PERFORMA ESTIMATOR LIU UNTUK MENGATASI MULTIKOLINEARITAS PADA ANALISIS REGRESI LOGISTIK

(Skripsi)

Oleh

EVA SELVIANA
1817031075

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

ABSTRACT

LIU ESTIMATOR PERFORMANCE TO OVERCOME MULTICOLLINEARITY IN LOGISTIC REGRESSION ANALYSIS

By

Eva Selviana

Logistic regression analysis is a statistical method for analyzing the relationship between the dichotomous dependent variable and several independent variables. In several logistic regression studies, problems are often found where the data contains multicollinearity. The MLE (*Maximum Likelihood Estimation*) method is a method often used in logistic regression. However, the MLE method will not be stable when the data contains multicollinearity. One method to overcome multicollinearity is the Liu estimator method proposed by Liu Kejian (1993). The advantage of the Liu method compared to other methods is that it has a constant value which is a linear function, so that the Liu method is easy to determine. This study aims to compare the performance of the Liu estimator and the MLE method in overcoming multicollinearity through Monte Carlo simulations with n = 50, 75, and 100 as well as 4 independent variables with a multicollinearity level of 0.99 in 100 repetitions. This study gives the result that Liu's method is considered capable of handling multicollinearity better than the MLE method, because it produces smaller MSE values and provides the conclusion that the smaller the SE value obtained will lead to better estimation of data parameters that contain multicollinearity.

Keywords: Liu Estimator, MLE, Multicollinearity, Monte Carlo Simulation, Logistic Regression

ABSTRAK

PERFORMA ESTIMATOR LIU UNTUK MENGATASI MULTIKOLINEARITAS PADA ANALISIS REGRESI LOGISTIK

Oleh

Eva Selviana

Analisis regresi logistik merupakan metode statistika untuk menganalisis hubungan antara variabel dependen yang bersifat dikotomik dan beberapa variabel independen. Dalam beberapa penelitian regresi logistik, sering ditemukan permasalahan dimana data mengandung multikolinearitas. Metode MLE (*Maximum Likelihood Estimation*) merupakan metode yang sering digunakan pada regresi logistik. Namun, metode MLE tidak akan stabil digunakan ketika data yang diteliti mengandung multikolinearitas. Salah satu metode untuk mengatasi multikolinearitas adalah metode Liu yang diusulkan oleh Liu Kejian (1993). Kelebihan metode Liu dibanding metode lainnya yaitu memiliki nilai tetapan yang merupakan fungsi linear, sehingga metode Liu mudah ditentukan. Penelitian ini bertujuan untuk mengetahui perbandingan performa estimator Liu dan metode MLE dalam mengatasi multikolinearitas melalui simulasi Monte Carlo dengan n = 50, 75, dan 100 serta 4 variabel independen dengan tingkat multikolinearitas sebesar 0,99 pada pengulangan sebanyak 100 kali. Penelitian ini memberikan hasil bahwa metode Liu dinilai mampu mengatasi multikolinearitas lebih baik dibanding metode MLE, karena menghasilkan nilai MSE yang lebih kecil serta memberikan kesimpulan bahwa semakin kecil nilai SE yang diperoleh akan menyebabkan pendugaan parameter data yang mengandung multikolinearitas semakin baik.

Kata Kunci: Estimator Liu, MLE, Multikolinearitas, simulasi Monte Carlo, Regresi Logistik

PERFORMA ESTIMATOR LIU UNTUK MENGATASI MULTIKOLINEARITAS PADA ANALISIS REGRESI LOGISTIK

Oleh

EVA SELVIANA

Skripsi

Sebagai Salah Satu Syarat Untuk Mencapai Gelar
SARJANA MATEMATIKA

Pada

Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2023

Judul Skripsi : **PERFORMA ESTIMATOR LIU UNTUK MENGATASI MULTIKOLINEARITAS PADA ANALISIS REGRESI LOGISTIK**

Nama Mahasiswa : **Eva Selviana**

Nomor Pokok Mahasiswa : **1817031075**

Jurusan : **Matematika**

Fakultas : **Matematika dan Ilmu Pengetahuan Alam**

MENYETUJUI

1. Komisi Pembimbing

Dr. Ir. Netti Herawati, M.Sc.
NIP 19650125 199003 2 001

Subian Saidi, S.Si., M.Si.
NIP 1980082 1200812 1 001

2. Ketua Jurusan Matematika

Dr. Aang Nuryaman, S.Si.,M.Si.
NIP. 19740316 200501 1 001

MENGESAHKAN

1. Tim Penguji

Ketua : **Dr. Ir. Netti Herawati, M.Sc.** ....

Sekretaris : **Subian Saidi, S.Si., 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 19711001 200501 1 002

Tanggal Lulus Ujian Skripsi : **23 Mei 2023**

PERNYATAAN SKRIPSI MAHASISWA

Yang bertanda tangan di bawah ini :

Nama : **Eva Selviana**

Nomor Pokok Mahasiswa : **1817031075**

Jurusan : **Matematika**

Judul Skripsi : **PERFORMA ESTIMATOR LIU UNTUK MENGATASI MULTIKOLINEARITAS PADA ANALISIS REGRESI LOGISTIK**

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

Bandar Lampung, 16 Mei 2023

Yang menyatakan,

METERAI TEMPEL

6D4ABAKX458286732

Eva Selviana

NPM. 1817031075

RIWAYAT HIDUP

Penulis yang bernama lengkap Eva Selviana, lahir di Sungai Langka pada 04 April 2000. Penulis lahir dari pasangan Bapak Subur Atmoko dan Ibu Supiyah dan merupakan anak pertama dari dua bersaudara. Penulis saat ini bertempat tinggal di Desa Sungai Langka, Kecamatan Gedongtataan, Kabupaten Pesawaran, Lampung.

