ESTIMASI PARAMETER MODEL REGRESI PROBIT RIDGE DENGAN METODE *MAXIMUM LIKELIHOOD*

SKRIPSI

OLEH
FINA AMALIA ISTIQOMAH
NIM. 10610101

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2015

ESTIMASI PARAMETER MODEL REGRESI PROBIT RIDGE DENGAN METODE *MAXIMUM LIKELIHOOD*

SKRIPSI

Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Sains (S.Si)

Oleh
Fina Amalia Istiqomah
NIM. 10610101

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2015

ESTIMASI PARAMETER MODEL REGRESI PROBIT RIDGE DENGAN METODE *MAXIMUM LIKELIHOOD*

SKRIPSI

Oleh
Fina Amalia Istiqomah
NIM. 10610101

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal 29 Desember 2014

Pembimbing I,	Pembimbing II,
Fachrur Rozi, M.Si NIP. 19800527 200801 1 012	Dr. H. Ahmad Barizi, MA NIP. 19731212 199803 1 001

Mengetahui,
Ketua Jurusan Matematika

Dr. Abdussakir, M.Pd
NIP. 19751006 200312 1 001

ESTIMASI PARAMETER MODEL REGRESI PROBIT RIDGE DENGAN METODE *MAXIMUM LIKELIHOOD*

SKRIPSI

Oleh
Fina Amalia Istiqomah
NIM. 10610101

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Sains (S.Si)
Tanggal 07 Januari 2015

Penguji Utama	:	Abdul Aziz, M.Si	....
Ketua Penguji	:	Dr. Sri Harini, M.Si	....
Sekretaris Penguji	:	Fachrur Rozi, M.Si	....
Anggota Penguji	:	Dr. H. Ahmad Barizi, MA	....

Mengetahui,
Ketua Jurusan Matematika

Dr. Abdussakir, M.Pd
NIP. 19751006 200312 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Fina Amalia Istiqomah

NIM : 10610101

Jurusan : Matematika

Fakultas : Sains dan Teknologi

Judul Skripsi : Estimasi Parameter Model Regresi Probit Ridge dengan Metode *Maximum Likelihood*.

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 11 Januari 2015

Yang membuat pernyataan,

Fina Amalia Istiqomah

NIM. 10610101

MOTO

اِلٰهِيْ أَنْتَ مَقْصُوْدِي وَ رِضَاكَ مَطْلُوْبِيْ

"Tuhanku, Engkaulah tujuan hamba dan ridlo-Mu-lah yang hamba cari"

وَلِلَّهِ ٱلْمَشْرِقُ وَٱلْمَغْرِبُ ۚ فَأَيْنَمَا تُوَلُّوا۟ فَثَمَّ وَجْهُ ٱللَّهِ ۚ إِنَّ ٱللَّهَ وَٰسِعٌ عَلِيمٌ ﴿١١٥﴾

"Dan kepunyaan Allah-lah timur dan barat, maka kemanapun kamu menghadap di situlah wajah Allah. Sesungguhnya Allah Maha Luas (rahmat-Nya) lagi Maha Mengetahui" (QS. al-Baqarah/2:115).

PERSEMBAHAN

Skripsi ini peneliti persembahkan untuk:

Almarhum ayahanda Abdul Ghofur, ibunda Nur Saidah,

kakak tersayang almarhum M. Nuruddin, M. Fauzan Abdillah, M. Syahdani Achdan, adik tercinta Shabrina Achda Laily,

serta Rudiansyah yang selalu memberikan dukungan dan nasihat tiada henti kepada peneliti.

KATA PENGANTAR

بسم الله الرحمن الرحيم

Assalamu'alaikum Wr.Wb.

Alhamdulillahirobbil'alamiin, puji syukur kepada Allah Swt., atas rahmat, taufiq, dan hidayah-Nya, peneliti dapat menyelesaikan penulisan skripsi sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains dalam bidang matematika di Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.

Ucapan terima kasih penulis sampaikan seiring doa dan harapan kepada semua pihak yang telah berpartisipasi dan membantu penyelesaian skripsi ini. Ucapan terima kasih yang sebesar-besarnya peneliti sampaikan kepada:

1. Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. drh. Hj. Bayyinatul Muchtaromah, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Dr. Abdussakir, M.Pd, selaku ketua Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Fachrur Rozi, M.Si, selaku dosen pembimbing I yang telah banyak memberikan arahan dan bimbingan dengan sabar selama penulisan skripsi ini.
5. Dr. H. Ahmad Barizi, MA, selaku dosen pembimbing II yang telah memberikan arahan, bimbingan, dan saran dalam pemilihan ayat al-Quran untuk penulisan skripsi ini.

6. Segenap sivitas akademika Jurusan Matematika, terutama seluruh dosen, terima kasih atas segenap ilmu dan bimbingannya.
7. Almarhum ayahanda Abdul Ghofur dan ibunda Nur Saidah yang selalu mendoakan, memberikan kasih sayang, semangat, dan motivasi kepada peneliti tanpa kenal lelah. Kakak tersayang Almarhum M. Nuruddin, M. Fauzan Abdillah, M. Syahdani Achdan, adik tercinta Shabrina Achda Laily, dan Rudiansyah yang selalu memberikan semangat, motivasi, dan kasih sayang hingga selesainya skripsi ini.
8. Teman-teman mahasiswa Jurusan Matematika angkatan 2010, khususnya Wardatul Jannah, Mayasaroh, Mahmuda, Silvia Anggraini, Lailatul Mubarokah, Alfi Fadliana, Ririn Zulaikah, dan Eva Kurniasih, terima kasih atas segala pengalaman berharga dan kenangan terindah saat menuntut ilmu bersama.
9. Semua pihak yang tidak dapat disebutkan satu-persatu, yang telah membantu peneliti dalam menyelesaikan penulisan skripsi ini.

Semoga Allah Swt. senantiasa melimpahkan rahmat dan karunia-Nya. Akhirnya, peneliti berharap semoga dengan rahmat dan izin Allah, mudah-mudahan skripsi ini dapat memberikan banyak manfaat bagi peneliti dan bagi pembaca. *Amin ya Robbal 'alamiin…*

Wassalamu'alaikum Wr.Wb.

Malang, Januari 2015

Peneliti

DAFTAR ISI

HALAMAN JUDUL

HALAMAN PENGAJUAN

HALAMAN PERSETUJUAN

HALAMAN PENGESAHAN

HALAMAN PERNYATAAN KEASLIAN TULISAN

HALAMAN MOTO

HALAMAN PERSEMBAHAN

KATA PENGANTAR viii

DAFTAR ISI x

DAFTAR TABEL xii

DAFTAR LAMPIRAN xiii

ABSTRAK xiv

ABSTRACT xv

ملخص xvi

BAB I PENDAHULUAN

1.1 Latar Belakang 1
1.2 Rumusan Masalah 5
1.3 Tujuan Penelitian 5
1.4 Batasan Masalah 6
1.5 Manfaat Penelitian 6
1.6 Sistematika Penulisan 7

BAB II KAJIAN PUSTAKA

2.1 Distribusi Normal dan Normal Standar 9
2.2 Ekspektasi dan Variansi 10
2.3 Estimasi Parameter 11
2.4 Metode *Maximum Likelihood* 12
2.5 Iterasi *Method of Scoring* 13
2.6 Regresi Probit 16
2.7 Pemusatan dan Penskalaan Data dalam Bentuk Matriks Korelasi 18
2.8 Multikolinieritas 20
2.9 Bentuk Kanonik Model Regresi 22
2.10 Regresi Ridge 23

2.11 *Generalized Linear Model* (GLM) 25
2.12 Konsep Analisis Regresi dalam Al-Quran 26

BAB III METODE PENELITIAN

3.1 Pendekatan Penelitian 30
3.2 Sumber Data 30
3.3 Variabel Penelitian 30
3.4 Metode Analisis 31

BAB IV PEMBAHASAN

4.1 Estimasi Parameter Model Regresi Probit 33
4.1.1 Deskripsi ke dalam Bentuk *Generalized Linear Model* (GLM) 34
4.1.2 Estimasi dengan Metode *Maximum Likelihood* 37
4.2 Estimasi Parameter Model Regresi Probit Ridge 40
4.3 Aplikasi pada Data yang Mengandung Multikolinieritas 44
4.3.1 Uji Multikolinieritas 44
4.3.2 Model Regresi Probit Ridge dari Data dan Estimasi Parameter 45
4.3.3 Uji Multikolinieritas dengan Parameter Ridge 48
4.3.4 Analisis Hasil Estimasi Curah Hujan 49
4.4 Konsep Multikolinieritas dalam Al-Quran 53

BAB V PENUTUP

5.1 Kesimpulan 55
5.2 Saran 56

DAFTAR PUSTAKA 57

LAMPIRAN-LAMPIRAN 59

RIWAYAT HIDUP 65

DAFTAR TABEL

Tabel 4.1 Nilai Koefisien Korelasi 44

Tabel 4.2 Estimasi Probabilitas Terjadinya Hujan Tahun 2007 49

Tabel 4.3 Estimasi Probabilitas Terjadinya Hujan Tahun 2008-2009 50

DAFTAR LAMPIRAN

Lampiran 1 Data Curah Hujan, Temperatur, dan Kelembaban Udara di Karangploso Malang Tahun 2007-2008 59

Lampiran 2 Data Curah Hujan, Temperatur, dan Kelembaban Udara di Karangploso Malang Tahun 2009 60

Lampiran 3 Hasil Normalisasi Data Temperatur Udara 61

Lampiran 4 Program MATLAB untuk Estimasi Parameter Model Regresi Probit Ridge dengan Metode *Maximum Likelihood* 62

Lampiran 5 Program MATLAB untuk Menentukan Nilai VIF 64

ABSTRAK

Istiqomah, Fina Amalia. 2015. **Estimasi Parameter Model Regresi Probit Ridge dengan Metode *Maximum Likelihood***. Skripsi. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Pembimbing: (I) Fachrur Rozi, M.Si. (II) Dr. H. Ahmad Barizi, MA.

Kata Kunci: regresi probit ridge, *maximum likelihood*, multikolinieritas

Model regresi probit merupakan suatu model regresi dimana variabel respon bersifat kualitatif, yang biasanya menunjukkan ada atau tidaknya kriteria suatu atribut, sehingga seringkali menggunakan nilai 0 atau 1 untuk menunjukkan ada tidaknya kriteria yang dimaksud. Terdapat banyak cara yang dapat digunakan untuk mengestimasi parameter model regresi probit, salah satunya dengan menggunakan metode *maximum likelihood*. Namun ketika terdapat multikolinieritas antar variabel prediktor, maka variansi akan semakin membesar sehingga estimasi dengan metode tersebut menjadi tidak efisien. Oleh karena itu, dalam penelitian ini akan digunakan metode *maximum likelihood* untuk mendapatkan estimator dari parameter model regresi probit ridge sehingga multikolinieritas dapat teratasi dengan baik.

Estimasi parameter model regresi probit ridge ini didapatkan dengan cara mengestimasi terlebih dahulu parameter model regresi probit dengan mendeskripsikan ke dalam bentuk *Generalized Linear Model* (GLM), kemudian dilanjutkan dengan estimasi dengan metode *maximum likelihood*. Estimator yang telah didapatkan kemudian digunakan untuk mengestimasi parameter regresi probit ridge, dengan estimasi parameter

$$\hat{\beta}_{RR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\hat{\beta}_{ML} \text{ dan } \hat{\beta}_{ML} = \left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right)$$

Aplikasi pada data curah hujan di Karangploso Malang menghasilkan model regresi $y_i^* = -1,355 + 1,268X_{1i} + 0,021X_{2i}$, dengan y_i^* merupakan variabel laten. Perhitungan VIF menghasilkan $\text{VIF} = 0,6036$, sehingga multikolinieritas antara temperatur (X_1) dengan kelembaban udara (X_2) dapat diatasi.

ABSTRACT

Istiqomah, Fina Amalia. 2015. **The Estimation of Parameters for Probit Ridge Regression Model with Maximum Likelihood Method**. Thesis. Department of Mathematics, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang, Advisors: (I) Fachrur Rozi, M.Si (II) Dr. H. Ahmad Barizi, MA.

