ESTIMASI PARAMETER MODEL REGRESI PROBIT DENGAN METODE *RIDGE* PADA DATA YANG MENGANDUNG MULTIKOLINIERITAS

SKRIPSI

SRI IRMA YANI

H 121 15 305

PROGRAM STUDI STATISTIKA DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

FEBRUARI 2019

ESTIMASI PARAMETER MODEL REGRESI PROBIT DENGAN METODE *RIDGE* PADA DATA YANG MENGANDUNG MULTIKOLINIERITAS

SKRIPSI

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

SRI IRMA YANI

H 121 15 305

PROGRAM STUDI STATISTIKA DEPARTEMEN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

FEBRUARI 2019

LEMBAR PERNYATAAN KEOTENTIKAN

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

Estimasi Parameter Model Regresi Probit Dengan Metode *Ridge* Pada Data yang Mengandung Multikolinieritas

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

Makassar, 25 Februari 2019

SRI IRMAYANI

NIM. H 121 15 305

ESTIMASI PARAMETER MODEL REGRESI PROBIT DENGAN METODE RIDGE PADA DATA YANG MENGANDUNG MULTIKOLINIERITAS

Disetujui Oleh:

Pembimbing Utama	Pembimbing Pertama
Anisa, S.Si., M.Si.	**Dr. Amran, S.Si., M.Si.**
NIP. 19730227 199802 2001	**NIP. 1970 1101 199802 1001**

Pada Tanggal: 25 Februari 2019

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Sri Irma Yani

NIM : H 121 15 305

Program Studi : STATISTIKA

Judul Skripsi : Estimasi Parameter Model Regresi Probit Dengan Metode *Ridge* Pada Data yang Mengandung Multikolinieritas

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

DEWAN PENGUJI

Tanda Tangan

1. Ketua : Anisa, S.Si., M.Si. (.................................)
2. Sekretaris : Dr. Amran, S.Si., M.Si. (.................................)
3. Anggota : Dr. Nurtiti Sanusi, S.Si., M.Si (.................................)
4. Anggota : Sitti Sahriman, S.Si., M.Si (.................................)

Ditetapkan di : Makassar

Tanggal : 25 Februari 2019

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

Segala puji bagi Allah *Subhanahu Wa ta'ala Rabb* semesta alam, shalawat serta salam semoga senantiasa tercurahkan kepada Nabi yang paling dimuliakan, pemimpin orang-orang bertakwa, Rasulullah Muhammad *Shallallahu Alaihi Wasallam* dan kepada keluarga serta sahabat beliau yang senantiasa kita rindukan perjumpaan dengannya. Amma ba'du.

Alhamdulillah, berkat pertolongan Allah akhirnya skripsi dengan judul "Estimasi Parameter Model Regresi Probit Dengan Metode *Ridge* Pada Data yang Mengandung Multikolinieritas" yang disusun sebagai salah satu syarat akademik untuk meraih gelar sarjana sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin ini dapat dirampungkan. Dalam penulisan skripsi ini, penulis dengan segala keterbatasan kemampuan dan pengetahuan dapat melewati segala hambatan dan masalah berkat bantuan dan dorongan dari berbagai pihak. Oleh karena itu penulis menyampaikan ucapan terima kasih yang tak terhingga kepada orang tua penulis, Ibunda **Husnawati** dan Ayahanda **Alm. Loe Rachman Sakka** sebagai madrasah pertama yang telah banyak memberikan cinta, kasih sayang, doa, nasehat, dan segala bentuk pelajaran serta pendidikan sebagai bekal menjalani kehidupan. Untuk kakakku, **Yusri, S.Pd.SD,** dan kakak ipar **Yusridawati, S.Pd.**, serta adikku **Dianita Salsa** yang sangat saya banggakan terima kasih atas segala bentuk bantuan dan motivasi yang telah diberikan kepada penulis.

Tidak lupa pula penulis ucapkan terima kasih kepada seluruh pihak yang senantiasa membantu baik berupa materi, tenaga dan dukungan moral selama proses penyelesaian tulisan ini:

1. Ibu **Prof. Dr. Dwia Aries Tina Palubuhu, MA** selaku Rektor Universitas Hasanuddin.
2. Bapak **Dr. Eng. Amiruddin** selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.
3. Bapak **Prof. Dr. Amir Kamar Amir, M.Sc** selaku Ketua Departemen Matematika, segenap dosen pengajar dan staf Departemen Matematika yang

telah membekali ilmu dan kemudahan-kemudahan kepada penulis dalam berbagai hal selama menjadi mahasiswa di Departemen Matematika.

4. Ibu **Anisa, S.Si., M.Si.** selaku Ketua Tim Penguji sekaligus dosen Pembimbing Utama yang telah dengan sabar dan ikhlas meluangkan begitu banyak waktunya untuk membimbing dan memberikan masukan serta motivasi dalam penulisan skripsi ini.
5. Bapak **Dr. Amran, S.Si., M.Si.** selaku Sekretaris Tim Penguji sekaligus dosen Pembimbing Pertama yang juga telah meluangkan waktu dan pemikirannya untuk memberikan arahan dan motivasi dalam penulisan skripsi ini.
6. Ibu **Dr. Nurtiti Sanusi, S.Si., M.Si.** selaku Anggota Tim Penguji sekaligus Penasehat Akademik yang telah meluangkan begitu banyak waktunya memberikan bimbingan, masukan, saran serta motivasi selama menjalani pendidikan di Departemen Matematika.
7. Ibu **Sitti Sahriman, S.Si., M.Si** selaku Anggota Tim Penguji yang telah memberikan saran dan kritikan yang membangun dalam penyempurnaan penyusunan skripsi ini serta waktu yang telah diberikan kepada penulis.

Ucapan terima kasih juga penulis sampaikan kepada orang-orang yang telah berperan besar serta istimewa kepada:

1. Keluarga besar **STATISTIKA 2015**: **Bayu, Dian, Ani, Ihza, Erdi, Jidil, Fadil, Aminah** serta teman-teman lain yang tak sempat disebutkan namanya, terima kasih atas kebersamaan dan kebahagiaannya.
2. Keluarga besar **KM FMIPA UNHAS** dan **HIMATIKA FMIPA UNHAS**, terima kasih atas ilmu yang mungkin tidak bisa didapatkan diproses perkuliahan. Penulis merasa bangga menjadi salah satu bagian dari organisasi/himpunan ini. **USE YOUR MIND BE THE BEST** dan **BRAVO HIMATIKA.**
3. Seluruh teman-teman **SIMETRIS 2015** terima kasih telah bersabar direpotkan dan terima kasih untuk cerita sekaligus kenangan selama kita erproses.

eman-teman **KKN Tematik Infrastruktur Pemukiman Wajo** di elurahan Pattirosompe: **Putri, Neneng, Nunu, Dila, Farida, Eva, Okky,**

Brily dan **Ambo**, terima kasih untuk kebersamaan dalam suasana kekeluargaan yang hangat.

