PERFORMA *BETA RIDGE REGRESSION ESTIMATOR* UNTUK MENGATASI MULTIKOLINEARITAS DALAM DISTRIBUSI BETA

(SKRIPSI)

Oleh

RATNA ARUM SARI

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025

ABSTRACT

PERFORMANCE OF BETA RIDGE REGRESSION ESTIMATOR TO OVERCOME MULTICOLLINEARITY IN BETA DISTRIBUTION

By

Ratna Arum Sari

Beta Ridge Regression (BRR) is a ridge method applied in the beta regression model used to overcome the problem of multicollinearity, which is a condition in which the independent variables in the regression model have a high correlation. This problem can cause parameter estimates to be unstable and less accurate. This study aims to determine the performance of BRR estimator in overcoming multicollinearity in simulated data with small sample size. The analysis is done by comparing the estimation results based on the Mean Squared Error (MSE) and Mean Absolute Error (MAE) values. The results show that the proposed BRR estimator has superior performance compared to the Maximum Likelihood Estimation (MLE) method, by producing lower MSE and MAE values than MLE.

Keywords: Multicollinearity, Beta Ridge Regression, Beta Distribution, Simulated Data, Mean Squared Error, Mean Absolute Error.

ABSTRAK

PERFORMA *BETA RIDGE REGRESSION ESTIMATOR* UNTUK MENGATASI MULTIKOLINEARITAS DALAM DISTRIBUSI BETA

Oleh

Ratna Arum Sari

Beta Ridge Regression (BRR) adalah metode *ridge* yang diterapkan dalam model regresi beta yang digunakan untuk mengatasi masalah multikolinearitas, yaitu kondisi di mana terjadi variabel bebas dalam model regresi memiliki korelasi yang tinggi. Masalah ini dapat menyebabkan estimasi parameter menjadi tidak stabil dan kurang akurat. Penelitian ini bertujuan untuk mengetahui performa BRR *estimator* dalam mengatasi multikolinearitas pada data simulasi dengan ukuran sampel kecil. Analisis dilakukan dengan membandingkan hasil estimasi berdasarkan nilai *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). Hasil penelitian menunjukkan bahwa BRR *estimator* yang diusulkan memiliki performa yang lebih unggul dibandingkan metode *Maximum Likelihood Estimation* (MLE), dengan menghasilkan nilai MSE dan MAE yang lebih rendah dibandingkan MLE.

Kata Kunci: Multikolinearitas, *Beta Ridge Regression*, Distribusi Beta, Data Simulasi, *Mean Squared Error*, *Mean Absolute Error*.

PERFORMA *BETA RIDGE REGRESSION ESTIMATOR* UNTUK MENGATASI MULTIKOLINEARITAS DALAM DISTRIBUSI BETA

Oleh

RATNA ARUM SARI

SKRIPSI

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Lampung

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2025

Judul Skripsi : **PERFORMA *BETA RIDGE REGRESSION ESTIMATOR* UNTUK MENGATASI MULTIKOLINEARITAS DALAM DISTRIBUSI BETA**

Nama Mahasiswa : **Ratna Arum Sari**

Nomor Pokok Mahasiswa : **2117031009**

Program Studi : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Prof. Ir. Netti Herawati, M.Sc., Ph.D.	**Misgiyati, S.Pd., M.Si.**
NIP. 196501251990032001	NIP. 198509282023212032

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si., M.Si.
NIP. 197403162005011001

MENGESAHKAN

1. Tim Penguji

Ketua	: **Prof. Ir. Netti Herawati, M.Sc., Ph.D.** ....
Sekretaris	: **Misgiyati, S.Pd., M.Si.** ....
Penguji Bukan Pembimbing	: **Dr. Khoirin Nisa, S.Si., M.Si.** ....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Dr. Eng. Heri Satria, S.Si., M.Si.
NIP 197110012005011002

Tanggal Lulus Ujian Skripsi: **11 Februari 2025**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini:

Nama Mahasiswa	: **Ratna Arum Sari**
Nomor Pokok Mahasiswa	: **2117031009**
Jurusan	: **Matematika**
Judul Skripsi	: **PERFORMA *BETA RIDGE REGRESSION ESTIMATOR* UNTUK MENGATASI MULTIKOLINEARITAS DALAM DISTRIBUSI BETA**

Dengan ini menyatakan bahwa penelitian ini adalah hasil kerja saya sendiri. Apabila kemudian hari terbukti bahwa skripsi ini merupakan salinan atau dibuat oleh orang lain, maka saya bersedia menerima sanksi sesuai dengan ketentuan akademik yang berlaku.

Bandar Lampung, 11 Februari 2025

Penulis,

Ratna Arum Sari

NPM. 2117031009

RIWAYAT HIDUP

Penulis bernama Ratna Arum Sari dilahirkan di Pringsewu pada 4 September 2003. Penulis merupakan anak ketiga dari empat bersaudara dari pasangan Bapak Hardi Dwi Handoko dan Ibu Sri Nuryati.

Penulis pertama kali menempuh pendidikannya di Taman Kanak-Kanak Muslimat NU Bulukarto pada tahun 2008-2009, kemudian melanjutkan pendidikannya ke Sekolah Dasar di SD Negeri 03 Bulukarto dari tahun 2009 hingga 2015. Setelah menyelesaikan pendidikan dasar, penulis melanjutkan pendidikannya di SMP Negeri 03 Pringsewu pada tahun 2015-2018, lalu melanjutkan ke SMA Negeri 01 Pringsewu pada tahun 2018-2021. Kemudian pada tahun 2021 penulis diterima sebagai mahasiswa Program Studi S1 Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN).

Selama menjalani pendidikan di Universitas Lampung, penulis aktif dalam berbagai organisasi. Pada tahun 2022, penulis tergabung dalam Himpunan Mahasiswa Matematika (HIMATIKA) FMIPA UNILA dan pada tahun 2023, penulis menjadi bagian dari Badan Eksekutif Mahasiswa (BEM) FMIPA UNILA. Pada tahun 2024 penulis melaksanakan Kerja Praktik (KP) di Badan Pusat Statistik Kabupaten Pesawaran serta melaksanakan Kuliah Kerja Nyata (KKN) di Desa Bandar Agung, Kecamatan Sribawono, Kabupaten Lampung Timur. Pada tahun yang sama, penulis berhasil meraih beasiswa Bank Indonesia dan tergabung dalam sebuah komunitas Generasi Baru Indonesia (GenBI) yaitu sebuah organisasi yang menaungi para penerima beasiswa Bank Indonesia.

