka

$$\mathrm{VIF_j} > 10$$

$$\frac{1}{1 - \mathrm{R_j^2}} > 10$$

$$\frac{1}{10} > 1 - \mathrm{R_j^2}$$

$$0{,}1 - 1 > -\mathrm{R_j^2}$$

$$\mathrm{R_j^2} > 0{,}9$$

dengan nilai $\mathrm{VIF_j} > 10$, maka koefisien determinasi $\mathrm{R_j^2} > 0{,}9$ menyatakan bahwa variabel bebas $\boldsymbol{x}_j$ dijelaskan sangat baik oleh variabel bebas lainnya, yang artinya terjadi multikolinearitas pada model regresi linier berganda (Montgomery dkk., 2001).

2.6 Pemusatan dan Penskalaan Data

Pemusatan dan penskalaan data merupakan bagian dari membakukan variabel. Modifikasi sederhana dari pembakuan variabel ini adalah transformasi korelasi (*Correlation Transformation*). Salah satu metode pemusatan dan penskalaan data adalah *unit length scalling* dengan rumus sebagai berikut:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{S_{jj}^{1/2}}$$

$$y_i^* = \frac{y_i - \bar{y}}{SS_T^{1/2}}$$

dengan

$$S_{jj} = \sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j\right)^2$$

$$SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Model regresi dengan transformasi variabel y_i^* dan z_{ij} yang didefiniskan dengan transformasi korelasi diatas disebut model regresi baku. Model tersebut adalah sebagai berikut:

$$y_i^* = \beta_1^* z_{i1} + \beta_2^* z_{i2} + \cdots + \beta_j^* z_{ij} + \cdots + \beta_p^* z_{ip} + \varepsilon_i \tag{2.4}$$

dengan,

$y_i^* =$ Variabel terikat untuk pengamatan ke-i yang telah distandarisasi

$z_{ij} =$ Variabel bebas untuk pengamatan ke-i variabel ke-j yang telah distandarisasi

$\beta_j^* =$ Parameter regresi ke-j

$\varepsilon_i =$ Sisaan pengamatan ke-i

Persamaan (2.4) dapat diubah ke dalam bentuk notasi matriks, sebagai berikut:

$$\boldsymbol{y}^* = \boldsymbol{Z}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon} \tag{2.5}$$

dengan,

$$\boldsymbol{y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{bmatrix} \qquad \boldsymbol{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{np} \end{bmatrix} \qquad \boldsymbol{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_p^* \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$\boldsymbol{y}^* =$ Vektor dari variabel terikat untuk data yang telah distandarisasi

$\boldsymbol{Z} =$ Matriks dari variabel bebas untuk data yang telah distandarisasi

$\boldsymbol{\beta}^* =$ Vektor parameter regresi yang telah distandarisasi

$\boldsymbol{\varepsilon} =$ Vektor sisaan pengamatan yang yang telah distandarisasi

hubungan antara parameter β_j^* pada model regresi baku dengan parameter $\beta_0, \beta_1, \dots, \beta_p$ pada model regresi awal adalah sebagai berikut:

$$\beta_j = \beta_j^* \left(\frac{SS_T}{S_{jj}}\right)^{1/2} \tag{2.6}$$

dan

$$\beta_0 = \bar{y} - \sum_{j=1}^{p} \beta_j \bar{x}_j \tag{2.7}$$

2.7 Metode Kuadrat Terkecil

Metode Kuadrat Terkecil (MKT) merupakan salah satu metode untuk mengestimasi parameter regresi dalam analisis regresi linier berganda. MKT bertujuan untuk meminimumkan jumlah kuadrat sisaan (Montgomery dkk, 2001). Misalkan model yang akan diestimasi adalah parameter dari persamaan (2.2). Untuk mendapatkan nilai estimasi $\boldsymbol{y}$ yaitu $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ diperlukan nilai estimator untuk parameter $\boldsymbol{\beta}$ yaitu $\widehat{\boldsymbol{\beta}}$. Parameter regresi adalah $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$ dan dalam bentuk vektor dilambangkan sebagai berikut:

$$\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

Untuk mengestimasi parameter regresi MKT dilakukan dengan cara meminimumkan jumlah kuadrat sisaan $\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\varepsilon}_{\boldsymbol{i}}^{\boldsymbol{2}}$. Misalkan fungsi kuadrat sisaan adalah:

$$S(\boldsymbol{\beta}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\varepsilon}_{\boldsymbol{i}}^{\boldsymbol{2}} = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\boldsymbol{y} - \boldsymbol{X\beta})'\,(\boldsymbol{y} - \boldsymbol{X\beta}) \tag{2.8}$$

persamaan (2.6) dapat dinyatakan sebagai berikut:

$$S(\boldsymbol{\beta}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X\beta} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}$$

$$S(\boldsymbol{\beta}) = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta}$$

kemudian, $S(\boldsymbol{\beta})$ diturunkan terhadap $\boldsymbol{\beta}$ sehingga diperoleh estimator kuadrat terkecil untuk parameter $\boldsymbol{\beta}$ sebagai berikut:

$$\left.\frac{\boldsymbol{\partial} S}{\boldsymbol{\partial\beta}}\right|_{\widehat{\boldsymbol{\beta}}} = \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{y} + \boldsymbol{\beta}'\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{0}$$

$$\boldsymbol{0} - 2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{0}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-\boldsymbol{1}}\boldsymbol{X}'\boldsymbol{y} \tag{2.9}$$

2.8 Regresi *Ridge*

Ketika metode kuadrat terkecil diaplikasikan pada data yang mengalami multikolinearitas, metode ini akan menghasilkan estimasi yang buruk.

Multikolinearitas dapat menyebabkan variansi dari estimator MKT akan menjadi besar. Oleh karena itu, pada kasus multikolinearitas dengan menggunakan metode kuadrat terkecil menjadi kurang tepat digunakan. Pada tahun 1970, Hoerl dan Kennard memperkenalkan metode alternatif penaksiran parameter model regresi linier berganda yang dapat mengatasi pengaruh multikolinearitas, yaitu metode *ridge*. Metode *ridge* menghasilkan estimasi parameter yang disebut estimator *ridge*. Estimator *ridge* ini bersifat bias dan bergantung pada suatu konstanta yang disebut konstanta estimator *ridge*. Secara umum, estimator *ridge* dituliskan sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_R = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y} \tag{2.10}$$

dengan k merupakan konstanta bias *ridge*. Adapaun karakteristik dari estimator *ridge* yaitu:

1. Bias estimator *ridge*

   Ekspektasi dari estimator *ridge* yaitu:

$$\begin{aligned} E\left(\widehat{\boldsymbol{\beta}}_R\right) &= E((\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'E(\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} \\ &= \boldsymbol{Z}_k\boldsymbol{\beta} \end{aligned}$$

   dengan $\boldsymbol{Z}_k = (\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}$. Karena $E\left(\widehat{\boldsymbol{\beta}}_R\right) \neq \boldsymbol{\beta}$, maka taksiran *ridge* merupakan taksiran yang bias untuk $\boldsymbol{\beta}$**.**

2. Kovariansi estimator *ridge*

   Kovariansi dari estimator *ridge* yaitu:

$$\begin{aligned} Cov\left(\widehat{\boldsymbol{\beta}}_R\right) &= E\left(\left(\widehat{\boldsymbol{\beta}}_R - E\left(\widehat{\boldsymbol{\beta}}_R\right)\right)\left(\widehat{\boldsymbol{\beta}}_R - E\left(\widehat{\boldsymbol{\beta}}_R\right)\right)'\right) \\ &= E\left((\boldsymbol{Z}_k\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_k\boldsymbol{\beta})(\boldsymbol{Z}_k\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_k\boldsymbol{\beta})'\right) \\ &= \boldsymbol{Z}_k E\left((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right)\boldsymbol{Z}_k' \\ &= \boldsymbol{Z}_k\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}_k' \\ &= \sigma^2\boldsymbol{Z}_k(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}_k' \end{aligned}$$