Penulis menempuh pendidikan taman kanak-kanak di TK Dharma Wanita Sungai Langka tahun 2006-2007, kemudian melanjutkan sekolah dasar di SDN 5 Sungai Langka tahun 2007-2012. Penulis lalu menempuh jenjang sekolah menengah pertama di SMPN 1 Gedongtataan tahun 2013-2015 dan selanjutnya ke jenjang sekolah menengah atas di SMAN 1 Gedongtataan tahun 2015-2018.

Pada tahun 2018 penulis terdaftar sebagai Mahasiswa Program Studi S1 Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Selama menjadi mahasiswa, penulis aktif mengikuti organisasi diantaranya menjadi pengurus UKMF Rois FMIPA Universitas Lampung sebagai anggota bidang kaderisasi periode 2018, Bendahara BSO BBQ periode 2019, dan Wakil Ketua Umum periode 2020, pengurus BEM FMIPA Universitas Lampung sebagai staf Dinas Kajian dan Aksi Strategis (Kastrat) periode 2019 dan Plt. Wakil Bendahara Eksekutif periode 2019, serta UKMU Birohmah Universitas Lampung sebagai Wakil Ketua Umum periode 2021.

Pada bulan Februari-Maret 2021, penulis melaksanakan Kerja Praktik di BPS Kabupaten Pesawaran sebagai bentuk penerapan ilmu yang telah diperoleh selama kuliah. Pada bulan Agustus-September 2021, penulis melaksanakan Kuliah Kerja

Nyata (KKN) di Desa Wiyono, Kecamatan Gedongtataan, Pesawaran sebagai bentuk pengabdian mahasiswa dan menjalankan Tri Dharma Perguruan Tinggi. Pada bulan Februari-Oktober 2021, penulis beserta tim mengikuti serangkaian proses PKM (Pekan Kreativitas Mahasiswa) bidang GFK dan berhasil sampai pada PIMNAS ke-33.

KATA INSPIRASI

"Wahai orang-orang yang beriman! Jika kamu menolong (agama) Allah, niscaya Dia akan menolongmu dan meneguhkan kedudukanmu"
(Q.S Muhammad: 7)

"Berangkatlah kamu baik dengan rasa ringan maupun dengan rasa berat, dan berjihadlah dengan harta dan jiwamu di jalan Allah. Yang demikian itu adalah lebih baik bagimu jika kamu mengetahui"
(Q.S At-Taubah: 41)

Jika kamu memulai sesuatu karena Allah, maka jangan berhenti karena manusia

I'll fight till the end and never give up
Keep holding on, in fact your enemy is not only yourself, but the age of your parents

Matematika memberi kita sebuah pelajaran, bahwa untuk mendapatkan hasil yang baik haruslah melalui proses yang baik pula. Matematika adalah ibu dari sains. Ia pasti namun banyak jalannya.

Segalanya cukup, asal Allah ridho

PERSEMBAHAN

Alhamdulillah, puji syukur kepada Allah SWT. yang telah memberikan petunjuk dan rahmat-Nya juga memberikan penerangan dalam ilmu pengetahuan. Hanya karena-Nya lah skripsi ini bisa penulis selesaikan dengan rasa syukur dan bahagia. Dengan segala kerendahan hati, penulis persembahkan karya sederhana ini kepada:

Orang Tua dan Adik Tercinta

Yang telah memberikan dukungan, doa, serta kasih sayangnya sehingga menjadi sumber kekuatan penulis untuk bertahan dan berjuang menyelesaikan skripsi ini. Sebuah perjuangan yang panjang dan penuh cerita yang tak lepas dari doa tulus Mamak, Bapak, dan Iqbal

Keluarga Besar Aktivis Dakwah Kampus angkatan 2018

Yang senantiasa menguatkan, mengingatkan, dan memberikan kenangan indah selama mengemban amanah sebagai sebenar-benarnya mahasiswa

Dosen Pembimbing dan Pembahas

Yang senantiasa memberikan bimbingan, arahan, dan ilmu yang bermanfaat bagi penulis.

Almamaterku Tercinta, Universitas Lampung

SANWACANA

Segala puji bagi Allah, Tuhan semesta alam, atas limpahan karunia dan rahmat-Nya lah, penulis dapat menyelesaikan skripsi yang berjudul “Performa Estimator Liu untuk Mengatasi Multikolinearitas pada Analisis Regresi Logistik”.