5. Teman-teman alumni **MAN Baraka**, **Armin, Fathul, Dini, Dila** dan seluruh teman XII IPA-1 serta kanda **Reno** yang telah banyak berjasa dan senantiasa menghibur, memberikan dukungan dan motivasi kepada penulis.
6. Kepada kak **Muh. Rusdi, S.Pd.I** yang sudah begitu berjasa membantu, meluangkan waktu, memberikan dorongan, motivasi dan segala bentuk bantuan yang tak ternilai dalam proses penyusunan tugas akhir ini.
7. Kepada semua pihak yang tidak dapat penulis sebutkan satu-persatu, semoga segala dukungan dan partisipasi yang diberikan kepada penulis bernilai ibadah disisi Allah *Subhanahu Wa ta'ala.*

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

Makassar, 25 Februari 2019

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama	: Sri Irma Yani
NIM	: H 121 15 305
Program Studi	: Statistika
Departemen	: Matematika
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:

"Estimasi Parameter Model Regresi Probit Dengan Metode *Ridge* Pada Data yang Mengandung Multikolinieritas"

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

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar, pada tanggal 25 Februari 2019

Yang menyatakan

ma Yani)

ABSTRAK

Model regresi probit merupakan suatu model regresi yang digunakan ketika variabel respon kategorik berupa variabel dikotomi yang menunjukkan ada atau tidaknya kriteria atribut dengan menggunakan nilai 0 atau 1. Terdapat banyak cara yang dapat digunakan untuk mengestimasi parameter model regresi probit, salah satunya dengan menggunakan metode *maximum likelihood estimation* (MLE). Namun, ketika terjadi multikolinieritas antar variabel prediktor, maka variansi semakin membesar yang menyebabkan estimasi dengan MLE menjadi tidak efisien. Salah satu metode untuk menangani multikolinearitas adalah regresi *ridge*. Penelitian ini bertujuan untuk memperoleh estimasi parameter model regresi probit dengan menggunakan metode *ridge* pada data yang mengandung multikolinearitas. Pendugaan paremeter pada model regresi probit dengan metode *ridge* melibatkan penambahan konstanta bias (k) ke setiap elemen diagonal matriks. Penelitian ini diaplikasikan pada data kemiskinan kabupaten/kota di Provinsi Sulawesi Selatan tahun 2017. Variabel respon yang bersifat kategorik dalam penelitian ini yaitu persentase penduduk miskin sesuai indikator kemiskinan dari Badan Pusat Statistik dan variabel prediktor yaitu faktor-faktor yang mempengaruhi kemiskinan yaitu pengeluaran perkapita, ketenagakerjaan, fasilitas perumahan dan pendidikan. Hasil yang diperoleh adalah nilai *mean square error* (MSE) dari penduga parameter dengan metode *ridge* sebesar 0.3672 sedangkan dengan metode MLE diperoleh MSE sebesar 4.5108 dan metode klasik, yaitu metode Ordinary Least Square diperoleh MSE sebesar 10.19. Hal ini menunjukkan bahwa metode *ridge* lebih efektif digunakan untuk mengatasi masalah multikolinearitas.

Kata Kunci : Regresi Probit, Multikolinieritas, *Maximum Likelihood Estimation*, *Ridge*, *Mean Square Error*.

ABSTRACT

The probit regression model is a regression model that is used when the categorical response variable is a dichotomous variable that indicates the presence or absence of attribute criteria using a value of 0 or 1. There are many ways that can be used to estimate the parameters of the probit regression model, one of them using the maximum method likelihood estimation (MLE). However, when there is multicollinearity between predictor variables, the variance will increase which will make estimation with MLE inefficient. One method for dealing with multicollinearity is ridge regression. The estimation of parameters in the probit regression model with the ridge method involves adding a bias constant *(k)* to each diagonal element of the matrix. This study aims to obtain parameter estimates of probit regression models using the ridge method on data containing multicollinearity. This study was applied to the poverty data of districts / cities in South Sulawesi Province in 2017. Categorical response variables in this study were the percentage of poor people according to poverty indicators from the Badan Pusat Statistik and predictor variables namely factors that influence poverty, namely expenditure per capita, employment housing and education facilities. The results obtained are the mean square error (MSE) of the parameter estimator with the ridge method of 0.3672 while the MLE method obtained by MSE is 4.5108 and the classical method, namely the Ordinary Least Square method obtained by MSE is 10.19. This shows that the ridge method is more effectively used to overcome multicollinearity problems.

Keywords : *Probit Regression, Multicollinearity, Maximum Likelihood Estimation, Ridge, Mean Square Error.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi adalah suatu teknik yang digunakan untuk membentuk suatu persamaan atau model yang menghubungkan antara variabel respon (Y) dengan satu atau lebih variabel prediktor (X), serta untuk menentukan nilai taksirannya. Kata taksiran identik dengan dugaan, begitu pula penaksiran identik dengan pendugaan atau estimasi. Proses estimasi penting dilakukan untuk mengetahui dampak yang terjadi akibat perubahan suatu variabel terhadap variabel lain sehingga dapat dilakukan antisipasi dalam menanggulangi dampak tersebut.

Analisis regresi telah berkembang dan mengalami perubahan yang semakin beragam, tidak hanya berpacu pada keberadaan variabel respon maupun variabel prediktor yang bersifat kuantitatif (Supranto, 2005). Selain dengan data kuantitatif, analisis regresi juga dapat dilakukan terhadap data kualitatif, dimana data kualitatif adalah data yang tidak bersifat numerik, tetapi dapat diolah dan dihitung dengan cara mengubah dari data kualitatif menjadi data kuantitatif.

Salah satu metode kuantifikasi atribut-atribut ini adalah dengan membentuk variabel-variabel artifisial yang menggunakan nilai-nilai 0 atau 1. Variabel-variabel yang mengasumsikan nilai-nilai seperti 0 dan 1 ini disebut dengan variabel buatan (*dummy variable*). Suatu model regresi yang hanya berisikan variabel-variabel prediktor *dummy* disebut dengan model analisis varians (Gujarati, 2006). Jika variabel yang bersifat *dummy* adalah variabel respon, maka salah satu pendekatan model yang dapat digunakan adalah model regresi probit. Sebuah model regresi dikatakan baik, jika memenuhi asumsi-asumsi model.