KATA INSPIRASI

"Sesungguhnya bersama kesulitan ada kemudahan."
(Q.S Al-insyirah:6)

"Jika berbuat baik, (berarti) kamu telah berbuat baik untuk dirimu sendiri.
Jika kamu berbuat jahat (kerugian dan kejahatan) itu akan
kembali kepada dirimu sendiri."
(Q.S Al-Isra:7)

"Tidak ada mimpi yang terlalu tinggi dan tidak ada mimpi yang patut diremehkan.
Lambungkan setinggi yang kau inginkan dan gapailah dengan
selayaknya yang kau harapkan."
(Maudy Ayunda)

"Terkadang, sesuatu yang paling kau takuti merupakan hal yang membebaskanmu."
(Leo Tolstoy)

"Orang lain gak akan paham struggle dan masa sulitnya kita, yang mereka ingin tahu hanya bagian succes storiesnya aja. Jadi berjuanglah untuk diri sendiri meskipun gak akan ada yang tepuk tangan. Kelak diri kita di masa depan akan sangat bangga dengan apa yang kita perjuangkan hari ini.
Jadi tetap berjuang ya!"
(Fardi Yandi)

"Jangan biarkan mimpi hanya menjadi bayangan dalam tidurmu. Bangun, perjuangkan, dan wujudkan. Jadikan setiap langkahmu sebagai jalan menuju keberhasilan."
(Ratna Arum Sari)

PERSEMBAHAN

Puji dan syukur kepada Allah SWT yang telah memberikan petunjuk, kemudahan, serta kelancaran dalam menyelesaikan skripsi ini. Dengan penuh rasa syukur dan kebahagiaan saya persembahkan karya sederhana ini kepada:

Orang Tua, Kakak dan Adik Tercinta

Yang senantiasa memberikan dukungan, doa, dan cintanya tanpa henti, sehingga menjadi sumber kekuatan dan motivasi bagi penulis untuk bertahan dan berjuang dalam menyelesaikan skripsi ini. Tanpa kalian, penulis tidak akan berada dititik ini. Sebuah perjuangan panjang yang penuh dengan cerita, yang tak lepas dari doa tulus Mamak, Bapak, Mas Arif, Mas Rizki, dan Laila. Terima kasih atas segala pengorbanan dan kasih sayang yang telah kalian berikan.

Dosen Pembimbing dan Penguji

Yang tulus memberikan motivasi, ilmu, dan arahan yang berharga bagi penulis.

Almameter Tercinta, Universitas Lampung

SANWACANA

Alhamdulillahi rabbil 'alamin, segala puji dan syukur penulis panjatkan kepada Allah SWT, yang telah melimpahkan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "Performa *Beta Ridge Regression Estimator* Untuk Mengatasi Multikolinearitas Dalam Distribusi Beta". Penyusunan skripsi ini, tidak akan terselesaikan dengan baik tanpa adanya bantuan, bimbingan, dan dukungan dari berbagai pihak.

Sehingga, dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Ibu Prof. Ir. Netti Herawati, M.Sc., Ph.D. selaku dosen pembimbing I dan dosen pembimbing akademik yang telah memberikan bimbingan, arahan dan masukan dalam proses menyelesaikan skripsi.
2. Ibu Misgiyati, S.Pd., M.Si. selaku dosen pembimbing II yang telah memberikan bimbingan, arahan, serta saran dalam proses penyusunan skripsi.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si. selaku dosen penguji yang telah memberikan kritik dan saran, sehingga skripsi ini dapat terselesaikan.
4. Bapak Dr. Aang Nuryaman, S.Si, M.Si. selaku Ketua Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
5. Bapak. Dr. Eng. Heri Satria, S.Si., M.Si. selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
6. Dua sosok yang sangat berharga, Ibu Sri Nuryati dan Bapak Hardi. Terima kasih telah menjadi sumber semangat dan alasan terbesar bagi penulis untuk terus bertahan dan berjuang. Segala pencapaian ini tidak akan terwujud tanpa doa, cinta, serta dukungan tanpa henti dari kalian. Setiap tetes keringat, pengorbanan, dan kasih sayang yang kalian berikan menjadi kekuatan terbesar dalam perjalanan ini.

7. Ketiga saudara kandung penulis, Mas Arif, Mas Rizki, dan Laila. Terima kasih telah menjadi saudara yang selalu ada di setiap langkah penulis, memberikan dukungan dalam segala bentuk, dari hal-hal kecil hingga yang paling berarti. Kalian telah menjadi bagian yang sangat penting dalam perjalanan ini. Terima kasih telah membawa keceriaan, kebahagiaan, dan semangat yang tak terhingga dalam hidup penulis.
8. Kepada Kakak ipar Yuni Lestari dan keponakan Salsabila, Terima kasih atas kasih sayang dan dukungan yang kalian berikan kepada penulis. Kehadiran kalian membawa kebahagiaan dan membuat hidup ini lebih berarti.
9. Ajeng, Amanda, Hania, Nindi, Nisa, Rafika, Sinta, dan Zahra (IK). Terima kasih atas kebersamaan yang tak terlupakan dari semester 1. Setiap momen, tawa, dan dukungan kalian adalah bagian dari perjalanan ini. *Good luck.*
10. Teman sejak kecil, Ayu, Diah, dan Mufi. Kalian adalah bagian dari kenangan terbaik dalam hidup penulis. Terima Kasih untuk setiap tawa, cerita, dan dukungan yang diberikan kepada penulis.
11. Teman sejalan sejak SMP, Muti, Sakinah, Senja, dan Zahra. Terima kasih telah menjadi bagian dari perjalanan ini sejak awal, dari tawa, cerita, hingga segala dukungan yang selalu diberikan kepada penulis.
12. Seluruh pihak terkait yang tidak bisa penulis sebutkan satu persatu yang telah terlibat dalam menyelesaikan skripsi ini.
13. Terakhir, terimakasi untuk diri sendiri, karena telah mampu berusaha dan berjuang sejauh ini, mampu mengendalikan berbagai tekanan dari luar dan tetap tidak pernah menyerah.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, namun penulis berharap semoga skripsi ini dapat memberikan manfaat kepada kita.

Bandar Lampung, 11 Februari 2025

Penulis

Ratna Arum Sari

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan suatu metode statistik yang digunakan untuk menguji pengaruh satu atau lebih variabel independen terhadap satu variabel dependen. Dengan regresi, kita dapat mengetahui seberapa besar pengaruh variabel bebas terhadap variabel terikat dan bentuk hubungan antara keduanya (Ghozali, 2016). Variabel dependen (variabel terikat) adalah variabel yang dipengaruhi oleh variabel lain, sedangkan variabel independen (variabel bebas) adalah variabel yang mempengaruhi variabel dependen. Variabel ini dianggap sebagai penyebab atau faktor yang menyebabkan perubahan pada variabel dependen.

Model regresi yang sering digunakan di berbagai bidang umumnya adalah model regresi klasik yang diasumsikan galat berdistribusi normal. Namun, dalam beberapa bidang penelitian, model ini tidak selalu dapat digunakan. Untuk mengatasi masalah tersebut, pendekatan *maksimum likelihood* dengan menggunakan fungsi likelihood dari distribusi binomial dapat diterapkan sebagai dasar dalam pemodelan regresi logistik. Pendekatan lain yang dapat digunakan adalah dengan pemodelan regresi beta (Hajarisman, 2012).

Menurut Swearingen, *et al*. (2011) model regresi beta menghasilkan penaksiran parameter yang lebih akurat dan efisien dibandingkan dengan metode kuadrat terkecil biasa, terutama ketika variabel respon memiliki distribusi yang tidak simetris, atau saat terjadi masalah heteroskedastisitas. Distribusi beta umumnya digunakan untuk memodelkan data yang berada dalam rentang (0,1) dan ditentukan oleh dua parameter yaitu a dan b. Kedua parameter ini memengaruhi bentuk fungsi kepadatan yang dapat berbentuk rata, simetris, asimetris ke kiri maupun ke kanan. Fungsi kepadatan distribusi beta dapat menampilkan bentuk yang bervariasi tergantung pada nilai parameter, yang membuat distribusi beta menjadi distribusi yang tepat untuk memodelkan data dalam interval (0,1) (Cribari-Neto & Vasconcellos, 2002).