Penulis menyadari bahwa dalam penyelesaian skripsi ini, tak lepas dari dukungan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ibu Dr. Ir. Netti Herawati, M.Sc., selaku dosen pembimbing akademik sekaligus pembimbing satu atas bimbingan, nasihat, arahan, dan motivasi yang membangun selama penulis menjalani perkuliahan.
2. Bapak Subian Saidi, S.Si., M.Si., selaku pembimbing dua yang telah memberikan bimbingan serta kemudahan dalam penyusunan skripsi ini.
3. Ibu Dr. Khoirin Nisa, S.Si., M.Si., selaku dosen penguji yang telah memberikan evaluasi dan saran bagi perbaikan skripsi penulis.
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. Seluruh dosen dan staf Jurusan Metematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.
7. Orang tua dan adik tercinta, Mamak, Bapak, dan Iqbal yang telah dengan sabar memberikan segala dukungan dan kepercayaan selama ini.
8. Keluarga besar Mbah Mujirah dan alm. Mbah Sukari yang senantiasa mendoakan dan memberikan semangat.
9. Mba Nilam, Mas Egi, Anis, Irvan, dan sepupu junior yang membawa nuansa

bahagia dalam perjalanan meraih ilmu.

10. Sahabatku, Putri Septiarini yang dengan baiknya berkenan membersamai lika-liku perjalanan hidup penulis sejak SMP hingga sekarang.
11. Bang Agung dan Yunda Yunna, yang dengan sabar membantu penulis saat sedang kesulitan menyelesaikan skripsi.
12. Rendi, Sherli, Ramona, Ajeng, Shofiyyah, Ridho, teman seperbimbingan yang telah saling membantu dan berbagi suka duka selama menyelesaikan perkuliahan ini.
13. Pimpinan dan keluarga besar Rois FMIPA Unila periode 2020 Kabinet Iltizam.
14. Pimpinan dan keluarga besar Birohmah Unila periode 2021 Kabinet Bianglala.
15. Teman-teman Jurusan Matematika angkatan 2018.
16. Ustadzah Adel selaku Waka asrama, Mba Mila, Fitria, dan keluarga besar SMAIT Permata Bunda yang telah memberi kesempatan untuk belajar dan berkembang secara profesional.
17. Semua pihak yang tidak bisa penulis sebutkan satu persatu yang telah membantu penulis dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, akan tetapi penulis berharap semoga skripsi ini dapat bermanfaat bagi kita semua.

Bandar Lampung, 16 Mei 2023
Penulis,

Eva Selviana

DAFTAR ISI

DAFTAR TABEL

Tabel Halaman

DAFTAR GAMBAR

Gambar Halaman

I. PENDAHULUAN

1.1 Latar Belakang dan Masalah

Analisis regresi merupakan salah satu metode statistika yang banyak digunakan untuk mengetahui hubungan antara satu atau lebih variabel bebas terhadap variabel tak bebas. Umumnya, data yang digunakan pada analisis regresi adalah data kontinu berdistribusi normal. Namun, dalam beberapa penelitian, variabel tak bebas (dependen) yang diteliti menggunakan data kategorik, yang menyatakan kejadian sukses dan gagal. Dalam analisis regresi, model yang digunakan dalam menganalisis hubungan antara variabel bebas dengan variabel tak bebas yang menggunakan data bersifat dikotomik ialah model regresi logistik. Menurut Hosmer & Lemeshow (2000), analisis regresi logistik biner digunakan untuk menjelaskan hubungan antara variabel tak bebas berupa data dikotomik/biner dengan variabel bebas yang berupa data kategorik atau kontinu.

Dalam kasus regresi logistik, apabila variabel tak bebas memiliki sifat kategorik, maka tidak dapat menggunakan metode OLS. Sehingga salah satu metode yang digunakan pada regresi logistik adalah metode *Maximum Likelihood Estimator* (MLE). Metode MLE berfungsi menduga parameter model pada regresi logistik dengan memaksimalkan fungsi *likelihood*. Untuk melakukan pendugaan model, agar penaksiran parameter dalam model bersifat BLUE (*Best Linear Unbiased Estimator*), maka salah satu asumsi yang harus terpenuhi adalah asumsi multikolinearitas (Suliadi, 2017). Namun, dalam beberapa penelitian regresi logistik, sering ditemukan permasalahan dimana data yang digunakan mengandung multikolinearitas.

Adanya multikolinearitas akan menyebabkan estimasi parameter regresi menjadi tidak akurat yang mengakibatkan nilai rata-rata kuadrat eror menjadi besar. Menurut Sembiring (1995), terjadinya multikolinearitas diantara variabel-variabel bebas dapat mengakibatkan konsekuensi penting bagi penafsiran dan penggunaan model regresi dugaan, karena dapat menyebabkan tanda dari koefisien regresi menjadi salah atau keputusan menjadi tidak signifikan.

Metode MLE yang digunakan untuk mengestimasi parameter pada regresi logistik tidak akan stabil digunakan ketika data yang diteliti mengandung multikolinearitas. Beberapa metode dikembangkan untuk dapat mengatasi permasalahan multikolinearitas, diantaranya yaitu metode *Ridge, Stein*, LASSO, *Electric Net,* dan Liu. Pada metode Stein yang ditemukan oleh James & Stein (1961), nilai tetapan yang digunakan merupakan fungsi linear, namun nilai variansi yang didapat besar. Hoerl & Kennard (1970) memperkenalkan metode Ridge dimana analisis regresi ridge dianggap memiliki sifat optimal dalam menangani kasus multikolinearitas. Liu (1993) menyarankan penduga lain yang merupakan kombinasi dari metode Stein dan metode Ridge, dimana metode tersebut memiliki sifat optimal seperti metode Ridge. Namun kelebihan metode Liu dibandingkan metode Ridge yaitu nilai tetapan pada metode Liu merupakan fungsi linear sehingga metode Liu mudah ditentukan (Qasim, dkk., 2019).