Keywords: probit ridge regression, maximum likelihood, multicollinearity

Probit regression model is a regression model where the response variable is qualitative, which usually indicates the presence or absence of an attribute criteria, so often use 0 or 1 to indicate whether or not the criteria are intended. There are many ways that can be used to estimate the parameters of the probit regression model, one of them is using maximum likelihood method. But when there is multicollinearity among predictor variables, the variance will be enlarged so that the estimation with that method becomes inefficient. Therefore, this research will use maximum likelihood method to obtain the estimator of probit ridge regression model that can fix the multicollinearity well.

The estimation of parameters for probit ridge regression model is obtained by estimating the parameters of probit regression model firs then describe it in the Generalized Linear Model (GLM), followed by the estimation with maximum likelihood method. Estimator that has been obtained then used to estimate the parameters of probit ridge regression, with the estimator of parameters

$$\hat{\beta}_{RR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\,\beta_{ML} \text{ and } \beta_{ML} = \left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right)$$

The application on rainfall data in Karangploso Malang obtain the regression model $y_i^* = -1,355 + 1,268X_{1i} + 0,021X_{2i}$, with y_i^* is a latent variable. The result of the calculation of VIF is $\mathrm{VIF} = 0,6036$, so that the multicollinearity between the temperature (X_1) and the air humidity (X_2) can be resolved.

ملخص

إستقامة، فينا عمليّة. ٢٠١٥. **تقدير المعلمة على نموذج الانحدار بروبيد ريدج بإستغرام طريقة *Maximum Likelihood***. النهائية بحث جامعي. قسم الرياضيات، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج، مشرف: (١) فحرالرازي ، الماجستير (٢) الدكتور الحاج أحمد باريزي، الماجستير.

كلمات البحث: الانحدار بروبيد، *maximum likelihood* ، الخطية المتعددة

نموذج إنحدار هو نموذج الانحدار الذي متغيره الاستجابة هو نوعي، والتي عادة يدل على أو عدم معيار من السمة، وغالبا تستخدم قيمة ٠ أو ١ لبيان وجود أو عدم المعيار. هناك العديد من الطرق التي يمكن استخدامها لتقدير المعلمات من النماذج انحدار بروبيد، أحدها هي باستخدام *Maximum Likelihood*. ولكن عندما يكون هناك الخطية المتعددة بين المتغيرات المتنبئ، سيتم توسيع التباين بحيث تقدير مع أسلوب يصبح غير فعال. لذلك، سيتم استخدامها في هذه الدراسة للحصول على أقصى قدر من طريقة احتمال المقدرات المعلمة ريدج نموذج الانحدار الاحتمالية التي يمكن التغلب عليها مع الخطية المتعددة جيد.

يتم الحصول على معلمة نموذج الانحدار, بتقدير المعلمات من نموذج الانحدار بروبيد موضح في شكل المعمم النموذج الخطي (GLM) أولا، تليها طريقة أقصى تقدير احتمال. مقدر التي تم الحصول عليها، نستخدمها لتقدير بطريقة *Maximum Likelihood* ، وتقديرات المعلمة

$$\beta_{ML}=\left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right) \text{ و } \hat{\beta}_{RR}=\left(X'\hat{W}X+kI\right)^{-1}X'\hat{W}X\,\beta_{ML}$$

التطبيقات على بيانات هطول الأمطار في كرنفلاصه مالانج إنتاج نموذج الانحدار $y_i^*=-1,355+1,268X_{1i}+0,021X_{2i}$ ، مع متغير y_i^* الكامنة. VIF المنتجات حساب $\text{VIF}=0,6036$ ، بحيث الخطية المتعددة بين درجة الحرارة (X_1) مع الرطوبة (X_2) يمكن تحليلها.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi merupakan salah satu metode statistik yang sering digunakan dalam kehidupan sehari-hari. Analisis regresi adalah suatu teknik yang digunakan untuk membangun suatu persamaan/model yang menghubungkan antara variabel respon (Y) dengan satu atau lebih variabel prediktor (X) dan sekaligus untuk menentukan nilai ramalan atau dugaannya. Pendugaan ini penting untuk mengetahui dampak yang terjadi akibat perubahan suatu variabel terhadap variabel lain, sehingga dapat dilakukan antisipasi dalam menghadapi dampak tersebut.

Sebuah model regresi dikatakan baik, jika memenuhi asumsi-asumsi sederhana yang sering disebut sebagai asumsi klasik. Asumsi yang pertama yaitu nilai rata-rata faktor kesalahan nol, atau $E(\varepsilon_i)=0$, untuk $i=1,2,3,...,n$. Asumsi kedua yaitu $Var(\varepsilon_i)=E(\varepsilon_i^2)=\sigma^2$ sama untuk semua faktor kesalahan (asumsi homoskedastis). Asumsi ketiga, tidak ada autokorelasi antara faktor kesalahan, yang berarti $kov(\varepsilon_i,\varepsilon_j)=0, i \neq j$. Asumsi keempat, variabel prediktor $X_1, X_2,..., X_k$ konstan dalam *sampling* yang terulang dan bebas terhadap kesalahan pengganggu ε_i. Asumsi kelima, tidak terjadi multikolinieritas antara variabel prediktor. Asumsi terakhir yaitu $\varepsilon_i \sim N(0,\sigma^2)$, yang berarti kesalahan pengganggu mengikuti distribusi normal dengan rata-rata nol dan varian σ^2 (Supranto, 2005a:29).

Model regresi yang sering dipelajari, baik variabel respon maupun variabel prediktor bersifat bilangan atau kuantitatif. Namun, ada kalanya variabel prediktor bersifat kualitatif dan variabel respon bersifat kuantitatif ataupun sebaliknya. Variabel kualitatif ini, yang sering dikenal sebagai variabel buatan atau variabel *dummy* atau variabel boneka (*dummy variable*), mempunyai beberapa istilah dalam literatur, seperti variabel indikator, variabel biner, variabel kategori, dan variabel dikotomi. Variabel-variabel kualitatif seperti itu biasanya menunjukkan ada atau tidaknya "kualitas" suatu atribut, seperti laki-laki atau perempuan, hitam atau putih, muslim atau non muslim, warga negara atau non warga negara.

Salah satu metode "kuantifikasi" atribut-atribut ini adalah dengan membentuk variabel-variabel artifisial yang memperhitungkan nilai-nilai 0 atau 1, 0 menunjukkan ketiadaan sebuah atribut dan 1 menunjukkan keberadaan (atau kepemilikan) atribut itu. Misalnya, 1 mungkin menunjukkan bahwa seseorang adalah wanita dan 0 mungkin menunjukkan pria. 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, 2006b:76). Jika variabel yang bersifat *dummy* adalah variabel respon, maka salah satu pendekatan model yang dapat digunakan adalah model regresi probit.

Dalam model regresi probit, variabel respon (Y) merupakan variabel *dummy* yang berdistribusi Bernoulli, yang berarti mempunyai dua nilai, yaitu 1 jika sukses dan 0 jika gagal. Konsep dari distribusi Bernoulli ini juga mempunyai relevansi dengan firman Allah Swt. dalam al-Quran surat al-Baqarah/2:137 yang berbunyi:

فَإِنْ ءَامَنُواْ بِمِثْلِ مَآ ءَامَنتُم بِهِۦ فَقَدِ ٱهْتَدَواْ ۖ وَّإِن تَوَلَّوْاْ فَإِنَّمَا هُمْ فِي شِقَاقٍ ۖ فَسَيَكْفِيكَهُمُ ٱللَّهُ ۚ وَهُوَ ٱلسَّمِيعُ ٱلْعَلِيمُ ﴿١٣٧﴾

"Maka jika mereka beriman kepada apa yang kalian telah beriman kepadanya, sungguh mereka telah mendapat petunjuk; dan jika mereka berpaling, sesungguhnya mereka berada dalam permusuhan (dengan kamu), maka Allah akan memelihara kamu dari mereka. dan Dia-lah yang Maha Mendengar lagi Maha Mengetahui" (QS. al-Baqarah/2:137).

Dalam ayat di atas, Allah Swt. berfirman, *"Maka jika mereka beriman,"* yakni orang-orang kafir dan ahli kitab serta lain-lainnya mau beriman, *"kepada apa yang kalian telah beriman kepadanya,"* hai orang-orang mukmin, yakni mereka beriman kepada semua kitab dan rasul Allah, serta tidak membedakan seorang pun di antara mereka, *"sungguh mereka telah mendapat petunjuk,"* yakni mereka telah menempuh jalan yang *haq* dan mendapat bimbingan ke arah-Nya (Katsir, 2000:127).

Allah Swt. juga berfirman, *"Dan jika mereka berpaling,"* yakni dari jalan yang benar dan menempuh jalan yang *bathil*, sesudah *hujjah* mematahkan alasan mereka, *"sesungguhnya mereka berada dalam permusuhan (dengan kamu),"* *"maka Allah akan memelihara kamu dari mereka,"* yakni Allah akan menolongmu dalam menghadapi mereka dan Dia akan memberikan kemenangan pada kalian atas mereka, *"dan Dia-lah yang Maha Mendengar lagi Maha Mengetahui"* (Katsir, 2000:127).

Berdasarkan tafsir ayat di atas, dapat diketahui bahwa ayat tersebut mempunyai relevansi dengan distribusi Bernoulli dalam matematika. Hal ini ditunjukkan pada kalimat

فَإِنْ ءَامَنُواْ بِمِثْلِ مَآ ءَامَنتُم بِهِۦ فَقَدِ ٱهْتَدَواْ

yang artinya, *"maka jika mereka beriman kepada apa yang kalian telah beriman kepadanya, sungguh mereka telah mendapat petunjuk,"* dan kalimat

وَّإِن تَوَلَّوۡاْ فَإِنَّمَا هُمۡ فِي شِقَاقٖ

yang artinya, *"dan jika mereka berpaling, sesungguhnya mereka berada dalam permusuhan (dengan kamu)."*

Pada kalimat pertama, terdapat kata *ihtadaw* yang bermakna mereka telah mendapat petunjuk, atau jalan keluar. Kata tersebut menunjukkan bahwa mereka adalah orang-orang yang berhasil, atau dalam distribusi Bernoulli dapat dikatakan bahwa mereka telah sukses dan bernilai 1. Selanjutnya pada kalimat kedua, terdapat kata *syiqaq* yang dapat diartikan dengan susah, sempit, atau suasana terhimpit karena telah berpaling dari jalan yang *haq*. Sehingga dapat dikatakan bahwa mereka telah gagal, dan dalam distribusi Bernoulli bernilai 0.

Permasalahan yang sering terjadi pada regresi probit saat variabel prediktor lebih dari satu adalah terjadi korelasi antar variabel-variabel prediktor tersebut yang disebut sebagai multikolinieritas. Jika terdapat multikolinieritas, berarti salah satu asumsi klasik tidak terpenuhi. Hal ini berarti bahwa penduga/estimator yang dihasilkan menjadi tidak efisien sehingga variansi dari koefisien regresi menjadi tidak minimum (Gujarati, 2006a:48).

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, yang di dalamnya telah diteliti bagaimana cara mendapatkan estimasi parameter model regresi probit dengan menggunakan metode *maximum likelihood*, beserta aplikasinya dalam data.

Berdasarkan latar belakang tersebut, akan diteliti estimasi parameter model regresi probit ridge yang kemudian dapat digunakan untuk mengestimasi model regresi probit yang mengandung multikolinieritas. Dalam penelitian ini, peneliti mengambil judul "Estimasi Parameter Model Regresi Probit Ridge dengan Metode *Maximum Likelihood*".

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana bentuk estimasi parameter model regresi probit ridge dengan menggunakan metode *maximum likelihood*?
2. Bagaimana hasil estimasi parameter model regresi probit ridge pada data curah hujan di Karangploso Malang?