3. *Mean square error* estimator *ridge*

Mean square error dari estimator *ridge* yaitu

$$MSE(\widehat{\boldsymbol{\beta}}_R) = Cov(\widehat{\boldsymbol{\beta}}_R) + \left(bias(\widehat{\boldsymbol{\beta}}_R)\right)\left(bias(\widehat{\boldsymbol{\beta}}_R)\right)'$$

$$= \sigma^2 \boldsymbol{Z}_k(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}_k' + (\boldsymbol{Z}_k - \boldsymbol{I})\boldsymbol{\beta}\left((\boldsymbol{Z}_k - \boldsymbol{I})\boldsymbol{\beta}\right)'$$

$$= \sigma^2 \boldsymbol{Z}_k(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}_k' + (\boldsymbol{Z}_k - \boldsymbol{I})\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{Z}_k - \boldsymbol{I})'$$

sehingga skalar *mean square error* dari estimator *ridge* yaitu:

$$mse(\widehat{\boldsymbol{\beta}}_R) = tr\ MSE(\widehat{\boldsymbol{\beta}}_R)$$

$$= tr\left(Cov(\widehat{\boldsymbol{\beta}}_R) + \left(bias(\widehat{\boldsymbol{\beta}}_R)\right)\left(bias(\widehat{\boldsymbol{\beta}}_R)\right)'\right)$$

$$= tr(\sigma^2 \boldsymbol{Z}_k(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{Z}_k{}') + tr\left((\boldsymbol{Z}_k - \boldsymbol{I})\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{Z}_k - \boldsymbol{I})'\right)$$

$$= \sigma^2 tr((\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1}) + tr\Big(((\boldsymbol{X}'\boldsymbol{X} +$$

$$k\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} - \boldsymbol{I})\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})^{-1} - \boldsymbol{I})\Big)$$

Berikut ini dijelaskan pemilihan konstanta k yang optimal pada skalar *mean square error* estimator *ridge*. Skalar *mean square error* estimator *ridge* dapat dipandang sebagai fungsi dari konstanta taksrian *ridge* yaitu:

$$f(k, \lambda, \beta) = \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + k^2 \sum_{j=1}^{p} \frac{\beta_j^2}{\left(\lambda_j + k\right)^2}$$

untuk mencari nilai k optimal yang membuat $f(k, \lambda, \beta)$ seminimal mungkin maka berdasarkan teorema tentang nilai minimum fungsi, dibentuk turunan pertama fungsi $f(k, \lambda, \beta)$ sama dengan nol. Turunan pertama dari fungsi $f(k, \lambda, \beta)$ terhadap k yaitu:

$$\frac{\partial f(k, \lambda, \beta)}{\partial k} = -2\sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^3} + 2 \sum_{j=1}^{p} \frac{k\lambda_j\beta_j^2}{\left(\lambda_j + k\right)^3}$$

sehingga,

$$-2\sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^3} + 2 \sum_{j=1}^{p} \frac{k\lambda_j\beta_j^2}{\left(\lambda_j + k\right)^3} = 0 \qquad (2.11)$$

untuk mendapatkan k optimal yang meminimumkan fungsi $f(k)$, maka k dinyatakan sebagai bentuk perkalian atau penjumalahan dari λ_j, β_j^2 dan σ^2. Akan tetapi, pada persamaan (2.11) terlihat bahwa k sulit dipisahkan dengan λ_j, β_j^2 dan σ^2 karena k dalam bentuk kuadrat dan pangkat tiga. Oleh karena itu, nilai k optimal meminimumkan fungsi $f(k)$ sulit ditentukan. Karena k optimal sulit ditentukan, maka cukup digunakan k yang membuat penurunan nilai komponen variansi dari estimator *ridge* lebih besar dari kenaikan nilai komponen bias estimator *ridge*, dengan komponen variansi dari estimator *ridge* yaitu:

$$\sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^2}$$

dan komponen bias dari estimator *ridge* yaitu:

$$k^2 \sum_{j=1}^{p} \frac{\beta_j^2}{\left(\lambda_j + k\right)^2}$$

(Pitrianingsih, skripsi 2013)