Berdasar pada penjelasan di atas, performa metode Liu akan diteliti menggunakan data simulasi pada model regresi logistik yang mengandung multikolinearitas. Kemudian akan dibandingkan dengan metode *Maximum Likelihood Estimation* (MLE) berdasarkan nilai MSE (*Mean Square Error*) dari masing-masing metode.

1.2 Tujuan Penelitian

Adapun tujuan dari penelitian tersebut adalah sebagai berikut:

1. Mengetahui performa dari metode estimator Liu untuk menangani masalah multikolinearitas pada regresi logistik.
2. Membandingkan metode estimator Liu dengan metode *Maximum Likelihood Estimation* (MLE) dan menentukan metode terbaik dalam menangani multikolinearitas dengan menggunakan kriteria nilai MSE.

1.3 Manfaat Penelitian

Manfaat dari penelitian ini adalah memberikan informasi dan masukan kepada peneliti maupun pembaca mengenai performa metode Liu untuk menganalisis data yang mengandung multikolinearitas pada regresi logistik serta menambah pengetahuan bagi penulis.

II. TINJAUAN PUSTAKA

2.1 Analisis Regresi Linear

Analisis regresi linear merupakan hubungan fungsional yang terjadi antara 2 variabel, variabel bebas (variabel yang nilainya dapat ditentukan atau diamati) dan variabel tak bebas (variabel yang nilainya ditentukan oleh variabel bebas) (Drapper & Smith, 1998). Analisis regresi linier berganda merupakan suatu algoritma yang digunakan untuk menelusuri pola hubungan antara variabel terikat dengan dua atau lebih variabel bebas (Uyanik & Guler, 2013). Dalam analisis regresi suatu persamaan regresi hendak ditentukan dan digunakan untuk menggambarkan pola atau bentuk fungsi hubungan yang terdapat antarvariabel (Montgomery & Peck, 1992).

Secara umum persamaan regresi linear berganda dengan i variabel prediktor dinyatakan dengan:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_i x_j + \varepsilon_i \tag{2.1}$$

dengan:

y_i = nilai variabel bebas pengamatan ke-*i*

β_0 = kontansta parameter

β_i = kemiringan atau *slope*

x_j = variabel bebas ke-*j*

ε_i = galat pengamatan ke- *i*

2.2 Multikolinearitas

Multikolinearitas adalah kondisi dimana terdapat hubungan linear atau korelasi yang tinggi antara satu variabel prediktor dengan variabel prediktor yang lain. Dalam model regresi, korelasi antarvariabel prediktor menyebabkan dugaan parameter regresi akan memiliki galat yang besar. Multikolinearitas terjadi bila terdapat dua atau lebih peubah bebas yang saling berkaitan (Sembiring, 2003).

Pada analisis regresi, model yang baik adalah model yang bebas dari multikolinearitas. Menurut Montgomery & Runger (2011), multikolinearitas dapat dideteksi menggunakan nilai faktor inflasi ragam atau *Variance Inflation Factor* (VIF). VIF digunakan sebagai kriteria untuk mendeteksi multikolinearitas pada regresi linier yang melibatkan lebih dari dua variabel bebas. Nilai VIF dapat dicari menggunakan rumus sebagai berikut:

$$VIF_j = \frac{1}{1-R_j^2}\ ,\ j = 1,2,\dots,n \qquad (2.2)$$

R_j^2 adalah koefisien determinasi yang dihasilkan apabila X_j dijadikan sebagai variabel respon dan diregresikan dengan variabel prediktor lainnya.

2.3 Analisis Regresi Logistik

Analisis regresi logistik merupakan analisis regresi dimana variabel tak bebas memiliki sifat biner atau dikotomus dengan satu atau lebih variabel bebas (Hosmer & Lemeshow, 2000). Variabel dikotomus atau biner adalah variabel yang hanya mempunyai dua kategori yaitu 0 dan 1. Variabel tak bebas disimbolkan dengan y. Oleh karena bersifat 2 kategori, misalkan kategori yang menyatakan kejadian sukses y = 1 dan kategori yang menyatakan kejadian gagal y = 0. Menurut Agresti (2002), variabel Y tersebut mengikuti distribusi bernoulli. Fungsi probabilitas untuk Y dengan parameter (x) adalah:

$$f(y) = \pi(x)^y \, (1 - \pi(x))^{1-y} \quad , y = 0,1 \tag{2.3}$$

Probabilitas variabel Y untuk nilai x yang diberikan, dinotasikan sebagai $\pi(x)$. Menurut Hosmer & Lemeshow (2000), model regresi logistik yang melibatkan p sebagai variabel prediktor yaitu:

$$\pi(x) = \frac{\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+\exp(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+ \beta_p x_p)} \tag{2.4}$$

Fungsi regresi logistik pada persamaan (2.4) menunjukkan bahwa hubungan antara variabel bebas dan probabilitas tidak linear, sehingga untuk memudahkan pendugaan parameter regresi dan menjadikan hubungan variabel menjadi linear dapat dilakukan transformasi yang biasa disebut transformasi logistik. Bentuk logistik dari ($\boldsymbol{x}$) dinyatakan dengan (x) sebagai berikut:

$$g(x) = ln\left(\frac{\pi(x)}{1- \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \tag{2.5}$$