1.3 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengetahui bentuk estimasi parameter model regresi probit ridge dengan menggunakan metode *maximum likelihood*.
2. Mengetahui hasil estimasi parameter model regresi probit ridge pada data curah hujan di Karangploso Malang.

1.4 Batasan Masalah

Sesuai rumusan masalah di atas, pembatasan masalah dalam penelitian ini yaitu, estimasi parameter khusus untuk model regresi probit yang mengandung multikolinieritas dan aplikasi ke dalam data hanya menggunakan dua variabel prediktor saja dengan menggunakan parameter ridge $k = \max\left(\frac{1}{q_j}\right)$ dengan

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

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah:

a. Bagi peneliti
   1. Memperdalam dan mengembangkan disiplin ilmu matematika, khususnya pada bidang statistika.
   2. Mengetahui bagaimana cara mengestimasi parameter dari model regresi probit yang mengandung multikolinieritas.

b. Bagi lembaga
   1. Sebagai sumbangan pemikiran keilmuan matematika, khususnya dalam bidang statistika, analisis regresi dan ekonometrika.
   2. Meningkatkan peran serta Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang dalam pengembangan wawasan keilmuan matematika.

c. Bagi Pembaca

Memberikan pengetahuan lebih mendalam tentang estimasi parameter dan menjadikan penelitian ini sebagai bahan rujukan dalam pengembangan pembelajaran statistika, analisis regresi, dan ekonometrika.

1.6 Sistematika Penulisan

Penulisan penelitian ini terdiri dari 5 bab yang masing-masing terbagi dalam subbab dengan sistematika penulisan sebagai berikut:

Bab I Pendahuluan

Bab ini meliputi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan sistematika penulisan.

Bab II Kajian Pustaka

Bab ini menjelaskan tentang teori-teori pendukung yang digunakan dalam penelitian, yang meliputi distribusi normal dan normal standar, ekspektasi dan variansi, estimasi parameter, metode *maximum likelihood*, iterasi *method of scoring*, regresi probit, pemusatan dan penskalaan data, multikolinieritas, bentuk kanonik model regresi, regresi ridge, *generalized linear model*, dan analisis regresi dalam al-Quran.

Bab III Metode Penelitian

Bab ini berisi tentang data dan metode yang digunakan peneliti yang meliputi sumber data, variabel penelitian, dan metode analisis.

Bab IV Pembahasan

Pada bab ini dijelaskan bagaimana mendapatkan estimator regresi probit dengan menggunakan metode *maximum likelihood*, dilanjutkan dengan bagaimana mendapatkan estimator regresi probit ridge. Estimator yang telah didapatkan, diaplikasikan ke dalam data curah hujan di Karangploso Malang setelah menguji adanya multikolinieritas antar variabel prediktor, hingga memperoleh model regresi probit ridge dari data.

Bab V Penutup

Bab ini berisi kesimpulan dari penelitian dan saran untuk penelitian selanjutnya.

BAB II

KAJIAN PUSTAKA

2.1 Distribusi Normal dan Normal Standar

Menurut Dudewicz dan Mishra (1995:153), suatu peubah acak X dikatakan berdistribusi normal, atau $X \sim N\left(\mu_X, \sigma_X{}^2\right)$ yang berarti berdistribusi normal dengan nilai rata-rata μ_X dan variansi σ_X^2, jika (untuk suatu $\sigma^2 > 0$ dan $-\infty < \mu < \infty$) berlaku

$$f_X\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty \tag{2.1}$$

Fungsi di atas menunjukkan nilai *Probability Density Function* (PDF) dari distribusi normal. Sehingga *Cumulative Distribution Function* (CDF) dari distribusi normal dapat dinyatakan sebagai berikut:

$$F_X\left(x\right) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \tag{2.2}$$

Persamaan di atas dapat ditransformasikan ke dalam distribusi normal standar yang dinyatakan dengan $Z = \dfrac{X-\mu}{\sigma} \sim N\left(0,1\right)$, dimana

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

Misalkan $w = \dfrac{x-\mu}{\sigma}$, maka $dw = \dfrac{1}{\sigma} dx$, sehingga

$$\begin{aligned} x &= \mu + z\sigma \\ w &= \frac{\mu + 2\sigma - \mu}{\sigma} \\ &= z \end{aligned}$$

maka diperoleh

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

Simbol $\Phi(z)$ dinotasikan untuk CDF yang berdistribusi normal standar. Dan turunan dari standar normal kumulatif disebut standar normal PDF yang dinyatakan sebagai berikut:

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

2.2 Ekspektasi dan Variansi

Definisi 2.1 Ekspektasi

Ekspektasi atau rataan dari suatu variabel acak X didefinisikan sebagai $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$, jika X kontinu mutlak dengan fungsi padat peluang $f_X(x)$, dan $E[X] = \sum x_i p_X(x_i)$, jika X diskrit dengan fungsi masa peluang $p_X(x)$ (Dudewicz dan Mishra, 1995:246).

Definisi 2.2 Variansi

Misalkan *X* suatu variabel acak dengan fungsi distribusi $F(x)$. Momen pusat ke-*n* dari *X* (bila nilai ekspektasi ini ada) adalah $\mu_n = E\left[X - E[X]\right]^n$. Variansi dari *X*, dinyatakan dengan $Var(X)$ atau $\sigma_X^2(X)$, μ_2 (momen pusat kedua dari *X*). Sehingga $Var(X) = E\left[X - E[X]\right]^2$ (Dudewicz dan Mishra, 1995:253).

2.3 Estimasi Parameter

Dalam statistika, estimasi (penaksiran) adalah suatu metode untuk mengetahui nilai-nilai suatu populasi dengan menggunakan nilai-nilai sampel. Dalam kasus sebuah variabel acak X diasumsikan berdistribusi normal dengan dua parameter, yaitu nilai rata-rata (μ_X) dan varians $(\sigma_X{}^2)$, dimana nilai dari kedua parameter ini tidak diketahui. Untuk menaksir nilai parameter yang tidak diketahui ini, dapat diasumsikan terdapat sampel acak sebesar n dari distribusi probabilitas yang diketahui dan menggunakan sampel tersebut untuk menaksir parameter yang tidak diketahui. Jadi, rata-rata sampel dapat disajikan sebagai taksiran atas rata-rata populasi dan varians sampel sebagai taksiran atas varians populasi (Gujarati, 2006a:12).

Prinsip penggunaan metode estimasi pada sebuah observasi *t*, dengan persamaan regresi $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$, dapat diperoleh nilai *error* sebagai berikut:

$$\varepsilon_t = Y_t - \beta_0 - \beta_1 X_t$$

Dari suatu sampel sebanyak n, akan diperoleh suatu *error* ke-*n*. Sehingga dapat diperoleh rata-rata *error* ke-*n* dari sampel sebagai berikut:

$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \beta_0 - \beta_1 X_t\right)$$

Diasumsikan nilai rata-rata *error* ke-*n* adalah nol, sehingga berakibat nilai parameter $\beta_1 = 0$. Karena pada model ini hanya memiliki satu parameter, yaitu β_0, maka diperoleh

$$\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{\beta}_0\right) = 0 \qquad (2.3)$$

Karena nilai β_0 tidak terikat indeks *t*, persamaan (2.3) dapat ditulis dengan

$$\frac{1}{n}\sum_{t=1}^{n}Y_t - \hat{\beta}_0 = 0$$

$$\hat{\beta}_0 = \frac{1}{n}\sum_{t=1}^{n}Y_t$$

dimana $\hat{\beta}_0$ adalah estimasi dari β_0.

2.4 Metode *Maximum Likelihood*

Suatu metode yang bersifat umum dari estimasi titik dengan beberapa sifat teoretis yang lebih kuat dibandingkan dengan metode penaksiran kuadrat terkecil (*least square estimation*) adalah metode kemungkinan terbesar (*maximum likelihood*) (Aziz, 2010:29).

Misalkan *y* variabel acak berdistribusi Bernoulli dengan parameter θ berukuran *n*. Metode *maximum likelihood* akan memilih nilai θ yang diketahui sedemikian hingga memaksimumkan nilai probabilitas (*likelihood*) dari gambaran sampel secara acak yang telah diperoleh secara aktual. Untuk $y=0$ atau $y=1$, dapat dihitung probabilitas sampel acak dari *joint* PDF untuk $y_1, y_2, ..., y_n$, yaitu

$$f(y_1=1,...,y_n=0)=f(1,...,0)=\prod_{i=1}^{n}\theta^{y_i}(1-\theta)^{1-y_i}$$

Jadi, fungsi *likelihood*-nya adalah

$$l(\theta\,|\,y)=f(y\,|\,\theta)=\theta^{\sum_{i=1}^{n}y_i}(1-\theta)^{\sum_{i=1}^{n}(1-y_i)}$$

sedangkan fungsi *ln-likelihood*-nya adalah

$$L(\theta\,|\,y)=\ln f(y\,|\,\theta)=\sum_{i=1}^{n}y_i\ln\theta+\sum_{i=1}^{n}(1-y_i)\ln(1-\theta)$$

Untuk memaksimumkan fungsi *likelihood*, diperlukan

$$\frac{dL}{d\theta}=\sum_{i=1}^{n}y_i\frac{1}{\theta}-\sum_{i=1}^{n}(1-y_i)\frac{1}{1-\theta}$$

$$\frac{d^2L}{d\theta^2}=-\sum_{i=1}^{n}y_i\frac{1}{\theta^2}-\sum_{i=1}^{n}(1-y_i)\frac{1}{(1-\theta)^2}$$

Menyamakan turunan pertama dengan nol dan menyelesaikannya menghasilkan $\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}y_i$ yang merupakan nilai rata-rata sampel. Sedangkan turunan kedua selalu bernilai negatif untuk $0<\theta<1$, sehingga θ merupakan nilai maksimum global untuk fungsi *ln-likelihood* (Aziz, 2010:30).

2.5 Iterasi *Method of Scoring*

Iterasi *method of scoring* adalah salah satu iterasi dari metode nonlinier *maximum likelihood* untuk mendapatkan estimasi parameter β dengan

$$\beta^{(n+1)}=\beta^{(n)}-\left(E\left(\frac{\partial^2L}{\partial\beta\partial\beta'}\big|_{\beta^{(n)}}\right)\right)^{-1}\frac{\partial L}{\partial\beta}\big|_{\beta^{(n)}}$$

Pandang fungsi padat peluang (PDF) dari y_i yang diberikan oleh X_i, β dan σ^2 berikut

$$f\left(y_i \mid X_i, \beta, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\left(y_i - f\left(X_i, \beta\right)\right)^2\right)$$
$$= l_i\left(\beta, \sigma^2\right)$$

Dengan menggunakan sifat dari PDF, yaitu

$$\int_{-\infty}^{+\infty} f\left(y_i \mid X_i, \beta, \sigma^2\right) dy_i = 1$$

maka

$$\frac{\partial}{\partial \beta} \int_{-\infty}^{+\infty} f\left(y_i \mid X_i, \beta, \sigma^2\right) dy_i = 0$$

atau

$$\int_{-\infty}^{+\infty} \frac{\partial f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta} dy_i = 0$$
$$\int_{-\infty}^{+\infty} \frac{\partial f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta} \frac{f\left(y_i \mid X_i, \beta, \sigma^2\right)}{f\left(y_i \mid X_i, \beta, \sigma^2\right)} dy_i = 0$$

Perhatikan

$$l_i\left(\beta, \sigma^2\right) = f\left(y_i \mid X_i, \beta, \sigma^2\right)$$

dan

$$L_i = \ln l_i\left(\beta, \sigma^2\right) = \ln f\left(y_i \mid X_i, \beta, \sigma^2\right)$$

maka

$$\frac{\partial L_i}{\partial \beta} = \frac{1}{l_i\left(\beta, \sigma^2\right)} \frac{\partial l_i\left(\beta, \sigma^2\right)}{\partial \beta}$$
$$= \frac{1}{f\left(y_i \mid X_i, \beta, \sigma^2\right)} \frac{\partial f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta}$$

atau

$$\frac{\partial L_i}{\partial \beta} = \frac{\partial \log l_i}{\partial \beta}$$

$$= \frac{\partial \log f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta}$$