Keberadaan regresi logistik tidak memerlukan asumsi kenormalan baik pada distribusi galatnya maupun pada variabel responnya serta tidak ada asumsi yang menyatakan bahwa hubungan antara variabel respon dengan variabel prediktor adalah linier (Hosmer dan Lemeshow, 2000). Namun, regresi logistik memiliki an yakni tidak menunjukkan secara pasti kecenderungan munculnya suatu yang dipengaruhi oleh beberapa faktor terutama variabel kontinu yang tak teramati. Oleh sebab itu, regresi probit berkembang sebagai salah satu alternatif yang digunakan untuk mengatasi masalah tersebut.

Model regresi probit adalah model tak linier yang digunakan untuk menganalisis hubungan antara satu variabel respon dan beberapa variabel bebas, dengan variabel responnya berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan suatu karakteristik dan bernilai 0 untuk menyatakan ketidakberadaan suatu karakteristik (Young, 2003).

Permasalahan yang sering terjadi pada regresi probit dengan variabel prediktor lebih dari satu adalah terjadi korelasi antar variabel-variabel prediktor tersebut yang disebut sebagai multikolinieritas. Hal ini mengakibatkan penduga/estimator yang dihasilkan menjadi tidak efisien sehingga variansi dari koefisien regresi menjadi tidak minimum (Gujarati, 2006). Selain itu, dapat pula mengakibatkan variabel prediktor tidak signifikan mempengaruhi variabel respon, meskipun nilai koefisien determinasinya *($\boldsymbol{R}^2$)* tinggi sehingga model yang didapatkan menjadi kurang layak. Oleh karena itu masalah multikolinearitas perlu diatasi agar estimasi parameter regresi menjadi optimal.

Terdapat beberapa metode untuk mengatasi masalah multikolinearitas, diantaranya yaitu metode *ridge* yang pertama kali diusulkan oleh Hoerl dan Kennard (1970) pada model regresi linear dengan cara menambahkan suatu konstanta positif (*k*) yang kecil pada elemen diagonal matriks $\boldsymbol{X}^t\boldsymbol{X}$, yang mengakibatkan matriks $\boldsymbol{X}^t\boldsymbol{X}$ menjadi matriks non-singular.

Beberapa penelitian terdahulu yang terkait dengan metode estimasi yang digunakan untuk mengestimasi model regresi probit yang mengandung multikolinieritas di antaranya *Performance of Some Ridge Parameters for Probit Regression* (Locking, dkk., 2011) serta *Improving the Estimators of the Parameters of a Probit Regression Model: A Ridge Regression Approach* (Kibria dan Saleh, 2011), dan skripsi Sa'adah (2011) yang berjudul *Analisis Regresi Dummy Variable Model Probit*.

Berdasarkan uraian diatas, pada penelitian ini penulis mengkaji tentang [illegible] parameter model regresi probit. Sehingga penulis tertarik untuk [illegible]bil judul penelitian "***Estimasi Parameter Model Regresi Probit Dengan*** [illegible]***idge Pada Data yang Mengandung Multikolinieritas***".

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana bentuk penduga parameter *ridge* pada model regresi probit untuk mengatasi masalah multikolinieritas ?
2. Bagaimana mengestimasi parameter model regresi probit dengan metode *ridge* pada data kemiskinan di Provinsi Sulawesi Selatan tahun 2017 ?

1.3 Batasan Masalah

Penelitian ini memfokuskan pada penanganan masalah multikolinearitas pada model regresi probit dengan metode *ridge* menggunakan parameter

$$k = max\left(\frac{1}{q_j}\right) \text{ dengan } q_j = \frac{\lambda_{max}}{(n-p)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}_j^2}, j = 1,2,3.$$

Penelitian ini menggunakan data kemiskinan kabupaten/kota Provinsi Sulawesi Selatan tahun 2017. Variabel prediktor yang digunakan hanya delapan yang tersedia dari empat belas variabel yang merupakan kriteria kemiskinan.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan yang akan dicapai dalam penelitian ini adalah:

1. Memperoleh penduga parameter *ridge* pada model regresi probit untuk mengatasi masalah multikolinieritas.
2. Memperoleh estimasi parameter model regresi probit dengan metode *ridge* pada data kemiskinan di Provinsi Sulawesi Selatan tahun 2017.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

a. Bagi Peneliti

Penulis mengetahui tentang metode dan hasil estimasi parameter model regresi probit dengan metode *ridge*. Dapat menjadi wacana baru dalam pengembangan ilmu pengetahuan khususnya ilmu matematika yang dapat nfaatkan dalam kehidupan sehari-hari.

Lembaga

ebagai sumbangan pemikiran keilmuan matematika, khususnya dalam bidang statistika, analisis regresi dan ekonometrika.

2. Meningkatkan peran serta Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin dalam pengembangan wawasan keilmuan matematika di bidang statistika.

c. Bagi Pembaca

Memberikan pengetahuan lebih mendalam tentang estimasi parameter model regresi probit dengan metode *ridge* dan menjadikan penelitian ini sebagai bahan rujukan dalam pengembangan pembelajaran statistika, analisis regresi, dan ekonometrika yang berkelanjutan.

BAB II

TINJAUAN PUSTAKA

2.1 Estimasi Parameter

Menurut Yitnosumarto (1990), *estimator* (penduga) adalah anggota peubah acak dari statistik yang mungkin untuk sebuah parameter (anggota peubah diturunkan). Estimasi merupakan suatu pernyataan untuk menduga hubungan mengenai parameter populasi yang tidak diketahui menggunakan sampel (statistik), dalam hal ini peubah acak yang diambil dari populasi yang bersangkutan. Jadi dengan estimasi, keadaan populasi dapat diketahui (Hasan, 2002). Besaran sebagai hasil penerapan estimasi terhadap data dari semua contoh disebut nilai taksir (*estimator value*).

Adapun sitat-sifat estimator yang baik adalah sebagai berikut :

1. *Unbiased*

   Suatu hal yang menjadi tujuan dalam mengestimasi adalah estimator harus mendekati nilai sebenarnya dari parameter yang diduga tersebut. Misalkan terdapat parameter $\bar{\beta}$. Jika $\hat{\beta}$ merupakan penaksir tak bias dari parameter $\bar{\beta}$, maka $E(\hat{\beta}) = \bar{\beta}$ (Yitnosumarto, 1990).