2.9 Regresi *Generalized Ridge*

Pada tahun yang sama, Hoerl dan Kennard memperkenalkan metode alternatif lain yang merupakan pengembangan dari metode regresi *ridge* yaitu metode *Generalized Ridge Regression* (GRR). Estimasi parameter yang dihasilkan oleh metode GRR ini adalah estimator *generalized ridge*. Jika dalam estimator *ridge* dibutuhkan satu nilai untuk konstanta estimator *ridge*, maka dalam estimator GRR dibutuhkan kosntanta sejumlah variabel bebasnya. Persamaan umum regresi dalam bentuk kanonik dapat ditulis sebagai berikut:

$$\boldsymbol{y} = \boldsymbol{C\gamma} + \boldsymbol{\varepsilon} \tag{2.12}$$

dengan $\boldsymbol{C} = \boldsymbol{XQ}$ dan $\boldsymbol{\gamma} = \boldsymbol{Q}'\boldsymbol{\beta}$, sehingga berdasarkan persamaan 2.12 diperoleh estimator estimator *generalized ridge* sebagai berikut:

$$\hat{\boldsymbol{\gamma}}_{GR} = (\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{K})^{-1}\boldsymbol{C}'\boldsymbol{y} \tag{2.13}$$

dengan $\boldsymbol{K} = \text{diag}(k_1, k_2, \dots, k_p)$ merupakan konstanta bias dari estimator *generalized ridge*. Adapun karakteristik dari estimator *generalized ridge* yaitu:

1. Bias estimator *generalizedd ridge*

   Ekspektasi dari estimator *generalized ridge* yaitu:

$$E(\hat{\boldsymbol{\gamma}}_{GR}) = E((\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'y})$$
$$= (\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'}E(\boldsymbol{y})$$

karena $E(\boldsymbol{y}) = \boldsymbol{C\gamma}$, maka

$$E(\hat{\boldsymbol{\gamma}}_{GR}) = (\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'C\gamma}$$
$$= \boldsymbol{Z_k\gamma}$$

dengan $\boldsymbol{Z_k} = (\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'C}$. Karena $E(\hat{\boldsymbol{\gamma}}_{GR}) \neq \boldsymbol{\gamma}$, maka estimator *generalized ridge* merupakan estimator yang bias untuk $\boldsymbol{\gamma}$.

2. Variansi estimator *generalized ridge*

   Variansi dari estimator *generalized ridge* yaitu:

$$Var(\hat{\boldsymbol{\gamma}}_{GR}) = Var((\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'y})$$

berdasarkan sifat variansi yaitu $\text{Var}(\boldsymbol{Ay}) = \boldsymbol{AVA'}$ dengan $Var(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, maka

$$Var(\hat{\boldsymbol{\gamma}}_{GR}) = (\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'}Var(\boldsymbol{y})((\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'y})'$$

karena $Var(\boldsymbol{y}) = \sigma^2\boldsymbol{I}$, maka

$$Var(\hat{\boldsymbol{\gamma}}_{GR}) = (\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'}\sigma^2\boldsymbol{I}((\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'y})'$$
$$= \sigma^2(\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'C}(\boldsymbol{C'C} + \boldsymbol{K})^{-1}$$

sehingga nilai skalar dari variansi $\hat{\boldsymbol{\gamma}}_{GR}$ adalah

$$var(\hat{\boldsymbol{\gamma}}_{GR}) = \sigma^2 tr(\boldsymbol{C'C} + \boldsymbol{K})^{-1}\boldsymbol{C'C}(\boldsymbol{C'C} + \boldsymbol{K})^{-1}$$
$$= \sigma^2 \sum_{j=1}^{p} \frac{1}{\lambda_j + k_j}\lambda_j \frac{1}{\lambda_j + k_j}$$
$$= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k_j)^2}$$