Parameter distribusi beta dapat diestimasi menggunakan bebagai pendekatan. Dishon dan Weiss (1980) membandingkan performa estimasi parameter distribusi beta pada sampel kecil menggunakan metode maksimum *likelihood* dan metode momen. Hasilnya penelitian mereka menunjukkan bahwa metode *maksimum likelihood* memberikan estimasi yang lebih unggul. Model regresi beta umumnya digunakan ketika menganalisis data yang berbentuk rasio atau persentase. Namun, masalah yang mungkin muncul ketika menganalisis data ini adalah multikolinearitas. Multikolinearitas merupakan masalah umum dalam pemodelan ekonometrik yang diperkenalkan oleh Frisch pada tahun 1934. Multikolinearitas terjadi ketika terdapat korelasi yang tinggi antar variabel bebas yang dapat menimbulkan masalah dalam menginterpretasikan dan memodelkan data. Terjadinya multikolinearitas menyebabkan penduga kuadrat terkecil memiliki variansi yang besar (Montgomery & Peck, 1992).

Menurut Abonazel & Taha (2021), masalah multikolinearitas dapat diatasi dengan menggunakan metode *beta ridge regression* (BRR). BRR merupakan suatu variasi dari regresi beta yang menggunakan teknik *ridge regression* dalam menangani masalah multikolinearitas. Dalam regresi *ridge*, multikolinearitas dapat diatasi dengan menambahkan suatu nilai ketetapan bias k.Penambahan bias ini membantu mengurangi nilai variansi dari penduga parameter.

Penelitian sebelumnya yang membahas mengenai masalah multikolinearitas menggunakan metode *beta ridge regression* dilakukan oleh Abonazel & Taha (2021) dengan kesimpulan bahwa parameter *ridge* k yang telah diusulkan dan semuanya mengungguli estimator MLE dalam hal kriteria *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). Di antara estimator yang diusulkan, salah satu parameter *ridge* k memberikan hasil yang lebih efisien untuk n besar pada data simulasi yang diadopsi dalam penelitian ini, sehingga peneliti ini merekomendasikannya kepada peneliti selanjutnya. Kemudian penelitian yang dilakukan oleh Qasim, *et al*. (2021) dengan kesimpulan bahwa parameter *ridge* k yang telah diusulkan, dan semuanya mengungguli estimator MLE dalam hal kriteria *Mean Squarred Error* dan *Median Squarred Error*.

Berdasarkan latar belakang tersebut, penulis tertarik untuk melakukan penelitian dengan menggunakan metode *beta ridge regression* untuk mengatasi masalah multikolinearitas pada data simulasi untuk sampel (n) kecil.

1.2 Tujuan

Adapun tujuan yang ingin dicapai pada penelitian ini adalah:

1. Mengetahui performa *beta ridge regression estimator* dalam mengatasi masalah multikolinearitas pada data simulasi untuk sampel (n) kecil.
2. Membandingkan nilai dugaan dengan menggunakan kriteria nilai MSE dan MAE terkecil pada *beta ridge regression* dengan beberapa nilai parameter *ridge* k yang digunakan.

1.3 Manfaat

Adapun manfaat dari penelitian ini adalah dapat meningkatkan pengetahuan dan wawasan bagi penulis dan pembaca mengenai penerapan *beta ridge regression* dalam menangani masalah multikolinearitas. Serta dapat menjadi referensi bagi pembaca yang ingin melakukan penelitian dalam mengangani masalah multikolinearitas dengan menggunakan metode *beta ridge regression.*

II. TINJAUAN PUSTAKA

2.1 Distribusi Beta

Distribusi beta adalah salah satu distribusi probabilitas kontinu yang berguna dalam berbagai bidang statistik, terutama sebagai model untuk variabel acak terbatas pada interval (0,1) dan memiliki dua parameter bernilai positif, dilambangkan dengan a dan b , yang berperan sebagai eksponen variabel acak dan memengaruhi bentuk dari distribusi beta (Degroot & Schervish, 2002). Oleh karena itu, ciri utama distribusi beta adalah bentuknya dapat dikontrol oleh parameter a dan b dan distribusi beta digunakan untuk memodelkan variabel acak yang nilainya berada dalam interval (0, 1). Parameter a dan b dalam distribusi beta dikenal sebagai parameter bentuk (*shape* parameter). Secara statistik, distribusi beta memiliki penerapan yang bervariasi. Misalnya, distribusi beta digunakan untuk mempelajari persentase perubahan dalam sampel yang berbeda.

Ada beberapa notasi untuk menunjukkan bahwa variabel acak kontinu diatur oleh distribusi beta yaitu sebagai berikut:

$$X \sim B(a, b)$$

$$X \sim Beta(a, b)$$

$$X \sim \beta_{a,b}$$

Distribusi beta dikenal cukup fleksibel dalam memodelkan data dengan respon berbentuk proporsi, karena pola densitasnya dapat bervariasi yang berbeda bergantung pada nilai-nilai dari parameter yang terdapat dalam distribusi beta tersebut. Fungsi beta, dilambangkan sebagai $B(a, b)$ dan didefinisikan sebagai fungsi integral yang menggambarkan hubungan antara dua variabel yang terletak dalam interval (0,1). Berikut adalah definisi dari fungsi beta:

$$B(a, b) = \int_0^1 y^{a-1}(1-y)^{b-1} dy \, ,$$

dengan:

a dan b = Parameter positif distribusi beta,

y = Variabel acak yang nilainya natar 0 dan 1.

Fungsi beta dapat dinyatakan dalam bentuk fungsi gamma yaitu sebagai berikut:

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

dengan:

$\Gamma(n)$ = Fungsi gamma yang didefinisikan sebagai:

$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt.$$

Fungsi distribusi beta adalah fungsi kepadatan probabilitas (PDF) dari distribusi beta. Ini menggambarkan distribusi probabilitas untuk variabel acak yang terletak dalam interval (0,1) dan dinyatakan sebagai berikut:

$$f(y; a, b) = \frac{y^{a-1}(1-y)^{b-1}}{B(a, b)}.$$

Fungsi densitas dari variabel acak yang mengikuti distribusi beta dalam bentuk fungsi gamma adalah sebagai berikut:

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \, y^{a-1}(1-y)^{b-1} \, ,$$

dengan:

$\Gamma(.)$ = Fungsi gamma,
a dan b = Parameter distribusi beta,
y = Variabel acak yang nilainya antara 0 dan 1.