2. Efisien

   Suatu estimator misalkan $\hat{\beta}$ dikatakan efisien bagi parameter $\bar{\beta}$ apabila penduga tersebut mempunyai variansi yang kecil. Apabila terdapat lebih dari satu penduga, penduga yang efisien adalah penduga yang mempunyai variansi terkecil (Supranto, 1986).

3. Konsisten

   Suatu estimasi dikatakan konsisten apabila nilai estimasi tersebut sama dengan parameter yang diestimasi. Misalnya $\hat{\beta}$ merupakan estimasi dari $\bar{\beta}$ dengan sampel acak berukuran $\bar{\beta}$ yang menuju tak hingga dan variansi mendekati 0 maka $\hat{\beta}$ mendekati $\bar{\beta}$ (Supranto, 1986).

lisis Regresi

lah regresi pertama kali diperkenalkan oleh Sir Francis Galton pada tahun alisis regresi adalah teknik analisis yang mencoba menjelaskan bentuk n antara peubah-peubah yang mendukung sebab akibat. Prosedur analisisnya didasarkan atas distribusi probabilitas bersama peubah-peubahnya. Bila hubungan ini dapat dinyatakan dalam persamaan matematik, maka kita dapat memanfaatkan untuk keperluan lain misalnya peramalan (Sembiring, 1995).

Tujuan utama dari analisi regresi adalah untuk mendapatkan dugaan dari suatu variabel dengan menggunakan variabel lain yang diketahui. Analisis regresi mempunyai dua jenis pilihan yaitu regresi linier dan regresi non linier.

2.2.1 Regresi Linier Berganda

Model regresi yang paling sederhana adalah model regresi linier. Model regresi linier sederhana terdiri dari satu variabel bebas. Persamaan model regresi linier dengan k variabel bebas diberikan sebagai

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon \tag{2.1}$$

dengan,

Y	= variabel respon
$X_1, X_2, \dots, X_p$	= variabel prediktor
$\beta_0, \beta_1, \beta_2, \dots, \beta_p$	= parameter
ε	= galat

Bila pengamatan $Y, X_1, X_2, \dots, X_p$ dinyatakan masing-masing dengan $Y_i, X_{i1}, X_{i2}, \dots, X_{ip}$ dan galatnya ε_i, maka Persamaan (2.1) dapat dituliskan sebagai

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i, i = 1,2, \dots, n \tag{2.2}$$

dengan mean $E(\varepsilon_i) = 0$ dan variansinya $\sigma^2(\varepsilon_i) = \sigma^2$, dan tidak berkorelasi sehingga kovariansinya $E(\varepsilon_i, \varepsilon_j) = 0, i \neq j, i = l = 1,2, \dots, n$. Apabila dinotasikan dalam bentuk matriks, menjadi

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1p} \\ 1 & X_{21} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_p \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.3}$$

Menurut Sembiring (1995) Persamaan (2.3) dapat dinyatakan sebagai

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{2.4}$$

dengan,

or variabel respon (ukuran $n \times 1$)

iks variabel prediktor ukuran $(n \times (p+1))$

or parameter (ukuran $(p+1) \times 1$)

$\boldsymbol{\varepsilon}$ = vektor galat (ukuran $n \times 1$)

Persamaan matriks (2.4) dikenal sebagai penyajian matriks model regresi linier (p-variabel).

2.2.2 Regresi Probit

Analisis regresi logistik merupakan salah satu metode yang digunakan untuk menggambarkan hubungan antara variabel respon dan prediktor, dimana variabel respon hanya memiliki 2 kemungkinan nilai/hasil *(dikotomus)*, misalnya ya/tidak, sukses/gagal, dan lain sebagainya. Regresi logistik termasuk dalam kelompok *Generalized Linear Models* atau GLM. Model linier umum merupakan pengembangan dari model linier klasik. Model linier umum mengasumsikan bahwa komponen acak dan variabel respon tidak harus mengikuti distribusi normal, tetapi harus termasuk dalam distribusi keluarga eksponensial (Pradita, 2011). Salah satu bentuk pengembangan model regresi logistik adalah regresi probit.

Menurut Candra (2009), regresi probit merupakan regresi nonlinier yang digunakan untuk menganalisis hubungan antara satu variabel respon dengan beberapa variabel prediktor, dengan variabel respon berupa data kualitatif dikotomi yaitu bernilai 1 untuk menyatakan keberadaan sebuah atribut dan bernilai 0 untuk menyatakan ketidakberadaan sebuah atribut.

Untuk menganalisis sifat-sifat variabel respon kategorik, diperlukan untuk memilih *Cumulative Distribution Function* (CDF) yang tepat. Dalam hal ini, model yang menggunakan CDF Normal disebut Model Probit (Djalal, 2004). Menurut Skrondal & Hesketh dalam Widhiarso (2012), regresi probit merupakan modifikasi regresi logistik dengan menetapkan persamaan regresi logit berdistribusi normal.

Jika terdapat variabel X mengikuti distribusi normal dengan rata-rata μ dan variansi σ^2 disebut *Probability Density Function* (PDF) jika

$$f(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2} \tag{2.5}$$

dan CDF diberikan sebagai berikut:

$$F(x) = \int_{-\infty}^{X_0} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2} dx \tag{2.6}$$

samaan (2.6) dapat ditransformasi ke dalam bentuk distribusi normal ang dinyatakan dengan

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1) \tag{2.7}$$

dengan

$$
\begin{aligned}
F(z) &= P[Z \leq z] \\
&= P\left[\frac{X-\mu}{\sigma} \leq z\right] \\
&= P[X \leq \mu + z\sigma] \\
&= \int_{-\infty}^{\mu+z\sigma} \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \qquad (2.8)
\end{aligned}
$$

misal $w = \frac{x-\mu}{\sigma}$, maka $dw = \frac{1}{\sigma} dx$, sehigga

$$
\begin{aligned}
x &= \mu + z\sigma \\
w &= \frac{(\mu + z\sigma) - \mu}{\sigma} \\
w &= z
\end{aligned}
$$

maka diperoleh

$$
\begin{aligned}
F(z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} dw \\
&= \Phi(z) \qquad (2.9)
\end{aligned}
$$