3. *Mean Square Error* estimator *generalized ridge*

Mean Square Error dari estimator *generalized ridge* yaitu:

$$\begin{aligned} MSE(\hat{\boldsymbol{\gamma}}_{GR}) &= E[(\hat{\boldsymbol{\gamma}}_{GR} - \boldsymbol{\gamma})'(\hat{\boldsymbol{\gamma}}_{GR} - \boldsymbol{\gamma})] \\ &= Var(\hat{\boldsymbol{\gamma}}_{GR}) + (\text{bias } \hat{\boldsymbol{\gamma}}_{GR})^2 \\ &= Var(\hat{\boldsymbol{\gamma}}_{GR}) + (E(\hat{\boldsymbol{\gamma}}_{GR}) - \boldsymbol{\gamma})^2 \\ &= \sigma^2(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{K})^{-1}\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{K})^{-1} + (\boldsymbol{Z_k}\boldsymbol{\gamma} - \boldsymbol{\gamma})'(\boldsymbol{Z_k}\boldsymbol{\gamma} - \boldsymbol{\gamma}) \end{aligned}$$

sehingga sklar dari *mean ssquare error* dari estimator *generalized ridge* yaitu:

$$\begin{aligned} mse(\hat{\boldsymbol{\gamma}}_{GR}) &= \sigma^2 tr(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{K})^{-1}\boldsymbol{C}'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C} + \boldsymbol{K})^{-1} + tr(\boldsymbol{Z_k}\boldsymbol{\gamma} - \boldsymbol{\gamma})'(\boldsymbol{Z_k}\boldsymbol{\gamma} - \boldsymbol{\gamma}) \\ &= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k_j)^2} + \sum_{j=1}^{p} \frac{\gamma_j^2 k_j^2}{(\lambda_j + k_j)^2} \end{aligned}$$

$$mse(\hat{\boldsymbol{\gamma}}_{GR}) = \sum_{j=1}^{p} \frac{\sigma^2\lambda_j + \gamma_j^2 k_j^2}{(\lambda_j + k_j)^2}$$

Berikut ini akan dijelaskan pemilihan konstanta k_j yang optimal pada skalar *mean square error* estimator *generalized ridge*. Nilai optimal untuk k_j diperole dengan meminimumkan *mean square error* dari estimator *generalized ridge*, yaitu Q.

$$mse(\hat{\boldsymbol{\gamma}}_{GR}) = Q = \sum_{j=1}^{p} \frac{\sigma^2\lambda_j + \gamma_j^2 k_j^2}{(\lambda_j + k_j)^2}$$

Fungsi Q diturunkan secara parsial sayu kalli terhadap k_j dan hasil turunannya disamakan dengan nol. Turunan dari fungsi Q terhadap k_j adalah:

$$\begin{aligned} \frac{\partial Q}{\partial k_j} &= \frac{2\gamma_j^2 k_j^2 (\lambda_j + k_j)^2 - 2(\sigma^2\lambda_j + \gamma_j^2 k_j^2)(\lambda_j + k_j)}{(\lambda_j + k_j)^4} \\ &= \frac{2(\lambda_j + k_j)[\gamma_j^2 k_j(\lambda_j + k_j) - \sigma^2\lambda_j + \gamma_j^2 k_j^2]}{(\lambda_j + k_j)^4} \\ &= \frac{2\lambda_j(\lambda_j + k_j)(k_j\gamma_j^2 - \sigma^2)}{(\lambda_j + k_j)^4} \end{aligned}$$

Berdasarkan definisi tentang matriks definit positif, suatu matriks simetris $\boldsymbol{X}'\boldsymbol{X}$ disebut matriks definit positif jika $\boldsymbol{v}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{v} > 0$ untuk semua $\boldsymbol{v} \neq \boldsymbol{0}$.

Misalkan untuk suatu vektor:

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_p \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

dengan mengalikan $\boldsymbol{X'X}$ dengan $\boldsymbol{v}$ dari arah kiri maupun kanan sebagai berikut:

$$\begin{aligned} \boldsymbol{v}'(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{v} &= \boldsymbol{v}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{v} = (\boldsymbol{X}\boldsymbol{v})'\boldsymbol{X}\boldsymbol{v} \\ &= \langle \boldsymbol{X}\boldsymbol{v}, \boldsymbol{X}\boldsymbol{v} \rangle = \|\boldsymbol{X}\boldsymbol{v}\|^2 > 0 \end{aligned}$$

sehingga matriks $\boldsymbol{X'X}$ merupakan matriks definit positif. Karena matriks simestris $\boldsymbol{X'X}$ definit positif maka $\lambda_j > 0, j = 1,2, \dots, p$. Nilai minimum diperoleh dengan $\frac{\partial Q}{\partial k_j} = 0$, sehingga

$$\begin{aligned} \left(k_j\gamma_j^2 - \sigma^2\right) &= 0 \\ k_j\gamma_j^2 &= \\ k_j &= \frac{\sigma^2}{\gamma_j^2} \end{aligned}$$