2.4 *Maximum Likelihood Estimation*

Maximum Likelihood Estimation (MLE) merupakan metode yang digunakan untuk mengestimasi parameter dalam regresi logistik. Dalam pemaksimalan fungsi *likelihood*, metode MLE memberikan parameter penduga β dan mengharuskan data mengikuti persebaran tertentu. Ketika y_i menyebar binomial, dari persamaan (2.5), maka didapatkan fungsi *likelihood* sebagai berikut:

$$\begin{aligned} L(\beta|y) &= \prod_{i=1}^{n} f(y|\beta) \\ &= \prod_{i=1}^{N} \pi_i^{y_i} \, (1-\pi_i)^{1-y_i} \\ &= (\pi_i)^{\sum y_i}(1-\pi_i)^{\sum 1-y_i} \end{aligned} \tag{2.6}$$

Selanjutnya dilakukan derivatif pada fungsi *log likelihood* agar lebih mudah. Persamaan (2.6) menjadi:

$$\ell(\theta|y_i) = \log\,[(\pi_i)^{\sum y_i}(1-\pi_i)^{\sum 1-y_i}]$$
$$= \sum_{i=1}^{n}\,[y_i \log(\pi_i) + (1-y_i)\log(1-\pi_i)]$$
$$= \sum_{i=1}^{n}\,[y_i \log(\pi_i) + \log(1-\pi_i) - y_i \log\,(1-\pi_i)]$$
$$= \sum_{i=1}^{n}\,[\log(1-\pi_i) + y_i\, log\,\frac{\pi_i}{(1-\pi_i)}]$$
$$= \sum_{i=1}^{n}\,\log(1-\pi_i) + \sum_{i=1}^{n}\, y_i\, log\,\frac{\pi_i}{(1-\pi_i)}$$
$$=\sum_{i=1}^{n}\log\left(1-\frac{\exp(\sum_{j=1}^{p}\beta_j x_{ij})}{1+\exp\,(\sum_{j=1}^{p}\beta_j x_{ij})}\right)+\sum_{i=1}^{n} y_i\, log\,\frac{\left(\frac{\exp(\sum_{j=1}^{p}\beta_j x_{ij})}{1+\exp\,(\sum_{j=1}^{p}\beta_j x_{ij})}\right)}{\left(1-\frac{\exp(\sum_{j=1}^{p}\beta_j x_{ij})}{1+\exp\,(\sum_{j=1}^{p}\beta_j x_{ij})}\right)}$$
$$= \sum_{i=1}^{n} log\left(\frac{1}{1+\exp\,(\sum_{j=1}^{p}\beta_j x_{ij})}\right) + \sum_{i=1}^{n} y_i \log\,\left(\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)\right)$$
$$= \sum_{i=1}^{n} -\log\,\left(1+\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)\right) + \sum_{i=1}^{n} y_i \log\,\left(\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)\right) \qquad (2.7)$$

Selanjutnya, untuk mendapatkan nilai kritis dari persamaan (2.7) dan memaksimalkan fungsi *log-likehood*, maka turunan pertama dan kedua persamaan (2.7) terhadap nilai β_j dapat dilakukan.

$$\frac{\partial l}{\partial \beta_j} = -\sum_{i=1}^{n}\frac{1}{1+exp\left(\sum_{j=1}^{p}\beta_i x_{ij}\right)}\left(\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)x_i\right) + \sum_{i=1}^{n} y_i x_{ij}$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{p}\pi_i x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{p} y_i x_{ij}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{p}(y_i\pi_i)x_{ij} \qquad (2.8)$$

$$\frac{\partial^2 l(\beta_j)}{\partial {\beta_j}^2} = -\sum_{i=1}^{n}\frac{x_{ij}^2\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right) + \left(1+exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right) - \left(x_{ij}\,exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)\right)^2\right)}{\left(1+exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)\right)^2}$$

$$= -\sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2\pi_i - x_{ij}^2{\pi_i}^2$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2\pi_i(1-\pi_i) \qquad (2.9)$$

Selanjutnya diketahui diagonal matriks berbobot W dengan elemen ke-I, sebagai berikut:

$$W_i = \left(Var(y_i)\left(\frac{\partial\eta_i}{\partial\pi_i}\right)^2\right)^{-1} \qquad (2.10)$$

untuk baris ke – i yang merupakan vektor *m* yaitu:

$$m_i = \left(\frac{\partial\eta_i}{\partial\mu_i}\right)(y_i - \pi_i) \qquad (2.11)$$

Pada regresi logistik menggunakan fungsi hubung yaitu $\eta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$, sehingga $\frac{\partial\eta_i}{\partial\pi_i} = \frac{\partial}{\partial\pi_i}\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \frac{1}{\pi_i(1-\pi_i)}$. Oleh karena itu, matriks W dapat ditulis sebagai berikut:

$$\begin{aligned} W_i &= \left(Var(y_i)\left(\frac{\partial\eta_i}{\partial\pi_i}\right)^2\right)^{-1} \\ &= \left(\left(\pi_i(1-\pi_i)\left(\frac{1}{\pi_i(1-\pi_i)}\right)\right)^2\right)^{-1} \\ &= \pi_i(1-\pi_i) \end{aligned} \qquad (2.12)$$