Notasi $\Phi(z)$ merupaan CDF dari distribusi normal standar. Selanjutnya, turunan dari standar normal kumulatif disebut PDF normal standar yang dinyatakan sebagai berikut:

$$
\begin{aligned}
f(z) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w^2} \\
&= \phi(z) \qquad (2.10)
\end{aligned}
$$

model Probit yang dikembangkan oleh McFadden, yang dituliskan sebagai berikut:

$$
\begin{aligned}
P_i &= P(Y_i = 1|X_i) \\
&= P(I^* \leq I_i) \\
&= P(Z_i \leq \beta_1 + \beta_2 X_i) \\
&= F(\beta_1 + \beta_2 X_i) \\
&= F(I_i) \qquad (2.11)
\end{aligned}
$$

dengan CDF normal dapat dituliskan

$$
F(I) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{I_i} e^{-Z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta_1+\beta_2 X_i} e^{-Z^2/2} dz \qquad (2.12)
$$

, 2006)

Persamaan yang didasari oleh distribusi normal (Φ) menunjukkan berlakunya fungsi invers dari distribusi normal standar (*inverse standard normal distribution*).

$$\begin{aligned} F(I) &= P(Y = 1) \\ &= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dZ \\ &= \Phi(Z) \end{aligned} \tag{2.13}$$

Z adalah suatu variabel kontinu yang tidak teramati (laten) karena merupakan suatu "kecenderungan" munculnya suatu kejadian. Misalnya data yang teramati adalah lulus (kode 1) dan tidak lulus (kode 0), maka nilai Z menunjukkan kecenderungan atau probabilitas untuk lulus. Dalam kasus lain, dimisalkan Z merupakan suatu kecenderungan untuk melakukan pembelian ulang. Semakin besar nilai Z, maka semakin besar pula kecenderungan pelanggan untuk melakukan pembelian ulang.

Menurut Locking, dkk (2011), model regresi probit ditunjukkan pada persamaan

$$y_i^* = x_i'\boldsymbol{\beta} + \varepsilon_i \tag{2.14}$$

dengan y_i^* merupakan variabel laten, x_i adalah baris ke-i dari $\boldsymbol{X}$ yang merupakan matriks berordo $n \times (p+1)$ dengan p merupakan banyaknya variabel prediktor, $\boldsymbol{\beta}$ adalah vektor koefisien $(p+1) \times 1$ dan ε_i adalah error yang diasumsikan berdistribusi normal. Variabel laten tidak dapat diamati secara langsung, namun dapat dianalisis melalui variabel dummy sebagai berikut:

$$y_i = \begin{cases} 1, & jika\ y_i^* > 0 \\ 0, & untuk\ y_i^*\ lainnya \end{cases}$$

y_i berdistribusi $Be(\pi_i)$, dengan $\pi_i = \Phi(x_i'\boldsymbol{\beta})$ dan Φ adalah fungsi distribusi normal standar.

2.3 Metode *Maximum Likelihood Estimator*

Suatu metode yang bersifat umum dari estimasi titik dengan beberapa sifat teoretis yang lebih kuat dibandingkan dengan metode penaksiran kuadrat terkecil ... *uare estimation*) adalah metode kemungkinan terbesar (*Maximum* ... *d*) (Aziz, 2010).

Metode *Maximum Likelihood Estimator* (MLE) merupakan metode yang digunakan dalam menduga parameter regresi logistik. Prinsip dari MLE adalah menemukan penduga $\boldsymbol{\beta}$ yang memaksimumkan fungsi *likelihood* disamakan dengan 0. Karena setiap observasi bersifat independen, maka bentuk umum fungsi *likelihood* dari distribusi Bernoulli adalah (Agresti, 2002):

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(y_i)$$

$$= \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \qquad (2.15)$$

dengan

$$\pi_i = \frac{\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik})}{1+\exp(\beta_0+\beta_1 x_{i1}+\beta_2 x_{i2}+\cdots+\beta_k x_{ik})}$$

$$= \frac{\exp(\sum_{j=0}^{k}\beta_j X_{ij})}{1+\exp(\sum_{j=0}^{k}\beta_j X_{ij})}$$

$$= \frac{\exp(\boldsymbol{X_i\beta})}{1+\exp(\boldsymbol{X_i\beta})}, i = 1,2,\ldots,n$$

dengan $\boldsymbol{X_i} = (1, x_{i1}, x_{i2}, \ldots, x_{ik})$ dan $\boldsymbol{\beta} = (\beta_0\ \beta_1\ \beta_2 \ldots \beta_k)^t$

Perhitungan lebih mudah dilakukan dengan memaksimumkan fungsi *likelihood* yang disebut fungsi *log-likelihood* berupa logaritma natural dari fungsi *likelihood* tersebut, sehingga dituliskan sebagai :

$$\ell = \ln L(\boldsymbol{\beta}) = \ln\left(\prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}\right)$$

$$= \sum_{i=1}^{n}\{-\ln(1+\exp(\boldsymbol{X_i\beta})) + y_i \boldsymbol{X_i\beta}\} \qquad (2.16)$$

Menurut McCullagh dan Nelder (1989) penduga varian dan kovarian diperoleh dari turunan kedua fungsi log *likelihood*, sebagai berikut:

$$\frac{\partial^2\ell}{\partial^2\beta} = -\sum_{i=1}^{n}(\boldsymbol{X}_i^2\pi_i(1-\pi_i)) = -\sum_{i=1}^{n}\left(\boldsymbol{X}_1^2\widehat{\boldsymbol{W}}_i\right)$$

$$\frac{\partial^2\ell}{\partial^2\beta} = -\boldsymbol{X}^t\widehat{\boldsymbol{W}}\boldsymbol{X} \qquad (2.17)$$

dengan $\widehat{\boldsymbol{W}} = diag[\pi_i(1-\pi_i)]$

Karena model regresi logistik merupakan fungsi nonlinear, maka proses perhitungan MLE dapat didekati dengan metode *Weighted Least Square (WLS)*. asi ini matriks pembobotnya berubah setiap iterasi. Adapun penduga WLS ulis sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{WLS} = \left(\boldsymbol{X}^t\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\widehat{\boldsymbol{W}}\widehat{\boldsymbol{Y}} \qquad (2.18)$$

Metode ini merupakan pengembangan dari metode *fisher scoring* (Agresti, 2002). Penduga parameter dengan metode *fisher scoring* pada iterasi ke-$t+1$ dalam proses iterasi t= 0, 1, 2 ,... adalah sebagai berikut :

$$\widehat{\boldsymbol{\beta}}^{m+1} = \widehat{\boldsymbol{\beta}}^{t} + \boldsymbol{I}^{-1}(\boldsymbol{\beta}^{m})\boldsymbol{S}(\boldsymbol{\beta}^{m}) \tag{2.19}$$

dengan

$\widehat{\boldsymbol{\beta}}^{m}$ dan $\widehat{\boldsymbol{\beta}}^{m+1}$: vektor untuk $\boldsymbol{\beta}$ pada iterasi ke-*t* dan ke-*t* + 1