Namun, nilai k_j yang optimal bergantung pada parameter σ^2 dan γ_j^2 yang tidak diketahui nilainya. Pendekatan iterasi akan digunakan untuk mendapatkan nilai k_j (Hoerl dan Kennard, 1970). Solusi dari MKT digunakan sebagai nilai awal dari k_j. Proses iterasi akan terus berlangsung hingga didapatkan nilai estimasi parameter yang stabil. Ukuran yang digunakan untuk mengukur kestabilannya adalah kuadrat dari $\hat{\boldsymbol{\gamma}}_{GR}'\hat{\boldsymbol{\gamma}}_{GR}$. Pada penelitian ini, apabila selisih kuadrat panjang $\hat{\boldsymbol{\gamma}}_{GR}'\hat{\boldsymbol{\gamma}}_{GR}$ dari iterasi ke-i dan iterasi ke $i-1$ kurang dari 10^{-4}, maka iterasi berhenti. Dengan menggunakan nilai k_j yang diperoleh dari iterasi, maka akan didapatkan nilai *mean square errror* estimator *generalized ridge* yang minimum (Montgomery, Peck & Vinning, 2001).
(Walidya, Skripsi 2013)

2.10 Regresi Liu

Regresi *ridge* menggunakan konstanta k yang sulit ditentukan dan tidak mengatasi pengaruh multikolinearitas secara signifikan. Pada tahun 1993, Liu Kejian mengusulkan metode penaksiran baru yang dapat memperbaiki kelemahan pada regresi *ridge* yang dikenal dengan metode Liu. Metode Liu menghasilkan estimasi parameter yang disebut estimator Liu. Estimator Liu ini bersifat bias dan bergantung pada suatu konstanta yang disebut konstanta estimator Liu. Secara umum, estimator Liu dituliskan sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_L = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{y} + d\widehat{\boldsymbol{\beta}}) \tag{2.14}$$

dengan d merupakan konstanta bias liu. Adapun karakteristik dari estimator *generalized ridge* yaitu:

1. Bias estimator Liu

   Ekspektasi dari estimator liu yaitu:

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_L) &= E\left((\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})\widehat{\boldsymbol{\beta}}\right) \\ &= (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})E(\widehat{\boldsymbol{\beta}}) \\ &= (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})\boldsymbol{\beta} \\ &= \boldsymbol{F}_d\boldsymbol{\beta} \end{aligned}$$

dengan $\boldsymbol{F}_d = (\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})$. Karena $E(\widehat{\boldsymbol{\beta}}_L) \neq \boldsymbol{\beta}$, maka estimator Liu merupakan estimator yang bias untuk $\boldsymbol{\beta}$.

2. Kovariansi estimator Liu

   Kovariansi dari estimator Liu yaitu:

$$\begin{aligned} Cov(\widehat{\boldsymbol{\beta}}_L) &= E\left((\boldsymbol{F}_d\widehat{\boldsymbol{\beta}} - \boldsymbol{F}_d\boldsymbol{\beta})(\boldsymbol{F}_d\widehat{\boldsymbol{\beta}} - \boldsymbol{F}_d\boldsymbol{\beta})'\right) \\ &= E\left(\left(\boldsymbol{F}_d(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)\left(\boldsymbol{F}_d(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)'\right) \\ &= E\left(\left(\boldsymbol{F}_d(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{F}_d{}'\right) \\ &= \boldsymbol{F}_d E\left((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\right)\boldsymbol{F}_d{}' \\ &= \boldsymbol{F}_d\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{F}_d{}' \end{aligned}$$

$$Cov(\widehat{\boldsymbol{\beta}}_L) = \sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d{}'$$