Dari

$$\int_{-\infty}^{+\infty} \frac{\partial f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta} \frac{f\left(y_i \mid X_i, \beta, \sigma^2\right)}{f\left(y_i \mid X_i, \beta, \sigma^2\right)} dy_i = 0$$

maka

$$\int_{-\infty}^{+\infty} \frac{\partial L_i}{\partial \beta} f\left(y_i \mid X_i, \beta, \sigma^2\right) dy_i = 0$$

Dengan melakukan turunan parsial pertama terhadap β' dan menyamakannya dengan nol, yaitu

$$\frac{\partial}{\partial \beta'}\left(\int_{-\infty}^{+\infty} \frac{\partial L_i}{\partial \beta} f\left(y_i \mid X_i, \beta, \sigma^2\right) dy_i\right) = 0$$

akan diperoleh

$$0 = \int_{-\infty}^{+\infty}\left(\frac{\partial^2 L_i}{\partial \beta \partial \beta'} f\left(y_i \mid X_i, \beta, \sigma^2\right) + \frac{\partial L_i}{\partial \beta} \frac{\partial f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta'}\right) dy_i$$

$$= \int_{-\infty}^{+\infty}\left(\frac{\partial^2 L_i}{\partial \beta \partial \beta'} f\left(y_i \mid X_i, \beta, \sigma^2\right) + \frac{\partial L_i}{\partial \beta} \frac{\partial f\left(y_i \mid X_i, \beta, \sigma^2\right)}{\partial \beta'} \frac{f\left(y_i \mid X_i, \beta, \sigma^2\right)}{f\left(y_i \mid X_i, \beta, \sigma^2\right)}\right) dy_i$$

$$= \int_{-\infty}^{+\infty}\left(\frac{\partial^2 L_i}{\partial \beta \partial \beta'} f\left(y_i \mid X_i, \beta, \sigma^2\right) + \frac{\partial L_i}{\partial \beta} \frac{\partial L_i}{\partial \beta'} f\left(y_i \mid X_i, \beta, \sigma^2\right)\right) dy_i$$

$$= \int_{-\infty}^{+\infty}\left(\left(\frac{\partial^2 L_i}{\partial \beta \partial \beta'} + \frac{\partial L_i}{\partial \beta} \frac{\partial L_i}{\partial \beta'}\right) f\left(y_i \mid X_i, \beta, \sigma^2\right)\right) dy_i$$

$$= E\left(\frac{\partial^2 L_i}{\partial \beta \partial \beta'} + \frac{\partial L_i}{\partial \beta} \frac{\partial L_i}{\partial \beta'}\right)$$

atau

$$E\left(\frac{\partial^2 L_i}{\partial\beta\partial\beta'}\right) = -E\left(\frac{\partial L_i}{\partial\beta}\frac{\partial L_i}{\partial\beta'}\right)$$

2.6 Regresi Probit

Analisis regresi merupakan suatu teknik yang digunakan untuk membangun suatu persamaan yang menghubungkan antara variabel respon (Y) dengan variabel prediktor (X) dan sekaligus untuk menentukan nilai ramalan atau dugaannya. Sedangkan persamaan regresi adalah suatu persamaan matematika yang mendefinisikan hubungan antara dua variabel (Suharyadi dan Purwanto, 2009:74).

Ketika satu atau lebih variabel respon dalam model regresi bersifat kualitatif, maka dapat digunakan metode regresi linier dengan teknik variabel *dummy* untuk mengestimasi model ini. Namun, mengestimasi model dengan variabel respon bersifat kualitatif sangat berbeda. Disinilah peran regresi probit. Menurut Candra (2009:3), 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.

Menurut Skrondal & Hesketh dalam Widhiarso (2012:1-2), regresi probit merupakan modifikasi regresi logistik dengan menetapkan persamaan regresi logit berdistribusi normal. Dengan menggunakan regresi probit, maka $\beta_0 + \beta_i X$ dilihat sebagai skor standar *Z* yang mengikuti distribusi normal, sehingga didapatkan

$$P = \frac{\exp(Z)}{1+\exp(Z)} \text{ atau } \ln\left(\frac{p}{1-p}\right) = Z$$

Persamaan ini didasari pada distribusi normal (Φ) di bawah ini sehingga regresi probit ditunjukkan dengan $\Phi(Z)$. Simbol Φ menunjukkan berlakunya fungsi standar deviasi distribusi normal (*inverse standard normal distribution*).

$$P(Y=1) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} dZ = \Phi(Z)$$

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. Contoh lainnya adalah data pelanggan, yaitu melakukan pembelian ulang (kode 1) dan tidak melakukan pembelian ulang (kode 0). Dalam kasus ini *Z* merupakan suatu kecenderungan pelanggan untuk melakukan pembelian ulang. Semakin besar nilai *Z* semakin besar kecenderungan pelanggan untuk melakukan pembelian ulang.

Menurut Locking, dkk (2011:1), model regresi probit ditunjukkan pada persamaan

$$y_i^* = x_i'\beta + \varepsilon_i$$

dimana y_i^* merupakan variabel laten, x_i adalah baris ke-*i* dari *X* yang merupakan matriks berordo $n \times (p+1)$ dengan p merupakan banyaknya variabel prediktor, β 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 variabel *dummy* sebagai berikut:

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

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

2.7 Pemusatan dan Penskalaan Data dalam Bentuk Matriks Korelasi

Pemusatan dan penskalaan data merupakan bagian dari membakukan variabel. Modifikasi sederhana dari membakukan variabel ini adalah transformasi korelasi. Pemusatan merupakan perbedaan antara masing-masing pengamatan dan rata-rata dari semua pengamatan untuk variabel. Penskalaan merupakan gambaran pengamatan pada unit standar deviasi dari pengamatan untuk variabel (Kutner, dkk, 2005:272).

Misalkan yang akan dibakukan adalah model regresi linier dengan 2 variabel prediktor sebagai berikut:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

Model di atas jika dalam bentuk matriks dengan ukuran matriks Y adalah $n\times 1$, ukuran matriks X adalah $n\times 2$ dan ukuran matriks β adalah 2×1 menjadi

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \qquad X = \begin{bmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \\ \vdots & \vdots \\ X_{1n} & X_{2n} \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Bentuk matriks $X'X$ dari matriks di atas adalah

$$X'X = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2n} \end{bmatrix} \begin{bmatrix} X_{11} & X_{21} \\ X_{12} & X_{22} \\ \vdots & \vdots \\ X_{1n} & X_{2n} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} X_{1i}^{2} & \sum_{i=1}^{n} X_{1i} X_{2i} \\ \sum_{i=1}^{n} X_{2i} X_{1i} & \sum_{i=1}^{n} X_{2i}^{2} \end{bmatrix}$$

Bentuk umum dari $X'X$ adalah sebagai berikut:

$$X'X = \begin{bmatrix} \sum_{i=1}^{n} X_{1i}^{2} & \sum_{i=1}^{n} X_{1i} X_{2i} & \cdots & \sum_{i=1}^{n} X_{1i} X_{ki} \\ \sum_{i=1}^{n} X_{2i} X_{1i} & \sum_{i=1}^{n} X_{2i}^{2} & \cdots & \sum_{i=1}^{n} X_{2i} X_{ki} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{n} X_{ki} X_{1i} & \sum_{i=1}^{n} X_{ki} X_{2i} & \cdots & \sum_{i=1}^{n} X_{ki}^{2} \end{bmatrix}$$

Misal U adalah matriks X yang sudah dipusatkan, maka

$$U = \begin{bmatrix} X_{11} - \bar{X}_1 & X_{21} - \bar{X}_2 \\ X_{12} - \bar{X}_1 & X_{22} - \bar{X}_2 \\ \vdots & \vdots \\ X_{1n} - \bar{X}_1 & X_{2n} - \bar{X}_2 \end{bmatrix}, \text{ dengan } \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_{ji} \text{ dimana } j = 1, 2$$

Bentuk $U'U$ dari matriks U adalah

$$U'U = \begin{bmatrix} X_{11} - \bar{X}_1 & X_{12} - \bar{X}_1 & \cdots & X_{1n} - \bar{X}_1 \\ X_{21} - \bar{X}_2 & X_{22} - \bar{X}_2 & \cdots & X_{2n} - \bar{X}_2 \end{bmatrix} \begin{bmatrix} X_{11} - \bar{X}_1 & X_{21} - \bar{X}_2 \\ X_{12} - \bar{X}_1 & X_{22} - \bar{X}_2 \\ \vdots & \vdots \\ X_{1n} - \bar{X}_1 & X_{2n} - \bar{X}_2 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right)^2 & \sum_{i=1}^{n} \left(X_{1i} - \bar{X}_1 \right) \left(X_{2i} - \bar{X}_2 \right) \\ \sum_{i=1}^{n} \left(X_{2i} - \bar{X}_2 \right) \left(X_{1i} - \bar{X}_1 \right) & \sum_{i=1}^{n} \left(X_{2i} - \bar{X}_2 \right)^2 \end{bmatrix}$$

Sedangkan penskalaan merupakan gambaran pengamatan pada unit standar deviasi dari pengamatan untuk variabel. Bentuk umum standar deviasi adalah akar dari variansi

$$S_{jj} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ji} - \bar{X}_j\right)^2$$

Transformasi korelasi adalah fungsi sederhana dari membakukan variabel. Dengan mengikuti bentuk matriks X di atas maka bentuk Z sebagai matriks yang sudah dibakukan adalah

$$Z = \begin{bmatrix} \frac{X_{11}-\bar{X}_1}{\sqrt{S_{11}}} & \frac{X_{21}-\bar{X}_2}{\sqrt{S_{22}}} \\ \frac{X_{12}-\bar{X}_1}{\sqrt{S_{11}}} & \frac{X_{22}-\bar{X}_2}{\sqrt{S_{22}}} \\ \vdots & \vdots \\ \frac{X_{1n}-\bar{X}_1}{\sqrt{S_{11}}} & \frac{X_{2n}-\bar{X}_2}{\sqrt{S_{22}}} \end{bmatrix}$$

$$Z'Z = \begin{bmatrix} \left(\frac{\sum_{i=1}^{n}\left(X_{1i}-\bar{X}_1\right)}{\sqrt{S_{11}}}\right)^2 & \left(\frac{\sum_{i=1}^{n}\left(X_{1i}-\bar{X}_1\right)\left(X_{2i}-\bar{X}_2\right)}{\sqrt{S_{11}}\sqrt{S_{22}}}\right) \\ \left(\frac{\sum_{i=1}^{n}\left(X_{2i}-\bar{X}_2\right)\left(X_{1i}-\bar{X}_1\right)}{\sqrt{S_{22}}\sqrt{S_{11}}}\right) & \left(\frac{\sum_{i=1}^{n}\left(X_{2i}-\bar{X}_2\right)}{\sqrt{S_{22}}}\right)^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \frac{S_{12}}{\sqrt{S_{11}}\sqrt{S_{22}}} \\ \frac{S_{21}}{\sqrt{S_{22}}\sqrt{S_{11}}} & 1 \end{bmatrix}$$

(Nisa’, 2014:21-24).

2.8 Multikolinieritas

Istilah multikolinieritas diciptakan oleh Ragner Frish di dalam bukunya yang berjudul: *Statistical Confluence Analysis by Means of Complete Regression Systems*. Istilah itu berarti 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, 2005a:19).

Multikolinieritas antara variabel prediktor X akan mengakibatkan determinan matriks $X'X$ pada estimator *ordinary least square* maupun *maximum likelihood* mendekati nol sehingga menjadi singular. Draper dan Smith (1992:142) menyatakan bahwa hal ini dapat diketahui dari matriks korelasi hasil pemusatan dan penskalaan matriks *X* sebagai berikut:

$$\hat{\beta}_{LS} = \left(X'X\right)^{-1} X'Y$$

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

dimana r_{12} adalah koefisien korelasi antara X_1 dan X_2. Nilai r_{12} yang membesar akan menyebabkan determinan matriks $X'X$ mendekati nol (multikolinieritas mendekati sempurna) atau sama dengan nol (multikolinieritas sempurna) sehingga mengakibatkan matriks menjadi singular (tidak memiliki invers).