$\boldsymbol{I}^{-1}(\boldsymbol{\beta}^{m})$: matriks informasi yang berisi negatif ekspektasi dari turunan kedua *ln-likelihood* terhadap $\boldsymbol{\beta}^{m}$

$\boldsymbol{S}(\boldsymbol{\beta}^{m})$: vektor turunan pertama ln-*likelihood* terhadap $\boldsymbol{\beta}^{m}$

Dari iterasi tersebut akan diperoleh penduga maksimum *likelihood* untuk $\widehat{\boldsymbol{\beta}}$ dan $\widehat{\boldsymbol{W}}$ yang dinotasikan dengan $\widehat{\boldsymbol{\beta}}_{ML}$ dan $\widehat{\boldsymbol{W}}$

$$\widehat{\boldsymbol{\beta}}_{ML} = \left(\boldsymbol{X}^{t}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{t}\widehat{\boldsymbol{W}}\hat{z} \tag{2.20}$$

dengan merupakan vektor yang setiap elemen ke-*i* bernilai

$$z_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right) + \frac{y_i-\pi_i}{\pi_i(1-\pi_i)} \tag{2.21}$$

2.4 Iterasi *Method of Scoring*

Iterasi *method of scoring* adalah salah satu iterasi dari metode nonlinier *maximum likelihood* untuk mendapatkan estimasi prameter β yang merupakan bagian dari metode *fisher scoring* dengan

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{m} - \left(E\left(\frac{\partial^2 \ell}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{m}}\Big|_{\boldsymbol{\beta}^{m}}\right)\right)^{-1}\frac{\partial \ell}{\partial\boldsymbol{\beta}}\Big|_{\boldsymbol{\beta}^{m}} \tag{2.22}$$

Selanjutnya PDF dari y_i yang diberikan oleh $\boldsymbol{X}_i$**,** $\boldsymbol{\beta}$**,** dan σ^2 berikut :

$$\begin{aligned} f(y_i|\boldsymbol{X}_i,\boldsymbol{\beta},\sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}}exp\left(-\frac{1}{2\sigma^2}(y_i - f(\boldsymbol{X}_i,\boldsymbol{\beta}))^2\right) \\ &= L_i(\boldsymbol{\beta},\sigma^2) \end{aligned} \tag{2.23}$$

selanjutnya digunakan sifat PDF, yang menyatakan

$$\int_{-\infty}^{\infty} f(y_i|\boldsymbol{X}_i,\boldsymbol{\beta},\sigma^2)d\,y_i = 1 \tag{2.24}$$

maka diperoleh

$$\frac{\partial}{\partial\beta}\int_{-\infty}^{\infty} f(y_i|\boldsymbol{X}_i,\boldsymbol{\beta},\sigma^2)d\,y_i = 0 \tag{2.25}$$

ana diketahui Persamaan (2.23), maka dapat dibentuk

$$\ell_i = \ln L_i(\boldsymbol{\beta},\sigma^2) = \ln\, f(y_i|\boldsymbol{X}_i,\boldsymbol{\beta},\sigma^2) \tag{2.26}$$

$$\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} = \frac{1}{L_i(\boldsymbol{\beta}, \sigma^2)} \frac{\partial L_i(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}}$$

$$= \frac{1}{f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2)} \frac{\partial f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \quad (2.27)$$

dari Persamaan (2.27) diperoleh

$$\int_{-\infty}^{\infty} \frac{\partial f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} d\, y_i = 0$$

$$= \int_{-\infty}^{\infty} \frac{\partial f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} \frac{f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2)}{f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2)} d\, y_i$$

$$= \int_{-\infty}^{\infty} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2) d\, y_i \quad (2.28)$$

Selanjutnya turunan parsial pertama Persamaan (2.28) terhadap $\boldsymbol{\beta}^m$ dan menyakaman persamaanya dengan nol sehingga diperoleh,

$$\frac{\partial}{\partial \boldsymbol{\beta}^t} \left(\int_{-\infty}^{\infty} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2) d\, y_i \right) = 0$$

$$= \int_{-\infty}^{\infty} \left(\left(\frac{\partial^2 \ell_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^m} + \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}^m} \right) f(y_i|\boldsymbol{X}_i, \boldsymbol{\beta}, \sigma^2) \right) dy_i$$

$$= E \left(\frac{\partial^2 \ell_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^m} + \frac{\partial \ell_i}{\partial \boldsymbol{\beta}} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}^m} \right)$$

atau dapat dituliskan sebagai

$$E \left(\frac{\partial^2 \ell_i}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^m} \right) = -E \left(\frac{\partial \ell_i}{\partial \boldsymbol{\beta}} \frac{\partial \ell_i}{\partial \boldsymbol{\beta}^m} \right) \quad (2.29)$$

2.5 Multikolinieritas

Istilah multikolinieritas diciptakan oleh Ragner Frish pada tahun 1934 dalam bukunya yang berjudul: *Statistical Confluence Analysis by Means of Complete Regression Systems*. Istilah itu menyatakan adanya hubungan linier yang sempurna atau eksak (*perfect or exact*) di antara variabel-variabel prediktor dalam model regresi. Istilah kolinieritas sendiri berarti hubungan linier tunggal (*single linear relationship*), sedangkan multikolinieritas menunjukkan adanya lebih dari satu hubungan linier yang sempurna. Dalam praktik sering tidak dibedakan baik satu hubungan atau lebih dipergunakan istilah multikolinieritas (Supranto, 2005).

ltikolinieritas antara variabel prediktor $\boldsymbol{X}$ akan mengakibatkan determinan $^t\boldsymbol{X}$ pada estimator *ordinary least square* maupun *maximum likelihood* ti nol sehingga menjadi singular. Draper dan Smith (1992) menyatakan bahwa hal ini dapat diketahui dari matriks korelasi hasil pemusatan dan penskalaan matriks $\boldsymbol{X}$ sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{LS} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{Y}$$

$$= \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix} \qquad (2.30)$$

dengan r_{12} adalah koefisien korelasi antara $\boldsymbol{X}_1$ dan $\boldsymbol{X}_2$. Nilai r_{12} yang membesar akan menyebabkan determinan matriks $\boldsymbol{X}^t\boldsymbol{X}$ mendekati nol (multikolinieritas mendekati sempurna) atau sama dengan nol (multikolinieritas sempurna) sehingga mengakibatkan matriks menjadi singular (tidak memiliki invers).