Dalam distribusi beta nilai rata rata dan varian dengan parameter a dan b adalah sebagai berikut:

$$E(Y) = \frac{a}{a+b} \qquad \text{dan} \qquad Var(Y) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Hal tersebut dapat dibuktikan berdasarkan penjabaran sebagai berikut:

$$\begin{aligned}
E(Y^k) &= \int_{-\infty}^{\infty} y^k\, f(y)\, dy \\
&= \int_0^1 y^k \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}\, dy \\
&= \frac{1}{B(a,b)} \int_0^1 y^{k+a-1}(1-y)^{b-1}\, dy \\
&= \frac{1}{B(a,b)} B((k+a),b) \\
&= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(k+a)\Gamma(b)}{\Gamma((k+a)+(b)} \\
&= \frac{\Gamma(a+b)\Gamma(k+a)}{\Gamma(a)\Gamma(k+a+b)}
\end{aligned}$$

$$\begin{aligned}
E(Y) &= \frac{\Gamma(a+b)\Gamma(1+a)}{\Gamma(a)\Gamma(1+a+b)} \\
&= \frac{(a+b-1)!\, a!}{(a-1)!\,(a+b)!} \\
&= \frac{(a+b-1)!\, a(a-1)!}{(a-1)!\,(a+b)(a+b-1)!} \\
&= \frac{a}{a+b}
\end{aligned}$$

$$
\begin{aligned}
E(Y^2) &= \frac{\Gamma(a+b)\Gamma(2+a)}{\Gamma(a)\Gamma(2+a+b)} \\
&= \frac{(a+b-1)!\,(a+1)!}{(a-1)!\,(a+b+1)!} \\
&= \frac{(a+b-1)!\,(a+1)(a)(a-1)!}{(a-1)!\,(a+b+1)(a+b)(a+b-1)} \\
&= \frac{a(a+1)}{(a+b)\,(a+b+1)}
\end{aligned}
$$

$$
\begin{aligned}
Var(Y) &= E(Y^2) - \big(E(Y)\big)^2 \\
&= \frac{a(a+1)}{(a+b)\,(a+b+1)} - \left(\frac{a}{a+b}\right)^2 \\
&= \frac{1}{a+b}\left(\frac{a^2+a}{a+b+1} - \frac{a^2}{a+b}\right) \\
&= \frac{1}{a+b}\left(\frac{(a^2+a)(a+b) - a^2(a+b+1)}{(a+b+1)(a+b)}\right) \\
&= \frac{1}{a+b}\left(\frac{ab}{(a+b+1)(a+b)}\right) \\
&= \frac{ab}{(a+b)^2(a+b+1)}.
\end{aligned}
$$

2.2 Regresi Beta

Regresi beta adalah sebuah metode analisis statistik yang digunakan untuk memodelkan hubungan antara variabel dependen dan independen, di mana variabel dependen mengikuti distribusi beta. Tujuan utama dari analisis regresi beta adalah untuk memprediksi nilai variabel dependen (Y) berdasarkan variabel independen (X) dan untuk mengetahui seberapa besar pengaruh variabel independen terhadap variabel dependen. Model ini termasuk dalam bagian dari model linear umum *Generalized Linear Model* (GLM) dan mengasumsikan bahwa variabel dependen mengikuti distribusi beta, yang merupakan bagian dari distribusi eksponensial. Model regresi beta pertama kali dikembangkan oleh

Ferrari dan Cribari-Neto dengan menghubungkan fungsi rata-rata variabel responsnya ke serangkaian prediktor linier melalui fungsi diferensial monoton yang disebut fungsi tautan.

Menurut Ferrari dan Cribari-Neto (2004) mendefinisikan parameterisasi untuk mengembangkan model regresi respons berdistribusi beta berdasarkan persamaan awal fungsi densitas regresi beta. Dengan memisalkan bahwa $\mu = \frac{a}{a+b}$ dan $\phi = a + b$ sehingga diperoleh bahwa $a = \mu\phi$ dan $b = \phi - \phi\mu = \phi(1 - \mu)$. Setelah dilakukan reparameterisasi tersebut, maka fungsi untuk variabel acak y yang mengikuti distribusi beta adalah sebagai berikut:

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma(\phi - \phi\mu)} \, y^{\mu\phi-1}(1 - y)^{\phi-\phi\mu-1},$$

dengan:

$0 < y < 1,\ 0 < \mu < 1,\ \phi > 0,$

$\Gamma(.)$ = Fungsi gamma

ϕ = Parameter.

Sehingga rata-rata dan varians Y didefinisikan pada parameterisasi baru yaitu:

$$E(Y) = \mu \qquad \text{dan} \qquad Var(Y) = \frac{\mu(1 - \mu)}{1 + \phi}.$$

Dengan μ adalah rata-rata dari variabel dependen dan ϕ merupakan parameter, dalam arti bahwa untuk μ tertentu, maka nilai dari ϕ lebih besar akan memberikan varians bagi variabel dependen yang lebih kecil.

Dalam regresi beta untuk menghubungkan rata-rata (μ) dengan kovariat $\boldsymbol{X_i}$ dinyatakan melalui fungsi hubung logit. Fungsi logit adalah fungsi yang mengubah proporsi $\mu \epsilon$ (0,1) ke skala bilangan riil $(-\infty, \infty)$. Adapun fungsi hubung logit berbentuk sebagai berikut:

$$g(\mu) = \log\left(\frac{\mu}{1-\mu}\right).$$

Dalam model regresi, kovariat $\boldsymbol{X_i}$ merujuk pada variabel independen yang kita gunakan untuk menjelaskan variabel dependen Y. Jika kita memiliki n observasi (data), setiap observasi i terdiri dari satu atau lebih kovariat adalah:

$$\boldsymbol{X_i} = (X_{i1}, X_{i2}, \dots, X_{ip})\,.$$

Jika kita memiliki konstanta atau *intercept* dalam model, kita biasanya menambahkan kolom 1 ke setiap $\boldsymbol{X_i}$ sebagai komponen pertama, sehingga vektor kovariat menjadi:

$$\boldsymbol{X_i} = (1, X_{i1}, X_{i2}, \dots, X_{ip})\,.$$

Model linear menghubungkan kovariat $\boldsymbol{X_i}$ dengan parameter regresi β. Adapun hubungan linear ini ditulis sebagai berikut:

$$g(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq}\,.$$

Dalam bentuk matriks ditulis sebagai berikut:

$$g(\mu_i) = \boldsymbol{X_i^T \beta}$$

Sehingga dapat ditulis model regresi beta dengan fungsi penghubung log sebagai berikut:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{X_i^T \beta} \rightarrow \mu_i = g^{-1}\left(\boldsymbol{X_i^T \beta}\right) = \frac{\exp\left(\boldsymbol{X_i^T \beta}\right)}{1+\exp(\boldsymbol{X_i^T \beta})}\,.$$

2.3 Estimasi Parameter Model Regresi Beta

Penaksiran parameter pada model regresi beta dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE), yang dilakukan dengan cara memaksimumkan fungsi *likelihood*. MLE merupakan metode untuk memperkirakan parameter distribusi dari suatu populasi dengan memilih nilai parameter yang memaksimalkan fungsi *likelihood* dari data yang diamati (Casella & Berger, 2002). Fungsi *likelihood* dalam regresi beta adalah sebagai berikut:

$$L(y;\ a, b) = \prod_{i=1}^{n} f(y_i; a, b)\ ,$$

di mana $f(y_i; \mu_i, \sigma^2)$ adalah fungsi densitas dari distribusi beta yang didefinisikan sebagai berikut:

$$f(y_i; a, b) = \frac{y_i^{a-1}(1 - y_i)^{b-1}}{B(a, b)}\ ,$$

dengan:

$$a = \mu\left(\frac{1-\sigma^2}{\sigma^2}\right)$$

$$\beta = (1 - \mu)\left(\frac{1-\sigma^2}{\sigma^2}\right)$$

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a + b)}\ .$$

Sehingga dapat ditulis sebagai berikut:

$$L(y_i; a, b) = \prod_{i=1}^{n} \frac{y_i^{a-1}(1 - y_i)^{b-1}}{B(a, b)}\ .$$

Untuk mempermudah analisis maka akan digunakan fungsi *log likelihood* sebagai berikut:

$$
\begin{aligned}
\ell(a,b) &= \log L(y_i; a,b) \\
&= \log\left(\prod_{i=1}^{n} \frac{y_i^{a-1}(1-y_i)^{b-1}}{B(a,b)}\right) \\
&= \sum_{i=1}^{n} \log\left(\frac{y_i^{a-1}(1-y_i)^{b-1}}{B(a,b)}\right) \\
&= \sum_{i=1}^{n} \left[(a-1)\log(y_i) + (b-1)\log(1-y_i) - \log\big(B(a,b)\big)\right].
\end{aligned}
$$