Menurut Setiawan dan Kusrini (2010:26), salah satu ukuran untuk menguji adanya multikolinieritas adalah *Variance Inflation Factors* (VIF). VIF merupakan elemen diagonal dari matriks $X'X$.

$$\left(X'X\right)^{-1} = \begin{bmatrix} \dfrac{1}{1-r_{12}^2} & \dfrac{-r_{12}}{1-r_{12}^2} \\ \dfrac{-r_{12}}{1-r_{12}^2} & \dfrac{1}{1-r_{12}^2} \end{bmatrix}$$

$$VIF = diag\left(X'X\right) = \frac{1}{1-R^2}$$

Menurut Nisa' (2014:30), pengujian multikolinieritas juga dapat dilakukan dengan menghitung nilai VIF dengan persamaan

$$VIF = \left(\frac{1}{n-1} (X'X) \right)^{-1}$$

Sedangkan nilai VIF dari estimator *generalized ridge regression* dapat dihitung melalui persamaan

$$VIF = \left(\frac{1}{n-1} (X'X) + DKD' \right)^{-1} \left(\frac{1}{n-1} (X'X) \right) \left(\frac{1}{n-1} (X'X) + DKD' \right)^{-1}$$

dengan K adalah matriks yang elemen diagonalnya merupakan parameter ridge $k \geq 0$.

2.9 Bentuk Kanonik Model Regresi

Misalkan terdapat suatu matriks ortogonal D dimana $D' = D^{-1}$ sedemikian sehingga $D'D = I$ dan $D'CD = \Lambda$, dengan $C = X'X$ dan Λ merupakan matriks $p \times p$ dimana anggota diagonal utamanya merupakan nilai eigen dari matriks $X'X$. Proses bentuk kanonik dari model regresi $Y = X\beta + \varepsilon$ adalah

$$\begin{aligned} Y &= X\beta + \varepsilon \\ &= XDD'\beta + \varepsilon \\ &= X^{*}\gamma + \varepsilon \end{aligned} \tag{2.4}$$

dengan $X^{*} = XD$, $\gamma = D'\beta$, dan $\beta = D\gamma$, maka estimasi parameter γ dari persamaan (2.4) adalah

$$\hat{\beta}_{ML} = \left(X'X \right)^{-1} X'Y$$

$$D\hat{\gamma}_{ML} = \left(X'X \right)^{-1} X'Y$$

$$\left(X'X\right)D\hat{\gamma}_{ML} = \left(X'X\right)\left(X'X\right)^{-1}X'Y$$

$$X'XD\hat{\gamma}_{ML} = IX'Y$$

$$D'X'XD\hat{\gamma}_{ML} = D'X'Y$$

$$\left(D'X'XD\right)^{-1}\left(D'X'XD\right)\hat{\gamma}_{ML} = \left(D'X'XD\right)^{-1}D'X'Y$$

$$I\hat{\gamma}_{ML} = \left(D'X'XD\right)^{-1}D'X'Y$$

$$\begin{aligned}\hat{\gamma}_{ML} &= \left(D'CD\right)^{-1}\left(XD\right)'Y \\ &= \left(D'CD\right)^{-1}X^{*\prime}Y \\ &= \Lambda^{-1}X^{*\prime}Y\end{aligned}$$

(Nisa', 2014:27-28).

2.10 Regresi Ridge

Menurut Hoerl dan Kennard (1970:3), estimasi ridge untuk koefisien regresi dapat diperoleh dengan menyelesaikan suatu bentuk dari persamaan normal regresi. Asumsikan bahwa bentuk standar dari model regresi linier ganda adalah sebagai berikut:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + \varepsilon$$

Parameter penting yang membedakan regresi ridge dari metode kuadrat terkecil adalah *k*. Parameter ridge *k* yang relatif kecil ditambahkan pada diagonal utama matriks $X'X$, sehingga koefisien estimator regresi ridge dipenuhi dengan besarnya parameter ridge *k*.

Dalam praktiknya, perhitungan estimator regresi ridge dengan menyelesaikan persamaan di atas sangatlah rumit, oleh sebab itu dilakukan penyederhanaan dengan membawanya ke dalam bentuk notasi matriks. Estimator ridge diperoleh dengan meminimumkan jumlah kuadrat *error* untuk model

$$Y = X\beta + \varepsilon$$

atau

$$\varepsilon = Y - X\beta$$

dengan menggunakan metode pengali Lagrange yang meminimumkan fungsi

$$\varepsilon'\varepsilon = (Y - X\beta_R)'(Y - X\beta_R)$$

dengan syarat pambatas

$$\beta_R{}'\beta_R - c^2 = 0$$
$$G = (Y - X\beta_R)'(Y - X\beta_R) + k\left(\beta_R{}'\beta_R - c^2\right)$$

yang memenuhi syarat

$$\left.\frac{\partial G}{\partial \beta_R}\right|_{\hat{\beta}_R} = 0$$

dimana k pengali Lagrange tidak bergantung pada β_R dan k konstanta positif berhingga.

Dicari $\hat{\beta}_R$ dengan memecah $\left.\frac{\partial G}{\partial \beta_R}\right|_{\hat{\beta}_R} = 0$

$$\begin{aligned} G &= (Y - X\beta_R)'(Y - X\beta_R) + k\left(\beta_R{}'\beta_R - c^2\right) \\ &= Y'Y - Y'X\beta_R - \beta_R{}'X'Y + \beta_R{}'X'X\beta_R + k\left(\beta_R{}'\beta_R - c^2\right) \\ &= Y'Y + 2\beta_R{}'X'Y + 2\beta_R{}'X'X\beta_R + k\left(\beta_R{}'\beta_R - c^2\right) \end{aligned}$$

$$\left.\frac{\partial G}{\partial \beta_R}\right|_{\hat{\beta}_R} = 0$$

$$\begin{aligned} -2X'Y + 2X'X\hat{\beta}_R + 2kI\hat{\beta}_R &= 0 \\ -X'Y + X'X\hat{\beta}_R + kI\hat{\beta}_R &= 0 \\ X'X\hat{\beta}_R + kI\hat{\beta}_R &= X'Y \end{aligned}$$

$$\left(X'X + kI\right)\hat{\beta}_R = X'Y$$
$$\hat{\beta}_R = \left(X'X + kI\right)^{-1} X'Y$$

dimana $\hat{\beta}_R = \left(X'X + kI\right)^{-1} X'Y$ dengan $0 \le k \le \infty$, itulah yang disebut sebagai estimator regresi ridge. $k \ge 0$ adalah nilai konstan yang dipilih sebagai indeks dari kelas estimator.

2.11 *Generalized Linear Model* (GLM)

Konsep penting yang mendasari GLM adalah *exponential family*. Semua anggota dari *exponential family* memiliki fungsi padat peluang untuk variabel respon *y* yang dapat dinyatakan ke dalam bentuk

$$f\left(y;\theta,\phi\right) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c\left(y,\phi\right)\right\} \tag{2.5}$$

dimana $a(.)$, $b(.)$ dan $c(.)$ adalah fungsi tertentu. Parameter θ adalah parameter natural, dan parameter ϕ sering disebut dengan parameter dispersi. Untuk beberapa anggota dari *exponential family* berlaku $\phi = 1$, seperti Binomial dan Poisson (Myers, dkk, 2010:203).

Untuk anggota dari *exponential family*,

$$E\left(\frac{\partial \ln L(\theta)}{\partial \theta}\right) = 0$$

dan

$$E\left(\frac{\partial^2 \ln L}{\partial \theta^2}\right) + E\left(\frac{\partial \ln L}{\partial \theta}\right)^2 = 0$$

yang kemudian didapatkan

$$\mu = E(y) = \frac{db(\theta)}{d\theta} = b'(\theta)$$

$$Var(y) = \frac{d^2 b(\theta)}{d\theta^2} a(\phi) = b''(\theta) a(\phi)$$

$$= \frac{d\mu}{d\theta} a(\phi)$$

(Myers, dkk, 2010:204).

Regresor dijelaskan dengan cara berikut. Definisikan *linear predictor*

$$\eta = x'\beta$$

dengan $\eta = g(\mu)$ adalah *link function* yang menghubungkan *linear predictor* dengan *mean* μ. Keadaan khusus dari *link function* yang sangat menarik adalah *canonical link function*, ketika

$$\eta = \theta$$

Canonical link function ini adalah cara paling mudah yang memudahkan fungsi padat peluang (2.5) dievaluasi pada $\theta = x'\beta$ (Cameron dan Trivedi, 1998:34).

Metode *maximum likelihood* untuk GLM, memaksimumkan

$$L_{GLM} = \sum_{i=1}^{n} \left\{ \frac{\theta\left(x_i'\beta\right) y_i - b\left(\theta\left(x_i'\beta\right)\right)}{a(\phi)} + c(y_i, \phi) \right\}$$

(Cameron dan Trivedi, 1998:34).

2.12 Konsep Analisis Regresi dalam Al-Quran

Al-Quran adalah wahyu Allah Swt. yang diturunkan kepada Nabi Muhammad Saw. sebagai pedoman hidup manusia untuk berhijrah dari jalan yang gelap gulita menuju jalan yang terang benderang yaitu jalan yang diridlai Allah Swt.. Karena pada hakikatnya, kehidupan dunia adalah kehidupan yang sementara, juga sebagai jalan untuk menuju kehidupan di akhirat. Selama di dunia, manusia diperintahkan untuk senantiasa beribadah kepada Allah Swt., menaati segala perintah-Nya dan menjauhi larangan-Nya. Amal perbuatan inilah yang kelak akan menentukan dimana tempat manusia di akhirat kelak, di surga atau di neraka. Allah berfirman dalam al-Quran surat al-Baqarah/2:82 sebagai berikut:

وَٱلَّذِينَ ءَامَنُواْ وَعَمِلُواْ ٱلصَّٰلِحَٰتِ أُوْلَٰٓئِكَ أَصۡحَٰبُ ٱلۡجَنَّةِۖ هُمۡ فِيهَا خَٰلِدُونَ ٨٢

"Dan orang-orang yang beriman serta beramal shalih, mereka itu penghuni surga, mereka kekal di dalamnya" (QS. al-Baqarah/2:82).

Ayat tersebut menyebutkan bahwa penghuni (ahli) surga adalah orang-orang yang beriman serta beramal shalih. Yakni beriman kepada Allah Swt. dan rasul-Nya serta mengamalkan amal-amal shalih yang sesuai dengan apa yang diperintahkan oleh syariat. Ayat ini juga bersesuaian dengan surat al-Nisa'/4:124 yang berbunyi

وَمَن يَعۡمَلۡ مِنَ ٱلصَّٰلِحَٰتِ مِن ذَكَرٍ أَوۡ أُنثَىٰ وَهُوَ مُؤۡمِنٞ فَأُوْلَٰٓئِكَ يَدۡخُلُونَ ٱلۡجَنَّةَ وَلَا يُظۡلَمُونَ نَقِيرٗا

١٢٤

"Barangsiapa yang mengerjakan amal-amal shalih, baik laki-laki maupun wanita sedang ia orang yang beriman, maka mereka itu masuk ke dalam surga dan mereka tidak dianiaya walau sedikitpun" (QS. al-Nisa'/4:124).

Kedua ayat tersebut memiliki keterkaitan dengan matematika, khususnya analisis regresi. Analisis regresi adalah suatu metode yang digunakan untuk menghubungkan suatu variabel respon (Y) dengan variabel prediktor (X). Sedangkan persamaan regresi adalah suatu persamaan matematika yang mendefinisikan hubungan antara dua variabel tersebut. Variabel prediktor inilah yang mempengaruhi nilai dari variabel respon. Ayat-ayat di atas menunjukkan bahwa iman dan amal shalih menjadi penentu apakah seseorang termasuk ahli surga atau bukan. Sehingga dalam regresi, ahli surga dapat dijadikan sebagai variabel respon (Y) karena ahli surga ini dipengaruhi oleh dua hal, yakni iman dan amal shalih. Jadi, iman dapat dikatakan sebagai variabel prediktor pertama (X_1) yang mempengaruhi apakah seseorang termasuk ahli surga, dan amal shalih sebagai variabel prediktor kedua (X_2).