Ada beberapa cara untuk mendeteksi adanya multikolinearitas pada model regresi, antara lain dengan menghitung matriks korelasi. Prosedur ini merupakan cara yang paling sederhana dan paling mudah untuk mendeteksi multikolinearitas yaitu pemeriksaan elemen diagonal r_{ij}. Jika variabel prediktor x_i dan x_j berkorelasi linear, maka $|r_{ij}|$akan mendekati satu (Myers,1990).

Menurut Setiawan dan Kusrini (2010), salah satu ukuran untuk menguji adanya multikolinieritas adalah *Variance Inflation Factors* (VIF). VIF merupakan elemen diagonal dari matriks $\boldsymbol{X}^t\boldsymbol{X}$.

$$(\boldsymbol{X}^t\boldsymbol{X})^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix} \qquad (2.31)$$

$$VIF_j = diag(\boldsymbol{X}^t\boldsymbol{X}) = \frac{1}{1-\boldsymbol{R}_j^2} \qquad (2.32)$$

Pengujian multikolinieritas juga dapat dilakukan dengan menghitung nilai *Tolerance* (TOL) dengan persamaan

$$TOL_j = \frac{1}{VIF_j} \qquad (2.33)$$

Nilai $TOL < 0.1$ mengindikasikan bahwa terjadi multikolinieritas antar variabel prediktor. Sedangkan nilai VIF dari estimator *generalized ridge regression* dapat dihitung melalui persamaan

$$VIF = diag\left(\frac{1}{n-1}(\boldsymbol{X}^t\boldsymbol{X}) - \boldsymbol{DKD}^t\right)^{-1}\left(\frac{1}{n-1}(\boldsymbol{X}^t\boldsymbol{X})\right)\left(\frac{1}{n-1}(\boldsymbol{X}^t\boldsymbol{X}) - \boldsymbol{DKD}^t\right)^{-1} \qquad (2.34)$$

dengan $\boldsymbol{K}$ adalah matriks yang elemen diagonalnya merupakan parameter ridge $k \geq$ ... nyatakan suatu matriks ortogonal dengan $\boldsymbol{D} = \boldsymbol{D}^{-1}$sedemikian sehingga ... dan $\boldsymbol{D}^t\boldsymbol{C}\boldsymbol{D} = \Lambda$, dengan $\boldsymbol{C} = \boldsymbol{X}^t\boldsymbol{X}$ dan Λ merupakan matriks $p \times p$ yang ... diagonal utamanya merupakan nilai eigen dari matriks $\boldsymbol{X}^t\boldsymbol{X}$.

2.6 Metode *Ridge*

Regresi *ridge* adalah salah satu metode yang dapat digunakan untuk mengatasi kasus multikolinieritas. Penduga *ridge* pertama kali dirumuskan oleh Hoerl dan Kennard (1970) untuk menangani masalah multikolinearitas pada regresi linear dengan menyajikan nilai regresi linear berganda yang mengakibatkan matriks $\boldsymbol{X}^t\boldsymbol{X}$ memiliki determinan mendekati nol yang menghasilkan nilai penduga parameter yang tidak stabil. Penduga parameter regresi *ridge* menggunakan metode *Least Square* (LS) yaitu dengan menambahkan bilangan positif kecil (k) pada diagonal matriks $\boldsymbol{X}^t\boldsymbol{X}$, sehingga bias yang terjadi dapat dikendalikan. Bilangan positif kecil (k) bernilai antara 0 dan 1, sehingga penduga regresi *ridge* akan bias terhadap parameter $\boldsymbol{\beta}$, tetapi cenderung lebih stabil (Sunyoto, 2009) .

Menurut Hoerl dan Kennard (1970), estimasi ridge untuk koefisien regresi dapat diperoleh dengan menyelesaikan suatu bentuk dari persamaan normal regresi. Penduga *ridge* dapat dituliskan sebagai berikut:

$$\widehat{\boldsymbol{\beta}}_{RR} = (\boldsymbol{X}^t\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^t\boldsymbol{Y} \qquad (2.35)$$

Pada dasarnya penduga *ridge* merupakan metode kuadrat terkecil. Perbedaannya adalah bahwa pada metode regresi *ridge*, nilai variabel bebasnya ditransformasikan dahulu melalui prosedur *centering* dan *rescaling* (Wasilaine, dkk, 2014).

Kemudian pada tahun 1984, R.L Schaefer, L.D Roi dan R.A Wolfe mengembangkan penduga *ridge* pada model regresi logistik untuk menangani masalah multikolinearitas. Penduga *ridge* untuk model regresi logistik diperoleh dengan metode Lagrange untuk meminimumkan fungsi *Weighted Sum of Square Error* (WSSE) berikut:

$$\boldsymbol{WSSE} = (\boldsymbol{Y} - \boldsymbol{X\beta})^t\boldsymbol{W}(\boldsymbol{Y} - \boldsymbol{X\beta}) \qquad (2.36)$$

Parameter penting yang membedakan regresi ridge dari metode kuadrat terkecil adalah k. Parameter ridge k yang relatif kecil ditambahkan pada diagonal utama matriks $\boldsymbol{X}^t\boldsymbol{X}$, sehingga koefisien estimator regresi ridge dipenuhi dengan besarnya parameter ridge k.

imator ridge diperoleh dengan meminimumkan jumlah kuadrat *error* del

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (2.37)$$

atau

$$\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X\beta} \qquad (2.38)$$

dengan menggunakan metode pengali Lagrange yang meminimumkan fungsi

$$\boldsymbol{\varepsilon}^t\boldsymbol{\varepsilon} = (\boldsymbol{Y} - \boldsymbol{X\beta}_{RR})^t(\boldsymbol{Y} - \boldsymbol{X\beta}_{RR}) \qquad (2.39)$$

dengan syarat pembatas

$$\boldsymbol{\beta}_{RR}^t\boldsymbol{\beta}_{RR} - c^2 = 0$$

$$\boldsymbol{G} = (\boldsymbol{Y} - \boldsymbol{X\beta}_{RR})^t(\boldsymbol{Y} - \boldsymbol{X\beta}_{RR}) + k(\boldsymbol{\beta}_{RR}^t\boldsymbol{\beta}_{RR} - c^2) \qquad (2.40)$$

yang memenuhi syarat $\frac{\partial \boldsymbol{G}}{\partial \boldsymbol{\beta}_{RR}}\Big| \widehat{\boldsymbol{\beta}}_{RR} = 0$

$$-2\boldsymbol{X}^t\boldsymbol{Y} + 2\boldsymbol{X}^t\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{RR} + 2k\boldsymbol{I}\widehat{\boldsymbol{\beta}}_{RR} = 0$$

$$\widehat{\boldsymbol{\beta}}_{RR} = (\boldsymbol{X}^t\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^t\boldsymbol{Y} \qquad (2.41)$$

dengan $\widehat{\boldsymbol{\beta}}_{RR} = (\boldsymbol{X}^t\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^t\boldsymbol{Y}$ dengan $0 \leq k \leq \infty$, itulah yang disebut sebagai estimator regresi ridge. $k \geq 0$ adalah nilai konstan yang dipilih sebagai indeks dari kelas estimator.