Vektor *m* dapat didefinisikan dalam persamaan (2.11) sebagai berikut:

$$\begin{aligned} m_i &= \left(\frac{\partial\eta_i}{\partial\pi_i}\right)(y_i - \pi_i) \\ &= \frac{1}{\pi_i(1-\pi_i)}(y_i - \pi_i) \\ &= \frac{(y_i-\pi_i)}{\pi_i(1-\pi_i)} \end{aligned} \qquad (2.13)$$

Sehingga, turunan pertama log-*likelihood* model Bernoulli dapat ditulis sebagai berikut:

$$\begin{aligned} S(\beta_{t-1}) &= \frac{\partial\ell}{\partial\beta_j} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{p}(y_i - \pi_i)x_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{p}\left(\pi_i(1-\pi_i)\frac{(y_i-\pi_i)}{\pi_i(1-\pi_i)}\right)x_{ij} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{p}W_i m_i\, x_{ij} \\ &= X'Wm \end{aligned} \qquad (2.14)$$

Dimana β_{t-1} merupakan vektor β pada iterasi ke-*t* dan ke- *t-1*. $X'Wm$ merupakan bentuk persamaan matriks. Untuk negatif ekspetasi turunan kedua log-*likelihood* model Bernoulli sebagai berikut.

$$-E\left(\frac{\partial^2\ell(\beta_j)}{\partial\beta_j\partial\beta_j}\right) = -\left(-\sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2\pi_i(1-\pi_i)\right)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2\pi_i(1-\pi_i) \qquad (2.15)$$

Selanjutnya, matriks informasi yang merupakan negatif ekspetasi turunan kedua log-*likelihood* terhadap β_{t-1} dapat ditulis sebagai berikut:

$$I^{-1}(\beta_{t-1}) = -E\left(\frac{\partial^2\ell(\beta_j)}{\partial\beta_j\partial\beta_j}\right)^{-1}$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2\pi_i(1-\pi_i)$$
$$= \sum_{i=1}^{n}\sum_{j=1}^{p} x_{ij}^2 W_i$$
$$= X'WX \qquad (2.16)$$

Dengan bantuan metode *Iteratively Weighted Least Square* (IWLS), didapatkan persamaan:

$$\beta_t = \beta_{t-1} + I^{-1}(\beta_{t-1})S(\beta_{t-1})$$
$$= \beta_{t-1} + [X'WX]^{-1}X'Wm$$
$$= [X'WX]^{-1}X'WX\beta_{t-1} + [X'WX]^{-1}X'Wm$$
$$= [[X'WX]^{-1}X'W][X\beta_{t-1} + m] \qquad (2.17)$$

Vektor *m* diketahui sebagaimana persamaan (2.11) dan fungsi hubung $E(y_i) = \pi_i$ dimana fungsi linear pada model Bernoulli untuk $x_i'\beta$ adalah log$\left[\left(\frac{\pi_i}{1-\pi_i}\right)\right]$ maka persamaan diatas menjadi:

$$= \left[\left[X^iWX\right]^{-1}X'W\right]\left[log\left(\frac{\pi}{1-\pi}\right) + \left(\frac{y-\pi}{\pi(1-\pi)}\right)\right]$$
$$= [X'WX]^{-1}X'Wz \qquad (2.18)$$

Persamaan (2.18) merupakan persamaan untuk mendapatkan beta duga MLE, sehingga dapat ditulis sebagai berikut:

$$\hat{\beta}_{MLE} = [X'WX]^{-1}X'W\hat{z}_i \quad (2.19)$$

dengan:

$\widehat{W} = diag[\widehat{\pi}_i(1-\widehat{\pi}_i)]$ dan $\hat{z}$ yaitu vektor dimana elemen ke-i bernilai $\hat{z}_i = \log(\widehat{\pi}_i) + \frac{y_i - \widehat{\pi}_i}{\widehat{\pi}_i(1-\widehat{\pi}_i)}$.

2.5 Metode Liu

Metode Liu merupakan metode alternatif regresi logistik berupa estimator susut bias dan penduga langsung generalisasi yang diusulkan untuk model regresi linier oleh Liu (1993) guna mengatasi masalah multikolinearitas. Metode ini pertama kali dikenalkan oleh Liu (1993) yang merupakan gabungan dari metode *Stein* dan *Ridge*. Estimator Liu memiliki keunggulan dibanding dengan metode yang lain, seperti memiliki nilai *Scalar Mean Square Error* (SMSE) yang lebih kecil dibandingkan dengan taksiran *Ridge.* Oleh karena itu, Liu (1993) menyarankan estimator lain dimana parameter yang diperoleh dari estimator ini memiliki manfaat sebagai fungsi linier dari parameter penyusutan d. Menurut Qasim, dkk (2019), taksiran nilai d yang dilakukan pada estimator Liu lebih mudah untuk ditentukan. Penyusutan parameter d dapat mengambil nilai antara nol dan satu dan ketika d kurang dari satu maka kita memiliki $\|\hat{\beta}_d\| \leq \|\hat{\beta}_{MLE}\|$. Karena MLE rata-rata, terlalu lama dengan adanya multikolinearitas, d diasumsikan berkinerja lebih baik daripada MLE dalam situasi seperti itu.