Selanjutnya dalam surat al-Nisa'/4:95 Allah Swt. berfirman:

لَّا يَسْتَوِى ٱلْقَٰعِدُونَ مِنَ ٱلْمُؤْمِنِينَ غَيْرُ أُو۟لِى ٱلضَّرَرِ وَٱلْمُجَٰهِدُونَ فِى سَبِيلِ ٱللَّهِ بِأَمْوَٰلِهِمْ وَأَنفُسِهِمْ ۚ فَضَّلَ ٱللَّهُ ٱلْمُجَٰهِدِينَ بِأَمْوَٰلِهِمْ وَأَنفُسِهِمْ عَلَى ٱلْقَٰعِدِينَ دَرَجَةً ۚ وَكُلًّا وَعَدَ ٱللَّهُ ٱلْحُسْنَىٰ ۚ وَفَضَّلَ ٱللَّهُ ٱلْمُجَٰهِدِينَ عَلَى ٱلْقَٰعِدِينَ أَجْرًا عَظِيمًا ﴿٩٥﴾

"Tidaklah sama antara mukmin yang duduk (yang tidak ikut berperang) yang tidak mempunyai 'udzur dengan orang-orang yang berjihad di jalan Allah dengan harta mereka dan jiwanya. Allah melebihkan orang-orang yang berjihad dengan harta dan jiwanya atas orang-orang yang duduk satu derajat. Kepada masing-masing mereka Allah menjanjikan pahala yang baik (surga) dan Allah melebihkan orang-orang yang berjihad atas orang yang duduk dengan pahala yang besar" (QS. al-Nisa'/4:95).

Ayat tersebut mengandung pengertian bahwa pahala dari setiap orang tergantung dari seberapa besar tingkat keimanan dan amal shalihnya. Hal ini menunjukkan bahwa tingkat keimanan dan amal shalih seseorang yang sebelumnya telah dikatakan sebagai variabel prediktor, dapat diukur dengan suatu parameter (β). Semakin besar nilai parameter ini, maka semakin besar pula pengaruh dari iman dan amal shalih ini terhadap ketentuan sebagai ahli surga. Sehingga dalam bentuk persamaan regresi dapat ditulis dengan

$$Y = \beta_1 X_1 + \beta_2 X_2$$

dengan

Y = Ahli surga

X_1 = Iman

X_2 = Amal shalih

β_1 = Parameter yang mengukur tingkat keimanan

β_2 = Parameter yang mengukur tingkat amal shalih

BAB III

METODE PENELITIAN

3.1 Pendekatan Penelitian

Penelitian ini dilakukan dengan menggunakan metode kepustakaan (*library research*) yaitu dengan cara mengumpulkan dan mengkaji teori-teori pendukung yang berkaitan dengan penelitian ini, yang meliputi teori tentang metode estimasi, analisis regresi, multikolinieritas, model regresi probit, dan regresi ridge untuk mendapatkan estimator dari parameter model regresi probit ridge. Untuk mengaplikasikan estimator tersebut, peneliti mengambil data curah hujan, temperatur, dan kelembaban udara di Karangploso Malang.

3.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah data curah hujan, temperatur, dan kelembaban udara, yang merupakan data bulanan selama 3 tahun, terhitung tahun 2007 hingga 2009. Data ini diambil dari skripsi Sa'adah (2011) yang berjudul *Analisis Regresi Dummy Variable Model Probit*.

3.3 Variabel Penelitian

Variabel dalam penelitian ini adalah variabel respon yang berupa curah hujan, dan variabel prediktor yang berupa temperatur dan kelembaban udara. Variabel respon menggunakan simbol 1 dan 0, dengan ketentuan 1 untuk bulan yang sering terjadi hujan, dengan kriteria curah hujan lebih dari 150 mm, dan 0 untuk bulan yang jarang terjadi hujan, dengan kriteria curah hujan kurang dari 150 mm.

3.4 Metode Analisis

Langkah-langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Mengestimasi parameter model regresi probit ridge.

   Estimasi ini dilakukan melalui beberapa tahap, yaitu:

   i. Mengestimasi parameter model regresi probit dengan menggunakan metode *maximum likelihood* setelah pendeskripsian ke dalam bentuk GLM.

   ii. Mengestimasi parameter model regresi probit ridge dengan menggunakan estimator model regresi probit yang telah didapatkan.

2. Aplikasi metode estimasi pada data curah hujan di Karangploso Malang dengan tahap sebagai berikut:

   i. Mendeteksi adanya multikolinieritas pada data dengan menghitung nilai korelasi antar variabel prediktor dan dengan menghitung nilai VIF. Perhitungan dilakukan dengan bantuan program SPSS 15 dan MATLAB.

   ii. Mengestimasi parameter model regresi probit ridge dari data dengan estimator yang telah didapatkan. Proses estimasi dilakukan dengan bantuan program MATLAB.

   iii. Memeriksa apakah multikolinieritas telah teratasi dengan menghitung kembali nilai VIF setelah penambahan parameter ridge. Perhitungan dilakukan dengan bantuan program MATLAB.

iv. Membandingkan hasil estimasi curah hujan dengan data curah hujan yang sebenarnya untuk mengetahui kesesuaian model yang didapatkan.

3. Membuat kesimpulan.

BAB IV

PEMBAHASAN

4.1 Estimasi Parameter Model Regresi Probit

Variabel respon y_i $(i=1,2,...,n)$ dalam model regresi probit berdistribusi Bernoulli dengan peluang sukses $\pi_i = \Phi(x_i'\beta)$, $i=1,2,...,n$, dengan Φ adalah fungsi distribusi normal standar, x_i adalah kolom ke-*i* dari matriks X berordo $p \times n$ dengan $p-1$ adalah banyaknya variabel prediktor, dan β adalah matriks berordo $p \times 1$. Dalam bentuk matriks, dapat dituliskan sebagai berikut:

$$Y = \underset{(n \times 1)}{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}$$

$$X = \underset{(p \times n)}{\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}}$$

dengan

$$x_i = \underset{(p \times 1)}{\begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{pi} \end{bmatrix}}$$

sehingga,

$$x_i' = \underset{(1 \times p)}{\begin{bmatrix} x_{1i} & x_{2i} & \cdots & x_{pi} \end{bmatrix}}$$

dan

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}_{(p\times 1)}$$

Karena y_i berdistribusi Bernoulli dengan peluang sukses $\pi_i = \Phi(x_i'\beta)$, maka dalam bentuk persamaan dapat dituliskan dengan

$$f(y_i) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \tag{4.1}$$

Untuk mengestimasi parameter β, peneliti menggunakan metode *maximum likelihood* setelah mendeskripsikan persamaan (4.1) ke dalam bentuk GLM terlebih dahulu.

4.1.1 Deskripsi ke dalam Bentuk *Generalized Linear Model* (GLM)

Langkah pertama yang harus dilakukan untuk mendeskripsikan fungsi distribusi dari variabel respon di atas ke dalam bentuk GLM yaitu menunjukkan bahwa fungsi distribusi tersebut merupakan anggota dari *exponential family*. Dengan menggunakan fungsi *ln*, maka persamaan (4.1) menjadi

$$\begin{aligned} \ln f(y_i) &= y_i \ln(\pi_i) + (1-y_i)\ln(1-\pi_i) \\ &= y_i \ln(\pi_i) + \ln(1-\pi_i) - y_i \ln(1-\pi_i) \end{aligned} \tag{4.2}$$

Dengan mengumpulkan y_i, maka persamaan (4.2) dapat ditulis dengan

$$\ln f(y_i) = y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) + \ln(1-\pi_i) \tag{4.3}$$

Dari persamaan (4.3) di atas, dapat diketahui bahwa persamaan tersebut sesuai dengan bentuk

$$\ln f(y_i) = \frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i,\phi)$$

dengan $\theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ dan $b(\theta_i) = -\ln(1-\pi_i)$. Karena untuk distribusi Binomial berlaku $\phi = 1$, maka $a_i(\phi) = \phi = 1$ dan $c(y_i,\phi) = 0$. Sehingga persamaan (4.3) dapat ditulis dengan

$$L_i = \ln f(y_i) = y_i\theta_i - b(\theta_i)$$

Dari $\theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$, dapat dicari π_i sebagai berikut:

$$\begin{aligned}
\theta_i &= \ln\left(\frac{\pi_i}{1-\pi_i}\right)\\
e^{\theta_i} &= \frac{\pi_i}{1-\pi_i}\\
e^{\theta_i}(1-\pi_i) &= \pi_i\\
e^{\theta_i} - e^{\theta_i}\pi_i &= \pi_i\\
e^{\theta_i} &= \pi_i + e^{\theta_i}\pi_i\\
e^{\theta_i} &= \pi_i\left(1+e^{\theta_i}\right)\\
\pi_i &= \frac{e^{\theta_i}}{1+e^{\theta_i}}
\end{aligned}$$

sehingga,

$$\begin{aligned}
b(\theta_i) &= -\ln(1-\pi_i)\\
&= -\ln\left(1-\frac{e^{\theta_i}}{1+e^{\theta_i}}\right)\\
&= -\ln\left(\frac{1}{1+e^{\theta_i}}\right)\\
&= -\left(\ln 1 - \ln\left(1+e^{\theta_i}\right)\right)\\
b(\theta_i) &= \ln\left(1+e^{\theta_i}\right) \qquad (4.4)
\end{aligned}$$

Dari persamaan (4.4) dapat dicari $E(y)$ dan $Var(y)$ sebagai berikut:

$$\begin{aligned} E(y_i) &= \frac{db(\theta_i)}{d\theta_i} \\ &= \frac{d}{d\theta_i}\ln\left(1+e^{\theta_i}\right) \\ &= \frac{e^{\theta_i}}{1+e^{\theta_i}} \\ &= \pi_i \end{aligned}$$

dan

$$\begin{aligned} Var(y_i) &= \frac{d^2 b(\theta)}{d\theta^2}a(\phi) \\ &= \frac{d}{d\theta}\left(\frac{e^{\theta_i}}{1+e^{\theta_i}}\right) \\ &= \frac{e^{\theta_i}\left(1+e^{\theta_i}\right) - e^{\theta_i}\left(e^{\theta_i}\right)}{\left(1+e^{\theta_i}\right)^2} \\ &= \frac{e^{\theta_i}}{\left(1+e^{\theta_i}\right)^2} \end{aligned}$$

yang dalam parameter π_i dapat ditulis dengan

$$Var(y_i) = \pi_i(1-\pi_i)$$

Selanjutnya adalah menentukan *link function*. Karena *link function* terbaik adalah *canonical link function*, maka dalam hal ini peneliti menggunakan fungsi $g(\mu_i) = \eta_i = \theta_i = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$ yang merupakan *logit link function*, dimana η_i adalah *linear predictor*, sehingga $\eta_i = x_i'\beta$.