3. *Mean Square Error* estimator Liu

Mean square error dari estimator Liu yaitu:

$$\begin{aligned} MSE(\widehat{\boldsymbol{\beta}}_L) &= \sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d' + \left(E(\widehat{\boldsymbol{\beta}}_L) - \boldsymbol{\beta}\right)\left(E(\widehat{\boldsymbol{\beta}}_L) - \boldsymbol{\beta}\right)' \\ &= \sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d' + (-(1-d)(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta})(-(1-d)(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta})' \\ &= \sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d' + \left((d-1)(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta}\right)\left((d-1)(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta}\right)' \\ &= \sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d' + (d-1)(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(d-1) \\ &= \sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d' + (d-1)^2(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1} \end{aligned}$$

sehingga nilai skalar *mean square error* dari estimator Liu yaitu:

$$\begin{aligned} mse(\widehat{\boldsymbol{\beta}}_L) &= tr\, MSE(\widehat{\boldsymbol{\beta}}_L) \\ &= tr(\sigma^2 \boldsymbol{F}_d (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{F}_d' + (d-1)^2(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}) \\ &= \sigma^2 tr((\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}(\boldsymbol{X}'\boldsymbol{X}+d\boldsymbol{I})^2(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^{-1}) \\ &\quad + (d-1)^2\boldsymbol{\beta}'(\boldsymbol{X}'\boldsymbol{X}+\boldsymbol{I})^2\boldsymbol{\beta} \end{aligned}$$

Berikut ini akan dijelaskan pemilihan konstanta d yang optimal pada skala estimator Liu. Skalar *mean square error* estimator Liu dapar dipandang sebagai fungsi dari konstanta estimator Liu yaitu:

$$f(d) = \sigma^2 \sum_{j=1}^{p} \frac{\left(\lambda_j + d\right)^2}{\lambda_j\left(\lambda_j + 1\right)^2} + (d-1)^2 \sum_{j=1}^{p} \frac{\beta_j^2}{\left(\lambda_j + 1\right)^2}$$

untuk mendapatkan d optimal yang meminimumkan fungsi $f(d)$, maka fungsu $f(d)$ diturunkan satu kali terhadap d dan hasil turunannya disamakan dengan 0. Turunan fungsi $f(d)$ terhadap d adalah

$$f'(d) = 2\sigma^2 \sum_{j=1}^{p} \frac{\left(\lambda_j + d\right)}{\lambda_j\left(\lambda_j + 1\right)^2} + 2(d-1) \sum_{j=1}^{p} \frac{\beta_j^2}{\left(\lambda_j + 1\right)^2} = 0$$

$$\sigma^2 \sum_{j=1}^{p} \frac{\left(\lambda_j + d\right)}{\lambda_j\left(\lambda_j + 1\right)^2} + (d-1) \sum_{j=1}^{p} \frac{\beta_j^2}{\left(\lambda_j + 1\right)^2} = 0$$

$$d_{opt} = \sum_{j=1}^{p} \frac{\left(\beta_j^2 - \sigma^2\right)}{\lambda_j\left(\lambda_j + 1\right)^2} \Big/ \sum_{j=1}^{p} \frac{\sigma^2 + \lambda_j \beta_j^2}{\lambda_j\left(\lambda_j + 1\right)^2}$$

dengan λ_j adalah nilai eigen dari matriks $\boldsymbol{X'X}$, $\beta_j, j = 1,2, \dots, p$ adalah parameter model regresi dan σ^2 adalah variansi *error* pada model regresi. Karena β_j dan σ^2 adalah parameter yang nilainya tidak diketahui maka β_j dan σ^2 diestimasi dengan menggunakan estimator MKT. Sehingga nilai d yang optimal dengan menggunakan $\widehat{\beta_J}$ dan $\hat{\sigma}^2$ adalah:

$$\hat{d}_{opt} = \sum_{j=1}^{p} \frac{\left(\widehat{\beta_J}^2 - \hat{\sigma}^2\right)}{\lambda_j\left(\lambda_j + 1\right)^2} \Big/ \sum_{j=1}^{p} \frac{\hat{\sigma}^2 + \lambda_j \widehat{\beta_J}^2}{\lambda_j\left(\lambda_j + 1\right)^2}$$

dengan menggunakan nilai $\widehat{\beta_J}^2$ maka akan didapatkan nilai skalar *mean square error* dari estimatior Liu yang minimum.