Dalam pengerjaaannya, Hoerl & Kennard (1970), Kibria (2003), dan Khalaf & Shukur (2005) mengusulkan ide di mana beberapa metode berbeda diusulkan untuk memperkirakan parameter penyusutan untuk regresi linear ridge. Namun, untuk estimator Liu operator maksimum lain juga digunakan yang akan memastikan bahwa nilai taksiran dari parameter penyusutan tidak negatif.

Adapun penaksiran untuk nilai *d* yang diusulkan oleh Hoerl & Kennard (1970) yaitu:

$$d_1 = max\left[0,\ \frac{\hat{a}_{j\,max}^2 - 1}{\frac{1}{\lambda_{j\,max}} + \widehat{a_{max}^2}}\right] \qquad (2.21)$$

Selanjutnya, estimator berikut yang didasarkan pada ide-ide dalam Kibria (2003), diusulkan:

$$d_2 = max\left[0,\ median\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\lambda_j} + \widehat{a_j^2}}\right] \qquad (2.22)$$

$$d_3 = max\left[0,\ \frac{1}{p}\sum_j^J\left(\frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \widehat{a_j^2}}\right)\right] \qquad (2.23)$$

Akhirnya, estimator berikut diusulkan dimana pada estimator ini kuantil lain selain median digunakan dan berhasil diterapkan oleh Khalaf dan Shukur (2005).

$$d_4 = max\left[0, max\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2}\right] \qquad (2.24)$$

Operasi maksimum dilakukan pada semua nilai d untuk memastikan nilai d berada pada interval 0 sampai 1. Nilai d yang menghasilkan MSE terkecil saat analisis metode Liu, maka nilai d itulah yang dipakai dalam rumus selanjutnya.

2.6 *Mean Square Error*

Mean Square Error (MSE) adalah salah satu pengukuran kesalahan yang terkenal dan mudah digunakan (Ghozali, 2006). Semakin kecil nilai MSE, maka semakin akurat nilai suatu pemodelan. Pada permasalahan multikolinearitas, metode terbaik merupakan metode yang dapat digunakan untuk memperbaiki multikolinearitas. Adapun perbaikan masalah multikolinearitas ini akan dilihat berdasarkan rata-rata dari *Mean Square Error* (MSE) dari hasil estimasi parameter $\hat{\beta}$ sebagai berikut:

$$\text{MSE } \hat{\beta} = \frac{1}{m} \sum_{j=1}^{m} \left(\widehat{\beta_J} - \beta_p\right)^2 \qquad ; j = 1,2, \ldots \ldots m \tag{2.25}$$

dengan:

$\widehat{\beta_J}$ = Penduga parameter regresi

β_p = Parameter regresi

m = Banyaknya ulangan

.

III. METODOLOGI PENELITIAN

3.1 Waktu dan Tempat Penelitian

Penelitian ini dilakukan pada semester genap tahun akademik 2022/2023, bertempat di jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

3.2 Data Penelitian

Data yang digunakan pada penelitian ini yaitu data simulasi yang mengandung multikolinearitas. Untuk data simulasi, data akan dibangkitkan dengan variabel bebas sebanyak p = 4 dan n observasi yaitu 50, 75, dan 100 dengan pengulangan sebanyak 100 kali.

Selanjutnya, untuk mendapat variabel terikat Y, akan dibangkitkan menggunakan probabilitas regresi logistik biner dengan distribusi binomial sebagai berikut:

$$P(y_p = 1)Y = \pi(x) = \frac{exp\ (\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)}{1+xp\ (\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_p x_p)} \tag{3.1}$$

dengan β_0 = 0 dan $\beta_1 = \beta_2 = \beta_3 \ldots = \beta_p = 1$, dimana data $X \sim b(n, p)$ dengan nilai n= 50,75,100 dan p atau probabilitas sebesar 0,5.

Sedangkan, untuk mendapatkan data multikolinearitas pada setiap himpunan data Xp, maka dibangkitkan menggunakan simulasi Monte Carlo dengan persamaan sebagai berikut:

$$X_p = \sqrt{(1-\rho^2)z_{ij}} + \rho z_{i(p+1)}, i = 1,2,\dots,n \text{ dan } j = 1,2,3,4 \qquad (3.2)$$

dengan z_{ij} ~N(0,1) and ρ = 0.99.

3.3 Metode Penelitian

Penelitian ini dilakukan dengan menggunakan metode studi pustaka, yaitu mempelajari buku penunjang dalam bentuk jurnal maupun karya ilmiah. Untuk mendapatkan hasil yang baik, penelitian ini menggunakan bantuan *software* R versi 4.2.1.

Langkah-langkah yang dilakukan dalam penelitian ini, yaitu:

1. Melakukan simulasi data regresi logistik dengan langkah-langkah sebagai berikut.
    - Membangkitkan sejumlah n data acak normal dengan rataan 0 dan variansi 1 sebanyak variabel bebas (p) dan menggunakan simulasi Monte Carlo pada persamaan (3.2) untuk mendapatkan data multikolinearitas pada setiap data himpunan X.
    - Menentukan nilai beta koefisien, dimana β_0 = 0 dan $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$.
    - Membangkitkan nilai variabel respon Y dengan menggunakan probabilitas model regresi logistik $\pi(x)$, kemudian dikonversi kedalam variabel dikotomus dengan kriteria probabilitas 0.5.
2. Mengidentifikasi multikolinearitas menggunakan VIF dengan persamaan (2.2) dan korelasi pada data simulasi.