4.1.2 Estimasi dengan Metode *Maximum Likelihood*

Metode *maximum likelihood* untuk GLM dilakukan dengan memaksimumkan fungsi peluang gabungan. Karena fungsi L_i adalah fungsi $\ln$, maka fungsi peluang gabungan dari fungsi L_i yaitu:

$$\begin{aligned} L &= \sum_{i=1}^{n} L_i \\ &= \sum_{i=1}^{n} \ln f\left(y_i\right) \\ &= \sum_{i=1}^{n} \left(y_i\theta_i - b\left(\theta_i\right)\right) \\ &= \sum_{i=1}^{n} y_i\theta_i - \sum_{i=1}^{n} b\left(\theta_i\right) \end{aligned}$$

dengan menggunakan aturan rantai,

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial L_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \beta_j}$$

dengan

$$\frac{\partial L_i}{\partial \theta_i} = y_i - b'\left(\theta_i\right) = y_i - \pi_i$$

$$\begin{aligned} \frac{\partial \theta_i}{\partial \pi_i} &= \frac{\partial}{\partial \pi_i} \ln\left(\frac{\pi_i}{1-\pi_i}\right) \\ &= \frac{1}{\pi_i} + \frac{1}{1-\pi_i} \\ &= \frac{1}{\pi_i\left(1-\pi_i\right)} \end{aligned}$$

$$\frac{\partial \pi_i}{\partial \beta_j} = \frac{\partial \pi_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \pi_i}{\partial \eta_i} \cdot x_{ij}$$

sehingga didapatkan

$$\frac{\partial L}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \pi_i}{\pi_i\left(1-\pi_i\right)} x_{ji} \left(\frac{\partial \pi_i}{\partial \eta_i}\right) \tag{4.5}$$

Setelah didapatkan persamaan (4.5), digunakan iterasi *method of scoring*,

$$\beta^{(n)} = \beta^{(n-1)} - \left(E\left[\frac{\partial^2 L}{\partial \beta_j \partial \beta_k} \right] \right)^{-1} \left. \frac{\partial L}{\partial \beta} \right|_{\beta^{(n-1)}}$$

$$= \beta^{(n-1)} - \left(-E\left[\frac{\partial L}{\partial \beta_j} \frac{\partial L}{\partial \beta_k} \right] \right)^{-1} \left. \frac{\partial L}{\partial \beta} \right|_{\beta^{(n-1)}}$$

$$= \beta^{(n-1)} + \left(E\left[\frac{\partial L}{\partial \beta_j} \frac{\partial L}{\partial \beta_k} \right] \right)^{-1} \left. \frac{\partial L}{\partial \beta} \right|_{\beta^{(n-1)}} \quad (4.6)$$

Selanjutnya, didefinisikan I_{jk} sebagai berikut:

$$I_{jk} = E\left[\frac{\partial L}{\partial \beta_j} \frac{\partial L}{\partial \beta_k} \right]$$

$$= E\left[\sum_{i=1}^{n} \frac{y_i - \pi_i}{\pi_i (1-\pi_i)} x_{ji} \left(\frac{\partial \pi_i}{\partial \eta_i} \right) \sum_{l=1}^{n} \frac{y_l - \pi_l}{\pi_l (1-\pi_l)} x_{kl} \left(\frac{\partial \pi_l}{\partial \eta_l} \right) \right]$$

Karena y_i saling bebas, maka $E\left[(y_i - \pi_i)(y_l - \pi_l) \right] = 0$ untuk $i \neq l$ sehingga menghasilkan $E\left[\left(\frac{y_i - \pi_i}{\pi_i (1-\pi_i)} x_{ji} \left(\frac{\partial \pi_i}{\partial \eta_i} \right) \right) \left(\frac{y_l - \pi_l}{\pi_l (1-\pi_l)} x_{kl} \left(\frac{\partial \pi_l}{\partial \eta_l} \right) \right) \right] = 0$ untuk $i \neq l$.

Jadi diperoleh

$$I_{jk} = \sum_{i=1}^{n} E\left[\frac{(y_i - \pi_i)^2}{\left(\pi_i (1-\pi_i) \right)^2} x_{ji} x_{ki} \left(\frac{\partial \pi_i}{\partial \eta_i} \right)^2 \right]$$

$$= \sum_{i=1}^{n} \frac{E\left[(y_i - \pi_i)^2 \right]}{\left(\pi_i (1-\pi_i) \right)^2} x_{ji} x_{ki} \left(\frac{\partial \pi_i}{\partial \eta_i} \right)^2$$

$$= \sum_{i=1}^{n} \frac{Var(y_i)}{\left(\pi_i (1-\pi_i) \right)^2} x_{ji} x_{ki} \left(\frac{\partial \pi_i}{\partial \eta_i} \right)^2$$

$$= \sum_{i=1}^{n} \frac{\pi_i (1-\pi_i)}{\left(\pi_i (1-\pi_i) \right)^2} x_{ji} x_{ki} \left(\frac{\partial \pi_i}{\partial \eta_i} \right)^2$$

$$= \sum_{i=1}^{n} \frac{x_{ji} x_{ki}}{\pi_i (1-\pi_i)} \left(\frac{\partial \pi_i}{\partial \eta_i} \right)^2 \quad (4.7)$$

dalam bentuk matriks, dapat ditulis sebagai berikut:

$$I = X'WX$$

W merupakan matriks diagonal berordo $n \times n$ dengan elemen diagonal

$$w_i = \frac{1}{\pi_i(1-\pi_i)}\left(\frac{\partial \pi_i}{\partial \eta_i}\right)^2$$

$$= \frac{\left(\phi(x_i'\beta)\right)^2}{\Phi(x_i'\beta)\left(1-\Phi(x_i'\beta)\right)}$$

dengan $\phi(x_i'\beta) = \frac{\partial \pi_i}{\partial \eta_i}$. Jadi, I adalah matriks berordo $p \times p$ yang elemen-elemennya adalah

$$I_{jk} = \sum_{i=1}^{n} w_i x_{ji} x_{ki}$$

karena $I_{jk} = E\left[\frac{\partial L}{\partial \beta_j}\frac{\partial L}{\partial \beta_k}\right]$, maka persamaan (4.5) dapat ditulis dengan:

$$\beta^{(n)} = \beta^{(n-1)} + \left(I_{jk}^{(n-1)}\right)^{-1} \left.\frac{\partial L}{\partial \beta}\right|_{\beta^{(n-1)}}$$

yang berarti

$$I_{jk}^{(n-1)}\beta^{(n)} = I_{jk}^{(n-1)}\beta^{(n-1)} + \left.\frac{\partial L}{\partial \beta}\right|_{\beta^{(n-1)}} \qquad (4.8)$$

Dengan mensubstitusikan persamaan (4.7) dan (4.5), ruas kanan dari persamaan (4.8) di atas, dapat dinyatakan sebagai berikut:

$$\sum_{l=1}^{m}\sum_{i=1}^{n}\frac{x_{ji}x_{ki}}{\pi_i(1-\pi_i)}\left(\frac{\partial \pi_i}{\partial \eta_i}\right)^2 \beta_k^{(n-1)} + \sum_{i=1}^{n}\frac{y_i-\pi_i}{\pi_i(1-\pi_i)}x_{ji}\left(\frac{\partial \pi_i}{\partial \eta_i}\right)$$

$$= \sum_{i=1}^{n}\left(\left(\frac{1}{\pi_i(1-\pi_i)}x_{ji}\left(\frac{\partial \pi_i}{\partial \eta_i}\right)^2\right)\left(\sum_{k=1}^{m}x_{ki}\beta_k^{(n-1)} + (y_i-\pi_i)\left(\frac{\partial \eta_i}{\partial \pi_i}\right)\right)\right) \qquad (4.9)$$

atau dalam bentuk matriks menjadi

$$X'Wz \tag{4.10}$$

dimana z adalah matriks berordo $n\times 1$ dengan elemen ke-*i* adalah

$$\begin{aligned} z_i &= \sum_{k=1}^{m} x_{ki}\beta_k^{(n-1)} + (y_i - \pi_i)\left(\frac{\partial \eta_i}{\partial \pi_i}\right) \\ &= \eta_i + (y_i - \pi_i)\left(\frac{\partial \theta_i}{\partial \pi_i}\right) \\ &= \eta_i + \frac{y_i - \pi_i}{\pi_i(1-\pi_i)} \end{aligned}$$

Jadi, persamaan (4.8) dapat ditulis sebagai berikut:

$$\begin{aligned} (X'WX)\beta &= X'Wz \\ \beta &= (X'WX)^{-1}(X'Wz) \\ \hat{\beta}_{ML} &= \left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right) \end{aligned} \tag{4.11}$$

dengan $\hat{W} = diag\left(\frac{\left(\phi(x_i'\beta)\right)^2}{\Phi(x_i'\beta)\left(1-\Phi(x_i'\beta)\right)}\right)$, $\hat{z} = \hat{\eta}_i + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$ dan $\hat{\pi}_i = \Phi(x_i'\beta)$ yang dievaluasi dari $\beta^{(n-1)}$.

4.2 Estimasi Parameter Model Regresi Probit Ridge

Pada model regresi probit, berlaku

$$y_i = \begin{cases} 1 & \text{jika } \mathrm{y}_i^* > 0 \\ 0 & \text{untuk } \mathrm{y}_i^* \text{ lainnya} \end{cases}$$

dan

$$y_i^* = x_i'\beta + \varepsilon \tag{4.12}$$

dimana y_i^* adalah variabel laten. Dengan menggunakan metode *maximum likelihood* telah didapatkan estimator sebagai berikut:

$$\hat{\beta}_{ML} = \left(X'\hat{W}X \right)^{-1} \left(X'\hat{W}\hat{z} \right)$$

dengan $\hat{W} = diag\left(\frac{\left(\phi\left(x_i'\beta\right)\right)^2}{\Phi\left(x_i'\beta\right)\left(1-\Phi\left(x_i'\beta\right)\right)} \right)$ dan $\hat{z} = \hat{\eta}_i + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i\left(1-\hat{\pi}_i\right)}$.

Ketika variabel-variabel prediktor mengandung multikolinieritas, maka estimator $\hat{\beta}_{ML}$ akan menghasilkan *error* yang lebih tinggi. Untuk itu, digunakan estimator $\hat{\beta}_{RR}$ sebagai alternatif. Langkah pertama untuk mendapatkan estimator $\hat{\beta}_{RR}$ yaitu dengan menuliskan model regresi dalam bentuk kanonik untuk mempermudah proses perhitungan. Asumsikan bahwa Λ merupakan matriks berorde $p \times p$ dimana anggota diagonal utamanya merupakan nilai eigen dari matriks $X'\hat{W}X$ dan D merupakan matriks orthogonal dimana $D' = D^{-1}$ sedemikian sehingga $D'D = DD' = I$ dan $D'X'\hat{W}XD = \Lambda$. Dengan mendefinisikan $X^* = XD$ dan $\gamma = D'\beta$, maka bentuk kanonik dari persamaan (4.12) yaitu

$$\begin{aligned} Y^* &= X\beta + \varepsilon_i \\ &= XDD'\beta + \varepsilon \\ &= X^*\gamma + \varepsilon \end{aligned}$$

sehingga estimasi parameter γ dengan metode *maximum likelihood* adalah

$$\begin{aligned} \hat{\beta}_{ML} &= \left(X'\hat{W}X \right)^{-1} \left(X'\hat{W}\hat{z} \right) \\ D\hat{\gamma}_{ML} &= \left(X'\hat{W}X \right)^{-1} \left(X'\hat{W}\hat{z} \right) \\ \left(X'\hat{W}X \right) D\hat{\gamma}_{ML} &= \left(X'\hat{W}X \right)\left(X'\hat{W}X \right)^{-1} \left(X'\hat{W}\hat{z} \right) \\ \left(X'\hat{W}X \right) D\hat{\gamma}_{ML} &= I\left(X'\hat{W}\hat{z} \right) \\ X'\hat{W}XD\hat{\gamma}_{ML} &= X'\hat{W}\hat{z} \\ D'X'\hat{W}XD\hat{\gamma}_{ML} &= D'X'\hat{W}\hat{z} \end{aligned}$$

$$(XD)'\hat{W}XD\hat{\gamma}_{ML} = (XD)'\hat{W}\hat{z}$$

$$X^{*\prime}\hat{W}X^{*}\hat{\gamma}_{ML} = X^{*\prime}\hat{W}\hat{z}$$

$$\left(X^{*\prime}\hat{W}X^{*}\right)^{-1}\left(X^{*\prime}\hat{W}X^{*}\right)\hat{\gamma}_{ML} = \left(X^{*\prime}\hat{W}X^{*}\right)^{-1} X^{*\prime}\hat{W}\hat{z}$$

$$I\hat{\gamma}_{ML} = \left(X^{*\prime}\hat{W}X^{*}\right)^{-1} X^{*\prime}\hat{W}\hat{z}$$