Substitusikan parameter a, b dan $B(a,b)$, sehingga:

$$
\begin{aligned}
\ell(\mu,\sigma^2) = & \sum_{i=1}^{n} \left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right) - 1\right)\log\ (y_i) + \left((1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right) - 1\right) \\
& \log\,(1-y_i) - \log\Gamma\left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) + \log\Gamma\left((1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) \\
& - \log\Gamma\left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right) + (1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) \\
= & \sum_{i=1}^{n} \left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right) - 1\right)\log\ (y_i) + \left((1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right) - 1\right) \\
& \log(1-y_i) - \log\Gamma\left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) + \log\Gamma\left((1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) \\
& - \log\Gamma\left(\frac{1-\sigma^2}{\sigma^2}\right) \\
= & \sum_{i=1}^{n} \log\Gamma\left(\frac{1-\sigma^2}{\sigma^2}\right) - \log\ \Gamma\left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) - \\
& \log\Gamma\left((1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right)\right) + \left(\mu_i\left(\frac{1-\sigma^2}{\sigma^2}\right) - 1\right)\ \log\ (y_i) + \\
& \left((1-\mu_i)\left(\frac{1-\sigma^2}{\sigma^2}\right) - 1\right)\log\ (1-y_i).
\end{aligned}
$$

Sehingga diperoleh fungsi *log likelihood* dari regresi beta adalah sebagai berikut:

$$\ell(\mu,\phi) = \sum_{i=1}^{n} [\, \log \Gamma(\phi) \; - \log \; \Gamma\left(\mu_i(\phi)\right) - \log \Gamma\left((1-\mu_i)(\phi)\right) +$$
$$(\mu_i(\phi) \; - 1) \log(y_i) + \left((1-\mu_i)(\phi) - 1\right) \log(1-y_i)] \, .$$

Kemudian untuk memperoleh nilai estimator *Maximum Likelihood* $\hat{\beta}$, vektor koefisien diestimasi menggunakan metode *Maximum Likelihood* dengan cara melakukan turunan pertama dari fungsi *log-likelihood* terhadap β atau dikenal dengan *score function*. Di mana μ_i tergantung pada β melalui fungsi hubung log. Adapun *score function* dapat ditulis sebagai berikut:

$$\mathcal{S}(\beta) = \frac{\partial \log \ell(\mu,\phi)}{\partial \beta}$$

$$= \frac{\partial \sum_{i=1}^{n} [\log \Gamma(\phi) \; - \log \; \Gamma\left(\mu_i(\phi)\right) - \log \Gamma\left(\left(1-\mu_i\right)(\phi)\right) + (\mu_i(\phi) \; - 1) \log\left(y_i\right) + \left(\left(1-\mu_i\right)(\phi) - 1\right) \log(1-y_i)]}{\partial \beta}$$

$$= \sum_{i=1}^{n} [-\psi\left(\mu_i(\phi)\right) \frac{\partial \mu_i(\phi)}{\partial \beta} - \psi\left((1-\mu_i)\phi\right) \frac{\partial((1-\mu_i)\phi)}{\partial \beta} + \frac{\partial}{\partial \beta}$$
$$[(\mu_i(\phi) - 1) \log(y_i)] + \frac{\partial}{\partial \beta} \left[\left((1-\mu_i)\phi - 1\right) \log(1-y_i)\right]$$

$$= \sum_{i=1}^{n} \left[\left(-\psi(\mu_i(\phi)) - \psi\left(1-\mu_i(\phi)\right) + \log(y_i) - \log\,(1-y_i)\right]\, .$$

Karena pada persamaan di atas berbentuk nonlinear dan tidak dapat diselesaikan secara analitik. Solusi $\mathcal{S}(\beta)$ sama dengan 0 dan dapat ditemukan melalui algoritma *Iterative Reweighted Least Square* (IRLS) yang berasal dari penerapan metode *Newton Raphson* atau *Fisher Scoring*. Persamaan umum *Newton-Raphson* untuk mencari akar dari *log-likelihood* adalah:

$$\beta_{m+1} = \beta_m + \{H(\beta_m)\}^{-1} \, \mathcal{S}(\beta_m)\, ,$$

dengan:

$H(\beta_m)$ = matriks Hessian (turunan kedua dari *log likelihood*)
$\mathcal{S}(\beta_m)$ = gradien (turunan pertama dari *log likelihood*).

Menurut McCullagh & Nelder (1989), gradien dan Hessian dievaluasi pada setiap iterasi m yang menghasilkan:

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\beta_m}) \approx \boldsymbol{X_m^T W_m}(\boldsymbol{z_m} - \boldsymbol{X_m \beta_m}) \text{ dan } \boldsymbol{H}(\boldsymbol{\beta_m}) \approx \boldsymbol{X_m^T W_m X_m} \, ,$$

dengan:

z_m = vektor respon yang disesuaikan pada iterasi m
W_m = matriks diagonal dari bobot.

Setelah didapatkan aproksimasi Hessian dan gradien, kemudian dimasukkan ke dalam persamaan *Newton-Raphson* sebagai berikut:

$$\boldsymbol{\beta_{m+1}} = \boldsymbol{\beta_m} + (\boldsymbol{X_m^T W_m X_m})^{-1} \, \boldsymbol{X_m^T W_m}(\boldsymbol{z_m} - \boldsymbol{X_m \beta_m}) \, .$$

Setelah mendapatkan penyederhanaan, bentuk akhir dari logaritma estimasi kemudian dikaitkan dengan *iterative reweighted least squares* sebagai berikut:

$$\boldsymbol{\beta_m} = (\boldsymbol{X_m^T W_m X_m})^{-1} \boldsymbol{X_m^T W_m z_m} \, ,$$

dengan:

$$W_m = diag\left(\frac{1+\phi}{\mu_i(1-\mu_i)}\right)$$
$$z_m = \log(\mu_i) + \frac{y_i + \mu_i}{\mu_i(1-\mu_i)} \, ,$$

Akan disimpulkan bahwa $\boldsymbol{\beta_m}$ konvergen ke $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}$ ketika $m \to \infty$, sehingga bentuk akhir dari *Maximum Likelihood Estimation* adalah sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}} = \left(\boldsymbol{X^T \widehat{W} X}\right)^{-1} \boldsymbol{X^T \widehat{W} z} \, .$$

Varians dari *Maximum Likelihood Estimation* adalah sebagai berikut:

$$\boldsymbol{V}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right) = \boldsymbol{\phi}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}.$$

Mean Squared Error dari *Maximum Likelihood Estimation* adalah sebagai berikut:

$$\boldsymbol{MSE}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right) = \boldsymbol{E}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}} - \boldsymbol{\beta}\right)^{\boldsymbol{T}}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}} - \boldsymbol{\beta}\right).$$

Karena $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}$ adalah estimator tak bias, artinya $\boldsymbol{E}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}) = \boldsymbol{\beta}$, maka MSE hanya bergantung pada varians (*trace*) yaitu sebagai berikut:

$$\begin{aligned}\boldsymbol{MSE}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right) &= trace\{\boldsymbol{\phi}(\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X})^{-\mathbf{1}}\} \\ &= \phi \sum_{j=1}^{p} \frac{1}{\lambda_j}\end{aligned}$$

Dengan λ_j adalah nilai eigen ke-j pada matriks $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}$. Matriks $\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}$ tidak terkondisikan dengan baik ketika regresor berkorelasi yang menyebabkan beberapa nilai eigen menjadi kecil, dan MSE dari MLE yang diestimasikan meningkat sehingga terjadi masalah multikolinearitas.