3. Melakukan penghitungan untuk menentukan koefisien parameter β dengan regresi logistik menggunakan *Maximum Likelihood Estimation* (MLE) dengan persamaan (2.19).
4. Menghitung nilai SE (*Standard Error*) penduga dan MSE (*Mean Square Error*) dari penduga β pada MLE dengan pengulangan sebanyak 100 kali.
5. Melakukan analisis Metode Liu untuk menentukan nilai parameter β dan MSE dengan Langkah-langkah sebagai berikut:
   - Menentukan matrik X'WX, λ, α, γ sebagai elemen dasar mencari β.
   - Melakukan penaksiran nilai *d* yang diusulkan Hoerl & Kennard (1970), Kibria (2003), Khalaf & Shukur (2005) sebagai acuan dimana penaksir *d* yang menghasilkan MSE terkecil adalah *d* yang dipilih untuk menentukan nilai β, menggunakan persamaan:

$$d_1 = max\left[0,\ \frac{\hat{a}_{j\,max}^2 - 1}{\frac{1}{\lambda_{j\,max}} + \widehat{a_{max}^2}}\right]$$

$$d_2 = max\left[0,\ median\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\lambda_j} + \widehat{a_j^2}}\right]$$

$$d_3 = max\left[0,\ \frac{1}{p}\sum_j^J\left(\frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \widehat{a_j^2}}\right)\right]$$

$$d_4 = max\left[0, max\ \frac{\hat{a}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2}\right]$$

   - Menentukan nilai MSE menggunakan persamaan (2.25).
   - Menentukan nilai β dengan persamaan:

$$\widehat{\beta_d} = \left(X'\widehat{W}X + I\right)^{-1}\left(X'\widehat{W}X + dI\right)\hat{\beta}_{MLE}$$

6. Menghitung nilai SE pada data simulasi menggunakan metode MLE dan Liu dengan pengulangan sebanyak 100 kali.
7. Membandingkan nilai parameter β, SE dan MSE metode MLE dengan metode Liu.
8. Menentukan metode terbaik yang dapat digunakan antara MLE dan metode Liu dengan melihat kriteria pada MSE terkecil.

V. KESIMPULAN

Berdasarkan penelitian ini, dapat disimpulkan bahwa:

1. Metode Liu merupakan metode alternatif yang dapat mengatasi masalah multikolinearitas pada regresi logistik.
2. Berdasarkan nilai MSE, performa metode Liu dalam mengatasi multikolinearitas lebih baik dibandingkan dengan metode MLE karena menghasilkan nilai MSE yang lebih kecil. Berdasarkan penelitian diatas, dengan menggunakan metode Liu didapatkan MSE untuk n=50 sebesar 0.0399, n=75 sebesar 0.0399, n=100 sebesar 0.03996. Semakin kecil nilai MSE maka semakin baik pula akurasi dalam menentukan nilai suatu model regresi.

DAFTAR PUSTAKA

Agresti, A. 2002. *Categorical Data Analysis*. Ed. ke-2. John willy and Sons, New York.

Drapper, N.R. & Smith, H. 1998. *Analisis Regresi Terapan*. Ed. Ke-3. Diterjemahkan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.

Ghozali, I. 2006. *Aplikasi Analisis Multivariat dengan Program SPSS*. Ed. ke-4. Badan Penerbit Universitas Diponegoro, Semarang.

Haloho, O., Sembiring, P., & Manurung, A. 2013. Saintia Matematika. *Penerapan Analisis Regresi Logistik pada Pemakaian Alat Kontrasepsi Wanita*. **1**(1): 51-61.

Hoerl, A.E. & Kennard, R.W. 1970. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics.* **12**:55–67.

Hosmer, D.W. & Lemeshow, S. 2000. *Applied Logistics Regression*. John Wiley & Sons, New York.

Khalaf, G. & Shukur, G. 2005. Choosing Ridge Parameter for Regression Problems. *Commun. Stat. Theory Methods*. **34**(5):1177–1182.

Kibria, B.M.G. 2003. Performance of some new ridge regression estimators. *Commun. Stat. Theory Methods.* **32**:419–435.

Kejian, L. 1993. A New Class of Biased Estimate in Linear Regression. *Communications in Statistics-Theory and Methods*. **22**(2): 393-402.

Mansson, K., Kibria, B.M.G., & Shukur, G. 2012. On Liu estimators for the logit regression model. *Econ. Modell*. **29**:1483–1488.

Montgomery, D.C. & Peck, A.E. 1992. *Introduction to Linear Regression Analysis*. A Wiley Intersection Publicatiom, New York.

Qasim, M., Amin, M., & Omer, T. 2019. Performance Of Some New Liu Parameters for the Linear Regression Model. *Communication In Statistics Theory And Methods*. **49**(17):4178-4196.

Sembiring, R.K. 2003. *Analisis Regresi*. Ed. Ke-3. ITB Press, Bandung.

Uyanik, G.K. & Guler, N. 2013. A study on multiple linear regression analysis. Proce*dia-Social and Behavioral Science*. **106**: 234-240.