(Pitrianingsih, skripsi 2013)

2.11 Pengujian Signifikansi Parameter Model Regresi

Pengujian signifikansi parameter model regresi bertujuan untuk mengetahui apakah parameter yang terdapat dalam model regresi telah menunjukkan hubungan yang tepat antara variabel bebas dengan variabel terikat. Pengujian signifikansi parameter model regresi juga sebagai salah satu cara mengevaluasi seberapa baik model regresi yang diperoleh. Ada dua tahapan dalam pengujian signifikansi parameter model regresi, yaitu pengujian secara simultan dengan menggunakan uji F dan pengujian secara parsial dengan menggunakan uji t.

2.11.1 Uji F

Pengujian parameter dengan menggunakan uji F bertujuan untuk mengevaluasi pengaruh semua variabel bebas terhadap variabel terikat. Adapun hipotesis yang digunakan dalam uji F yaitu:

H_0: $\beta_{LRR_1} = \beta_{LRR_2} = \cdots = \beta_{LRR_p} = 0$

H_1: Ada $\beta_{LRR_j} \neq 0,\ j = 1, \dots, p$

dengan statistik uji sebagai berikut:

$$F_{hitung} = \frac{\text{Kuadrat Tengah Regresi (KTR)}}{\text{Kuadrat Tengah Sisaan (KTS)}} \tag{2.15}$$

untuk menghitung nilai KTR dan KTS dapat digunakan bantuan tabel seperti yang dinyatakan pada Tabel (2.1).

Tabel 2. 1 Analisis Ragam Untuk Pengujian Signifikansi Parameter

Sumber	Derajat Bebas	Jumlah Kuadrat	Kuadrat Tengah	F_{hitung}
Regresi	p	$JKR = \sum_{i=1}^{n} \left(\hat{y}_{lrr_i} - \bar{y}\right)^2$	$KTR = \frac{JKR}{p}$	
Sisaan	$n-p-1$	$JKS = \sum_{i=1}^{n} \left(y_i - \hat{y}_{lrr_i}\right)^2$	$KTS = \frac{JKS}{n-p-1}$	$F_{hitung} = \frac{\text{KTR}}{KTS}$
Total	$n-1$	$JKT = \sum_{i=1}^{n} (y_i - \bar{y})^2$		

dengan kriteria pengujian yaitu:

Jika $F_{hitung} \geq F_{tabel}$ maka tolak H_0 $\left(F_{tabel} = F_{(\alpha,p,n-p-1)}\right)$

2.11.2 Uji t

Uji t digunakan untuk menguji pengaruh setiap variabel bebas secara satu persatu terhadap variabel terikatnya. Adapun hipotesis yang digunakan dalam uji t yaitu:

$H_0: \beta_{LRR_j} = 0, j = 1,2, \dots, p$

$H_1: \beta_{LRR_j} \neq 0$

dengan statistik uji sebagai berikut:

$$t_{hitung_j} = \frac{\hat{\beta}_{LRR_j}}{SE(\hat{\beta}_{LRR_j})} \tag{2.16}$$

dengan $\text{SE}\left(\hat{\beta}_{LRR_j}\right) = \sqrt{\text{Var}\hat{\beta}_{LRR_j}}$. Dengan kriteria pengujian yaitu:

Jika $\left|t_{hitung}\right| \geq t_{tabel}$ maka tolak H_0 $\left(t_{tabel} = t_{\left(\frac{\alpha}{2},\text{n}-\text{p}-1\right)}\right)$

2.12 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, 2018) atau secara matematis dapat dituliskan menjadi sebagai berikut:

$$\mathrm{AKB} = \frac{\text{Jumlah Kematian Bayi}}{\text{Jumlah Kelahiran Bayi}} \times 1000$$

Angka kematian bayi merupakan salah satu indikator penting dalam menentukan tingkat kesehatan masyarakat karena dapat menggambarkan kesehatan penduduk secara umun dan termasuk didalam 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 (SDK1) tahun 2019 yang menunjukkan bahwa Angka Kematian Bayi (AKB) sebesar 6.02 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.