2.4 Multikolinearitas

Multikolinearitas pertama kali dikemukakan oleh Ragnar Frisch pada tahun 1934 yang awalnya terdapat hubungan linear antara beberapa atau semua variabel prediktor dari model regresi (Gujarati & Porter, 2009). Pada multikolinearitas terdapat korelasi yang tinggi antara dua atau lebih variabel independen dalam model regresi yang dapat mengakibatkan estimasi yang tidak stabil. Masalah multikolinearitas merupakan masalah serius dalam penelitian terapan yang menyebabkan varians tinggi, interval kepercayaan lebih luas, dan estimasi parameter yang tidak stabil (Qasim, *et al*, 2021). Model yang baik adalah model yang bebas dari multikolinearitas (Yaldi, *et al*, 2022).
Terdapat 3 cara untuk mendeteksi multikolienaritas menurut Montgomery, *et al*. (2012), diantaranya sebagai berikut:

1. Koefisien korelasi

   Dalam masalah multikolinearitas dapat diketahui melalui koefisien korelasi dengan melihat nilai koefisien korelasinya. Nilai koefisien korelasi yang mendekati 1 atau -1 menunjukkan korelasi yang sangat kuat antara dua variabel. Semakin dekat nilai koefisien korelasi ke 1 atau -1, semakin tinggi tingkat multikolinearitas.

2. *Variance Inflation Factor* (VIF)

   Salah satu cara yang paling dikenal untuk mendeteksi multikolienaritas adalah dengan melihat nilai VIF. Jika nilai VIF $>$ 10, hal ini menunjukkan adanya masalah multikolienaritas yang signifikan. Pada penelitian ini digunakan nilai VIF untuk mengindikasikan adanya multikolinearitas. Nilai VIF dapat dihitung menggunakan persamaan berikut:

$$VIF = \frac{1}{1 - R_j^2},$$

   dengan:

   R_j^2 $=$ Koefisien determinasi yang diperoleh dari variabel prediktor ketika X_j diregresikan dengan variabel prediktor lainnya

   j = 1, 2, …, k (k merupakan banyaknya variabel prediktor)

3. *Tollerance* (TOL)

   Masalah multikolienaritas juga dapat dilihat melalui nilai *Tollerance* (TOL). Jika nilai *Tollerance* (TOL) $<$ 0.10 maka terjadi multikolinearitas. Nilai TOL dapat dihitung dengan persamaan sebagai berikut:

$$TOL = (1 - R_j^2)$$

Dampak yang diperoleh jika terjadi multikonearitas adalah:

1. Estimasi parameter regresi yang dihasilkan menjadi tidak efisien. Ini berarti bahwa koefisien regresi yang diperoleh mungkin tidak akurat dan dapat menyebabkan kesalahan dalam interpretasi hasil.
2. Interval kepercayaan yang lebih lebar karena nilai *standar error* yang besar.
3. Nilai koefisien determinasi (R^2) masih dapat relative tinggi, walaupun pada nilai statistik pada uji t tidak signifikan.

2.5 *Beta Ridge Regr*ession (BRR)

*Beta ridge regr*ession (BRR) merupakan suatu modifikasi untuk menganalisis permasalahan multikolinearitas atau mengalami korelasi yang tinggi pada variabel independen dengan metode *ridge* pada model regresi beta yang merupakan generalisasi dari Hoerl dan Kennard. *Beta ridge regr*ession digunakan untuk memperkirakan parameter model regresi beta ketika variabel penjelas berkorelasi. Model regresi beta adalah model statistik yang digunakan untuk menganalisis data yang berbentuk rasio atau persentase. Masalah multikolinearitas dalam regresi *ridge* dapat diatasi dengan menambahkan suatu nilai ketetapan bias k, yang dapat mengurangi nilai varian penduga. Menurut Dereny & Rashwan (2011), metode *ridge* didasarkan pada pertambahan konstanta bias k pada diagonal matriks $\boldsymbol{X^T X}$, sehingga koefisien penduga *ridge* dipengaruhi oleh besarnya tetapan bias k.

2.6 Estimasi Parameter *Beta Ridge Regr*ession (BRR)

Pada metode *beta ridge regr*ession prinsip *maximum likelihood* digunakan untuk memperkirakan parameter estimasi dengan cara yang lebih akurat dan stabil dengan minimum jumlah kuadrat sisaan *Weight Sum of Square* (WSSE). Pada estimator $\boldsymbol{\hat{\beta}_{MLE}}$ yang diperoleh akan dipandang sebagai estimator optimal dalam

WSSE. Misalkan kita akan memilih estimator sembarang $\widehat{\boldsymbol{B}}$ selain $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}$, di mana $\widehat{\boldsymbol{B}}$ ini merupakan suatu vektor dari $\boldsymbol{\beta}$ maka dapat didefinisikan WSSE dari estimator ini adalah sebagai berikut:

$$\begin{aligned}\boldsymbol{\theta} &= \left(\boldsymbol{y}-\widehat{\boldsymbol{B}}\right)^{\boldsymbol{T}}\left(\boldsymbol{y}-\widehat{\boldsymbol{B}}\right)\\ &= \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}+\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}-\widehat{\boldsymbol{B}}\right)\right)^{\boldsymbol{T}}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}+\boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}-\widehat{\boldsymbol{B}}\right)\right)\\ &= \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)+\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)\\ &= \boldsymbol{\theta}_{\boldsymbol{min}}+\boldsymbol{\theta}\left(\widehat{\boldsymbol{B}}\right).\end{aligned}$$

Di mana $\boldsymbol{\theta}_{\boldsymbol{min}}$ adalah nilai minimum, dan $\boldsymbol{\theta}\left(\widehat{\boldsymbol{B}}\right)>0$ adalah kenaikan tetap yang meningkatkan WSSE ketika estimator $\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}$ digantikan oleh estimator $\widehat{\boldsymbol{B}}$. Menurut Hoerl & Kennard (1970), untuk mendapatkan estimasi parameter pada *beta ridge regr*ession yaitu dengan meminimumkan fungsi lagrange (Q) menggunakan $\widehat{\boldsymbol{B}}^{\boldsymbol{T}}\widehat{\boldsymbol{B}}$ dengan $\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)=\theta_0$ yang kemudian akan dibuat dalam bentuk *lagrangian* sebagai berikut:

$$\begin{aligned}\boldsymbol{Minimum}\ (Q) &= \left(\boldsymbol{y}-\widehat{\boldsymbol{B}}\right)^{\boldsymbol{T}}\left(\boldsymbol{y}-\widehat{\boldsymbol{B}}\right)\\ &= \left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\left(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)+\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)\\ &= \widehat{\boldsymbol{B}}^{\boldsymbol{t}}\boldsymbol{B}+\left(\frac{1}{k}\right)\left\{\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)-\theta_0\right\},\end{aligned}$$

dengan:

$\left(\frac{1}{k}\right)$ = *Multiple Lagrangian*

θ_0 = Jumlah kuadrat error (konstanta non-negatif)

Kemudian persamaan *lagrangian* tersebut diturunkan terhadap $\widehat{\boldsymbol{B}}$ dan hasilnya disamadengankan 0 yaitu sebagai berikut:

$$\begin{aligned}\frac{\boldsymbol{\partial Q}}{\boldsymbol{\partial}\widehat{\boldsymbol{B}}} &= \frac{\boldsymbol{\partial}}{\boldsymbol{\partial}\widehat{\boldsymbol{B}}}\left[\widehat{\boldsymbol{B}}^{\boldsymbol{T}}\boldsymbol{B}+\left(\frac{\boldsymbol{1}}{\boldsymbol{k}}\right)\left\{\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)-\theta_0\right\}\right]\\ &= \boldsymbol{2}\widehat{\boldsymbol{B}}+\frac{\left\{\boldsymbol{2X}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{X}\left(\widehat{\boldsymbol{B}}-\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\right)\right\}}{\boldsymbol{k}}=0.\end{aligned}$$

Pada persamaan di atas $\widehat{\boldsymbol{B}}$ dianggap sebagai estimator *beta ridge regr*ession **($\widehat{\boldsymbol{\beta}}_{\boldsymbol{BRR}}$)**, sehingga persamaan estimator metode *beta ridge regr*ession dapat didefinisikan sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{BRR}} = (\boldsymbol{kI} + \boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{WX})^{-\boldsymbol{1}}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{WX}\widehat{\boldsymbol{\beta}}_{\boldsymbol{MLE}}\,,$$

dengan:

k = Parameter penyusutan

I = Matriks identitas dengan orde $p \times p$.

2.7 Estimasi Parameter *Ridge k*

Pada *beta ridge regression* untuk menangani masalah multikolinearitas, diperlukan parameter *ridge* yang disebut sebagai parameter ridge k. Parameter ini digunakan untuk mengurangi varian estimasi dan mengurangi efek multikolinearitas pada model regresi. Berdasarkan karya Alkhamisi, *et al* (2006) dan Kibria (2003), kita dapat memodifikasi estimasi k mereka untuk model regresi beta yaitu sebagai berikut:

1. $k_1 = max\left(\frac{1}{\phi\hat{\alpha}_j^2}\right)$
2. $k_2 = median\left(\frac{1}{\phi\hat{\alpha}_j^2}\right)$
3. $k_3 = mean\left(\frac{1}{\phi\hat{\alpha}_j^2}\right)$

Berdasarkan Abonazel & Taha (2021) memberikan estimasi k untuk model regresi beta yaitu sebagai berikut:

4. $k_4 = \left(\frac{\lambda_{max}}{\phi\hat{\alpha}_{max}^2}\right)$

5. $k_5 = \left(\frac{\lambda_{min}}{\phi\hat{\alpha}_{min}^2}\right)$

Dengan λ_{max} dan λ_{min} nilai eigen maksimum dan minimum pada matriks $\boldsymbol{X^T W X}$, α_j^2 merupakan nilai dari $Q^T\hat{\beta}_{MLE}$, dimana Q^T merupakan elemen vektor eigen dari matriks $\boldsymbol{X^T W X}$.

2.8 Performa Model

Salah satu tujuan dalam analisis regresi adalah untuk menemukan model yang paling tepat dalam menjelaskan hubungan antara variabel dependen (variabel yang ingin diprediksi) dan variabel independen (variabel prediktor). Model yang baik tidak hanya memiliki kemampuan prediksi yang tinggi, tetapi juga sederhana dan mudah diinterpretasikan. Model yang baik adalah model yang seluruh koefisien regresinya signifikan dan memiliki kriteria model terbaik optimum. Untuk melihat model yang baik dapat dilihat dari kriteria *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE).

2.8.1 *Mean Squared Error* (MSE)

Menurut James, *et al* (2013), *Mean Squared Error* (MSE) adalah ukuran kesalahan antara nilai aktual dan nilai prediksi dengan cara menghitung rata-rata dari kuadrat selisih antara keduanya. Pemilihan model terbaik salah satunya yaitu berdasarkan MSE. Semakin rendah nilai MSE, maka semakin baik model yang digunakan. MSE mengukur rata-rata kuadrat kesalahan antara nilai yang diramalkan dengan nilai aktual. Jika nilai MSE semakin kecil, artinya nilai prediksi semakin mendekati nilai sebenarnya yang menunjukkan kualitas yang lebih baik dalam melakukan estimasi. Nilai MSE dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$MSE = \frac{1}{L}\sum_{i=1}^{L}\left(\widehat{\beta_\iota} - \beta\right)^2$$

dengan:

$\widehat{\beta_\iota}$ = Penduga parameter regresi

β_i = Parameter regresi

L = Banyaknya ulangan

2.8.2 *Mean Absolute Error* (MAE)

Menurut Hyndman & Athanasopoulos (2018), *Mean Absolute Error* (MAE) ukuran rata-rata absolut dari selisih antara nilai aktual dan nilai prediksi. MAE mengukur seberapa besar kesalahan prediksi yang dihasilkan oleh model tanpa memperhatikan arah kesalahan tersebut. Dalam memilih model terbaik, salah satu kriterianya adalah berdasarkan nilai MAE, di mana semakin kecil nilai MAE, semakin baik model dalam memprediksi data. MAE tidak mengkuadratkan perbedaan, sehingga tidak memberikan penalti yang lebih besar pada error besar seperti yang terjadi pada MSE. Dengan demikian, MAE lebih mudah diinterpretasikan dalam konteks tertentu, terutama ketika distribusi error yang lebih ekstrem tidak diinginkan untuk memberikan dampak besar pada evaluasi performa model. Nilai MAE dapat dihitung dengan menggunakan persamaan sebagai berikut:

$$MAE = \frac{1}{L}\sum_{i=1}^{L}\left|\widehat{\beta_\iota} - \beta\right|$$

dengan:

$\widehat{\beta_\iota}$ = Penduga parameter regresi

β_i = Parameter regresi

L = Banyaknya ulangan

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2024/2025 di Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan dalam penelitian ini adalah data simulasi yang mengandung multikolinearitas yang dihasilkan dengan menggunakan *Software RStudio*. Adapun data simulasinya yakni data yang dibangkitkan dengan variabel bebas sebanyak 6 (p = 6) dengan tingkatan korelasi antar variabel bebas sebesar 0.6 dan 0.99 ($\rho = 0.6, 0.99$) dan jumlah sampel yang digunakan adalah $n =$ 10, 20, 30, 50, dan 75 dengan pengulangan sebanyak 1000 kali dan parameter ϕ yang digunakan adalah ($\phi = 1, 2, 3$), ditentukan 3 nilai parameter ϕ berbeda untuk melihat pengaruh pada metode penduga.