Terdapat beragam metode yang telah dikemukakan peneliti sebelumnya dalam pemilihan nilai tetapan k. Berikut metode pemilihan nilai tetapan yang sering digunakan dalam penelitian :

1. Hoerl dan Kennard dalam Muliati (2018)

   $$k_{HK} = \frac{s^2}{\hat{\alpha}_{max}^2}$$

   Merupakan penduga *ridge* yang paling klasik diusulkan oleh Hoerl dan Kennard, di mana $\hat{\alpha}_{max}^2$ merupakan elemen maksimum dari $\boldsymbol{\gamma\beta}_{ML}$ dimana $\boldsymbol{\gamma}$ merupakan vektor eigen dari $\boldsymbol{X}^t\boldsymbol{WX}$ dan s^2 adalah variansi.

2. Schaefer et al. dalam Muliati (2018)

   $$k_{SRW} = \frac{1}{\hat{\alpha}_{max}^2}$$

   Merupakan sebuah versi modifikasi dari Hoerl dan Kennard untuk regresi logistik yang menunjukkan bahwa nilai optimal dari k sama dengan $\frac{1}{\hat{\alpha}_{max}^2}$ bukan $\frac{s^2}{\hat{\alpha}_{max}^2}$.

sson and Shukur dalam Muliati (2018), menggeneralisasi pendekatan yang
eda untuk menduga parameter *ridge* yang diusulkan oleh Muniz et.al yang
yatakan bahwa nilai tetapan K13 dan K15 merupakan pilihan terbaik

sebagai parameter *ridge*. K13 dan K15 memiliki MSE yang lebih kecil ketika terjadi korelasi yang tinggi maupun rendah.

$$\text{K13}= \prod_{j=1}^{p}\left(\frac{1}{q_j}\right)^{\frac{1}{p}}$$

$$\text{K15}= median\left(\frac{1}{q_j}\right)$$

Adapun model regresi probit dengan metode *ridge* menggunakan parameter ridge

$$k = max\left(\frac{1}{q_j}\right) \tag{2.42}$$

$$q_j = \frac{\lambda_{max}}{(n-p)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}_j^2} \tag{2.43}$$

2.7 Kemiskinan

Masalah kemiskinan merupakan salah satu persoalan mendasar yang menjadi pusat perhatian pemerintah di negara manapun. Salah satu aspek penting untuk mendukung Strategi Penanggulangan Kemiskinan adalah tersedianya data kemiskinan yang akurat. Pengukuran kemiskinan yang dapat dipercaya dapat menjadi instrumen tangguh bagi pengambil kebijakan dalam memfokuskan perhatian pada kondisi hidup orang miskin. Data kemiskinan yang baik dapat digunakan untuk mengevaluasi kebijakan pemerintah terhadap kemiskinan, membandingkan kemiskinan antar waktu dan daerah, serta menentukan target penduduk miskin dengan tujuan untuk memperbaiki kondisi mereka (BPS, 2018).

Kemiskinan telah menjadi masalah di hampir semua negara, baik negara maju atau negara yang sedang berkembang. Tingkat kekompleksitas tiap negara berbeda dalam menyelesaikan masalah kemiskinan. Indonesia sebagai salah satu negara berkembang, angka kemiskinan masih cukup tinggi. Pemerintah melalui Badan Pusat Statistik (BPS) membuat kriteria kemiskinan, agar dapat menyusun secara lengkap pengertian kemiskinan sehingga dapat diketahui dengan pasti jumlahnya dan cara tepat menanggulanginya. Berikut 14 kriteria yang menjadikan sebagai indikator keluarga miskin menurut standar BPS:

1. Luas lantai bangunan tempat tinggal kurang dari 8 per orang.

is lantai tempat tinggal terbuat dari tanah/bambu/kayu murahan.

is dinding tempat tinggal dari bambu/rumbia/kayu berkualitas dah/tembok tanpa diplester.

4. Tidak memiliki fasilitas buang air besar/bersama-sama dengan rumah tangga lain.
5. Sumber penerangan rumah tangga tidak menggunakan listrik.
6. Sumber air minum berasal dari sumur/mata air tidak terlindung/sungai/air hujan.
7. Bahan bakar untuk memasak sehari-hari adalah kayu bakar/arang/minyak tanah.
8. Hanya mengkonsumsi daging/susu/ayam dalam satu kali seminggu.
9. Hanya membeli satu stel pakaian baru dalam setahun.
10. Hanya sanggup makan sebanyak satu/dua kali dalam sehari.
11. Tidak sanggup membayar biaya pengobatan di puskesmas/poliklinik.
12. Sumber penghasilan kepala rumah tangga adalah: petani dengan luas lahan 500 m^2, buruh tani, nelayan, buruh bangunan, buruh perkebunan dan atau pekerjaan lainnya dengan pendapatan dibawah Rp 600.000,-per bulan.
13. Pendidikan tertinggi kepala rumah tangga: tidak sekolah/tidak tamat SD/tamat SD.
14. Tidak memiliki tabungan/barang yang mudah dijual dengan minimal Rp. 500.000,- seperti sepeda motor kredit/non kredit, emas, ternak, kapal motor, atau barang modal lainnya.

Badan Pusat Statistik (BPS) mengukur Kemiskinan dengan 3 indikator yakni P_0, P_1, dan P_2. Penjelasan masing-masing dapat dilihat di sebagai berikut :

1. Persentase Penduduk Miskin (*Headcount Index*-P_0) yaitu persentase penduduk miskin yang berada di bawah Garis Kemiskinan. *Headcount Index* secara sederhana mengukur proporsi penduduk yang dikategorikan miskin.
2. Indeks Kedalaman Kemiskinan (*Poverty Gap Index*–P_1) merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis kemiskinan.
3. Indeks Keparahan Kemiskinan (*Poverty Severity Index*–P_2) memberikan gambaran mengenai penyebaran pengeluaran di antara penduduk miskin.