Untuk mendapatkan data multikolinearitas pada setiap himpunan data X_p dibangkitkan dengan menggunakan simulasi Monte Carlo berdasarkan McDonald dan Galarneau (1975) dengan persamaan sebagai berikut:

$$X_p = \sqrt{1-\rho^2}z_{ij} + \rho z_{i(p+1)}$$

dengan:

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

$j = 1, 2, \dots, p$

$z_{ij} \sim Normal\ (0,1);\ \rho = 0.6$ dan 0.99

Untuk membangkitkan data y yang mengikuti distribusi beta, kita perlu menggunakan distribusi beta yang memiliki dua parameter *shape*, seringkali disebut sebagai a dan b, yang mengontrol bentuk distribusi. Untuk membangkitkan data y yang mengikuti distribusi beta dalam simulasi, kita dapat menggunakan rumus sebagai berikut:

1. Tentukan nilai μ_i (ekspektasi y) sebagai fungsi dari variabel prediktor X_i.

$$\mu_i = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_q)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_q)}$$

   dengan menetapkan nilai $\beta_0 = 0$ dan $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 1$.

2. Tentukan parameter ϕ yang mengontrol variabilitas y disekitar μ_i.
3. Hitung parameter shape a dan b dari distribusi beta dengan rumus:

$$\text{Shape } a_i = \mu_i \phi$$

$$\text{Shape } b_i = (1 - \mu_i)\phi$$

4. Bangkitan data y_i dari distribusi beta:

$$y_i \sim Beta(a_i, b_i)$$

3.3 Metode Penelitian

Penelitian ini dilakukan secara studi pustaka, yaitu dengan mempelajari buku-buku teks penunjang dan karya ilmiah yang disajikan dalam bentuk jurnal. Dalam menyelesaikan penelitian ini, penulis menggunakan *Software Rstudio* versi 4.4.1 untuk mempermudah perhitungan dan mendapatkan hasil yang akurat. Adapun langkah-langkah yang dilakukan adalah sebagai berikut:

1. Membangkitkan data dengan n = 10, 20, 30, 50 dan 75.
2. Mengidentifikasi multikolinearitas dengan melihat nilai VIF pada data.
3. Menduga model regresi beta dengan metode MLE.
4. Mencari beberapa bentuk nilai parameter ridge k pada *beta ridge regression* yang digunakan dalam penelitian.
5. Merealisasikan nilai k yang telah didapatkan untuk mendapatkan *beta ridge regression estimator*.
6. Menghitung nilai MSE dan MAE pada beberapa bentuk parameter k yang digunakan pada *beta ridge regression estimator*.
7. Mengulangi langkah 1-6 sebanyak 1000 kali.
8. Merumuskan kesimpulan berdasarkan hasil analisis yang telah dilakukan.

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

1. Berdasarkan hasil penelitian menunjukkan bahwa performa *beta ridge regression estimator* dalam mengatasi masalah multikolinearitas yaitu dibuktikan dengan nilai MSE dan MAE pada kelima parameter *ridge k* $(k_1, k_2, k_3, k_4, k_5)$ lebih kecil dibandingkan MLE untuk semua parameter $\phi = 1, 2, 3$ dan pada tingkat korelasi rendah maupun tinggi.
2. Pada $n = 10$ dan 20, performa terbaik selalu terdapat pada parameter *ridge* k_4 dalam kondisi apapun. Pada $n = 30$ dan 50, performa terbaik bisa terjadi pada parameter *ridge* k_4 atau k_5, tergantung pada data. Sedangkan pada $n = 75$, performa terbaik selalu tercapai pada parameter *ridge* k_5.

DAFTAR PUSTAKA

Abonazel, M. R. & Taha, I. M. 2021. Beta ridge regression estimators: simulation and application. *Communications in Statistics-Simulation and Computation*. **52**(9): 4280-4292.

Alkhamisi, M., Khalaf, G. & Shukur, G. 2006. Some modifications for choosing ridge parameters. *Communications in Statistics-Theory and Methods*. **35**(11): 2005-2020.

Bayer, F. M. & Cribari-Neto, F. 2017. Model Selection criteria in beta regression with varying dispersion. *Communications in Statistics-Simulation and Computation.* **46**(1): 729-746.

Casella, G & Berger, R. L. 2002. *Statistical Inference.* 2nd ed. Duxbury Press, Pacific Grove.

Cribari-Neto, F. & Zeilis, A. 2010. Beta Regression in R. *Journal of Statistical Software.* **34**(2): 1-24.

Cribari-Neto, F. & Vasconcellos, K. L. 2002. Nearly unbiased maximum likelihood estimation for the beta distribution. *Journal of Statistical Computation and Simulation*. **72**(2): 107-118.

Degroot, M. H. & Schervish, M. J. 2002. *Probability and Statistics.* Addison-Wesley, Boston.

Dereny, M. & Rashwan, N. I. 2011. Solving multicollinearity problem using ridge regression models. *International Journal of Contemporary Mathematical Sciences*. **6**(12): 585-600.

Ferrari, S. & Cribari-Neto, F. 2004. Beta regression for modelling rates and proportions. *Journal of applied statistics*. **31**(7): 799-815.

Ghozali, I. 2016. *Aplikasi analisis multivariate dengan program SPSS.* Badan Penerbit Universitas DIponegoro, Semarang.

Gujarati, D. N. & Porter, D. C. 2009. *Basic econometrics*. McGraw-Hill Irwin, New York.

Hajarisman, N. 2012. Penaksiran Parameter Model Regresi Beta untuk Memodelkan Data Proporsi. *Statistika.* **12**(1): 9-18.

Hoerl, A. E. & Kennard, R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*. **12** (1): 55-67.

Hyndman, R. J. & Athanasopoulos, G. 2018. *Forecasting: Principles and Practice*. OTexts, Melbourne.

James, G., Witten, D., Hestin, T. & Tibshirani, R. 2013. *An Introduction to Statistical Learning: with Applications in R.* Springer, New York.

Kibria, B. G. 2003. Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*. **32**(2): 419-435.

McCullagh, P & Nelder, J. A. 1989. *Generalized linear models.* 2nd ed. CRC Press, New York.

McDonald, G. C., & Galarneau, D. I. 1975. A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association* **70**(350): 407-416.

Montgomery, D. C. & Peck, E. A. 1992. *Introduction to Linear Regression Analysis*. 2nd ed. Wiley, New York.

Montgomery, D. C., Peck, E. A. & Vining, G. G. 2012. *Introduction to linear regression analysis*. John Wiley & Sons, Amerika Serikat.

Qasim, M., Mansson, K. & Kibria, B. G. 2021. On some beta ridge regression estimators: method, simulation and application. *Journal of Statistical Computation and Simulation*. **91**(9): 1699-1712.

Swearingen, C. J., Tilley, B. C., Adams, R. J., Rumboldt, Z., Nicholas, J. S., Bandyopadhyay, D. & Woolson, R. F. 2011. Application of beta regression to analyze ischemic stroke volume in NINDS rt-PA clinical trials. *Neuroepidemiology*. **37**(2): 73-82.

Yaldi, E., Pasaribu, J. P. K., Suratno, E., Kadar, M., Gunardi, G., Naibaho, R. & Aryati, V. A. 2022. Penerapan uji multikolinearitas dalam penelitian manajemen sumber daya manusia. *Jurnal Ilmiah Manajemen Dan Kewirausahaan.* **1**(2): 94-102.