$$\hat{\gamma}_{ML} = \left(X^{*\prime}\hat{W}X^{*}\right)^{-1} X^{*\prime}\hat{W}\hat{z} \tag{4.13}$$

$$X^{*\prime}\hat{W}\hat{z} = \left(X^{*\prime}\hat{W}X^{*}\right)\hat{\gamma}_{ML}$$

diasumsikan $D'X'\hat{W}XD = \Lambda$, maka persamaan (4.13) menjadi

$$\hat{\gamma}_{ML} = \left(X^{*\prime}\hat{W}X^{*}\right)^{-1} X^{*\prime}\hat{W}\hat{z}$$

$$= \left((XD)'\hat{W}(XD)\right)^{-1} X^{*\prime}\hat{W}\hat{z}$$

$$= \left(D'X'\hat{W}XD\right)^{-1} X^{*\prime}\hat{W}\hat{z}$$

$$\hat{\gamma}_{ML} = \Lambda^{-1}X^{*\prime}\hat{W}\hat{z} \tag{4.14}$$

dengan menambahkan matriks K yang merupakan matriks diagonal yang elemen diagonalnya berupa parameter ridge $k \geq 0$ pada Λ, maka persamaan (4.14) menjadi

$$\hat{\gamma}_{RR} = (\Lambda + K)^{-1} X^{*\prime}\hat{W}\hat{z} \tag{4.15}$$

Didefinisikan $A = \Lambda + K$, sehingga persamaan (4.15) menjadi

$$\hat{\gamma}_{RR} = (\Lambda + K)^{-1} X^{*\prime}\hat{W}\hat{z}$$

$$= A^{-1}X^{*\prime}\hat{W}\hat{z}$$

$$= A^{-1}\left(X^{*\prime}\hat{W}X^{*}\right)\hat{\gamma}_{ML}$$

$$= A^{-1}\left((XD)'\hat{W}XD\right)\hat{\gamma}_{ML}$$

$$= A^{-1}\left(D'X'\hat{W}XD\right)\hat{\gamma}_{ML}$$

$$= A^{-1}\Lambda\hat{\gamma}_{ML}$$

Karena $\gamma = D'\beta$, maka $\hat{\gamma}_{RR} = D'\hat{\beta}_{RR}$ dan $\hat{\beta}_{RR} = D\hat{\gamma}_{RR}$. Sehingga,

$$
\begin{aligned}
\hat{\beta}_{RR} &= D\hat{\gamma}_{RR} \\
&= D\left(A^{-1}\Lambda\hat{\gamma}_{ML}\right) \\
&= D\left((\Lambda + K)^{-1}\Lambda\left(\Lambda^{-1}X^{*\prime}\hat{W}\hat{z}\right)\right) \\
&= D\left(\left(D'X'\hat{W}XD + K\right)^{-1} ID'X'\hat{W}\hat{z}\right) \\
&= D\left(\left(D'X'\hat{W}XD + D'DKD'D\right)^{-1} D'X'\hat{W}\hat{z}\right) \\
&= D\left(\left(D'\left(X'\hat{W}X + DKD'\right)D\right)^{-1} D'X'\hat{W}\hat{z}\right) \\
&= DD^{-1}\left(X'\hat{W}X + DKD'\right)^{-1}(D')^{-1} D'X'\hat{W}\hat{z} \\
&= I\left(X'\hat{W}X + DKD'\right)^{-1} IX'\hat{W}\hat{z} \\
&= \left(X'\hat{W}X + DKD'\right)^{-1} X'\hat{W}\hat{z} \\
&= \left(X'\hat{W}X + DKD'\right)^{-1} X'\hat{W}X\beta_{ML}
\end{aligned}
\tag{4.16}
$$

Misalkan $K = kI,\ k \geq 0$, maka persamaan (4.16) menjadi

$$
\begin{aligned}
\hat{\beta}_{RR} &= \left(X'\hat{W}X + DkID'\right)^{-1} X'\hat{W}X\beta_{ML} \\
&= \left(X'\hat{W}X + DkD'\right)^{-1} X'\hat{W}X\beta_{ML} \\
&= \left(X'\hat{W}X + kDD'\right)^{-1} X'\hat{W}X\beta_{ML} \\
&= \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\beta_{ML}
\end{aligned}
\tag{4.17}
$$

Sehingga, dari persamaan (4.17) didapatkan estimasi parameter β dari model regresi probit ridge

$$\hat{\beta}_{RR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\beta_{ML} \tag{4.18}$$

4.3 Aplikasi pada Data yang Mengandung Multikolinieritas

Variabel yang akan diteliti dalam penelitian ini yaitu curah hujan sebagai variabel respon (Y) dengan simbol 1 untuk curah hujan lebih dari 150 mm, dan 0 untuk curah hujan kurang dari 150 mm, temperatur udara sebagai variabel prediktor pertama (X_1), dan kelembaban udara sebagai variabel prediktor kedua (X_2). Pada skripsi Sa'adah (2011), telah dilakukan normalisasi data pada data X_1. Data dapat dilihat pada lampiran 2.

4.3.1 Uji Multikolinieritas

Sebelum mengestimasi parameter model regresi probit ridge, terlebih dahulu harus dilakukan uji multikolinieritas untuk mengetahui apakah terdapat multikolinieritas antara variabel prediktor. Untuk mengetahui adanya multikolinieritas pada data X_1 dan X_2, dilakukan perhitungan nilai koefisien korelasi (uji Pearson). Dengan menggunakan SPSS 15, diperoleh hasil sebagai berikut:

Tabel 4.1 Nilai Koefisien Korelasi

		X1	X2
X1	Pearson Correlation	1	.467**
	Sig. (2-tailed)		.004
	N	36	36
X2	Pearson Correlation	.467**	1
	Sig. (2-tailed)	.004	
	N	36	36

Nilai koefisien korelasi pada tabel di atas signifikan dengan taraf signifikansi sebesar $\alpha(0,05)$. Berdasarkan Tabel 4.1 tersebut dapat diketahui bahwa terdapat korelasi yang cukup tinggi antara variabel X_1 dan X_2, yakni sebesar 0,467 yang berarti terdapat multikolinieritas antara variabel X_1 dan X_2.

Pengujian multikolinieritas juga dapat dilakukan dengan menghitung nilai VIF dengan persamaan

$$VIF = \left(\frac{1}{n-1}(X'X) \right)^{-1} \tag{4.19}$$

Perhitungan dengan persamaan (4.19) melalui program MATLAB menghasilkan nilai VIF sebesar 44,7419. Karena nilai VIF lebih dari 10, maka dapat dikatakan bahwa terdapat multikolinieritas antar variabel prediktor X_1 dan X_2.

4.3.2 Model Regresi Probit Ridge dari Data dan Estimasi Parameter

Perhitungan nilai koefisien korelasi dan nilai VIF sebelumnya mengindikasikan adanya multikolinieritas antara temperatur dan kelembaban udara, sehingga data ini dapat dimodelkan dengan model regresi probit ridge. Estimasi parameter model regresi probit ridge $\left(\hat{\beta}_{RR}\right)$ dari data curah hujan, temperatur, dan kelembaban udara ini dilakukan dengan menggunakan bantuan program MATLAB yang terlampir pada lampiran 3.

Untuk mendapatkan $\hat{\beta}_{RR}$, terlebih dahulu dilakukan estimasi dengan menggunakan metode *maximum likelihood*. Metode ini menghasilkan estimator

$$\beta_{ML} = \left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right)$$

dengan $\hat{W} = diag\left(\frac{\left(\phi\left(x_i'\beta\right)\right)^2}{\Phi\left(x_i'\beta\right)\left(1-\Phi\left(x_i'\beta\right)\right)}\right)$ dan $\hat{z} = \hat{\eta}_i + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i\left(1-\hat{\pi}_i\right)}$ dengan $\hat{\pi}_i = \Phi\left(x_i'\beta\right)$ yang dievaluasi dari $\beta^{(n-1)}$. Sehingga langkah pertama yang harus dilakukan yakni menentukan $\beta^{(0)}$, yaitu nilai awal β yang digunakan untuk menghitung $\hat{\pi}_i^{(0)}$. Estimasi $\hat{\pi}_i^{(0)}$ yang diperoleh, digunakan untuk perhitungan $\hat{W}^{(0)}$ dan $\hat{z}^{(0)}$, sehingga $\beta^{(1)}$ dapat dihitung. Selanjutnya, $\beta^{(1)}$ digunakan untuk menghitung $\hat{\pi}_i^{(1)}$. $\hat{\pi}_i^{(1)}$ digunakan untuk menghitung $\hat{W}^{(1)}$, $\hat{z}^{(1)}$, dan $\beta^{(2)}$. Perhitungan berlanjut hingga β yang dihasilkan telah konvergen, atau dapat dikatakan $\beta^{(n)} - \beta^{(n-1)} \approx 0$. Adapun perhitungan dengan menggunakan program MATLAB menghasilkan β_{ML} sebagai berikut:

$$\beta_{ML} = \begin{bmatrix} -4{,}598496 \\ 2{,}053346 \\ 0{,}082596 \end{bmatrix}$$

Langkah selanjutnya setelah mendapatkan β_{ML} yaitu menghitung $\hat{\beta}_{RR}$ dengan rumus

$$\hat{\beta}_{RR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\beta_{ML} \qquad (4.20)$$

Untuk mendapatkan $\hat{\beta}_{RR}$, maka harus ditentukan parameter ridge (k) terlebih dahulu. Dalam hal ini peneliti menggunakan

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

dimana $q_j = \dfrac{\lambda_{\max}}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_j^2}$, dengan $\lambda_{\max}$ adalah nilai eigen maksimum dari $X'\hat{W}X$, $\hat{\alpha}_j^2$ didefinisikan sebagai elemen ke-*j* dari $\gamma\beta_{ML}$, dimana γ adalah vektor eigen sedemikian hingga $X'\hat{W}X = \gamma'\Lambda\gamma$, dengan Λ yang merupakan matriks diagonal dengan elemen diagonal berupa λ_j, dan $\hat{\sigma}^2$ merupakan jumlah kuadrat sisa dibagi dengan derajat bebas, atau dapat dituliskan dengan

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(\sqrt{-2\left(y_i \log \hat{\pi}_i + (1-y_i)\log(1-\hat{\pi}_i)\right)}\right)^2}{n-p-1}$$

Karena

$$q = \begin{bmatrix} -0{,}20547 \\ -0{,}78571 \\ 5{,}26301 \end{bmatrix}$$

maka didapatkan

$$k = \max \frac{1}{q_j} = 0{,}19$$

Setelah memperoleh parameter ridge, maka dilakukan perhitungan $\hat{\beta}_{RR}$ dengan menggunakan persamaan (4.15). Melalui program MATLAB dihasilkan

$$\hat{\beta}_{RR} = \begin{bmatrix} -1{,}355 \\ 1{,}268 \\ 0{,}021 \end{bmatrix}$$

Sehingga model regresi untuk probabilitas terjadinya hujan yang dipengaruhi oleh temperatur dan kelembaban udara adalah

$$y_i^* = -1{,}355 + 1{,}268X_{1i} + 0{,}021X_{2i}$$

dimana y_i^* merupakan variabel laten sedemikian hingga

$$\hat{y}_i = \begin{cases} 1 & \text{jika } y_i^* > 0 \\ 0 & \text{jika } y_i^* \le 0 \end{cases}$$

dengan

$\hat{y}_i$ = Estimasi terjadinya hujan

X_1 = Temperatur udara

X_2 = Kelembaban udara

4.3.3 Uji Multikolinieritas dengan Parameter Ridge

Untuk mengetahui apakah multikolinieritas antar variabel prediktor telah teratasi, maka dilakukan uji multikolinieritas kembali setelah perhitungan parameter ridge pada subbab sebelumnya. Dalam hal ini, peneliti akan menghitung nilai VIF setelah didapatkan parameter ridge $k = 0{,}19$. Nilai VIF dapat diperoleh dari persamaan

$$VIF = \left(\frac{1}{n-1}(X'X) + DKD'\right)^{-1}\left(\frac{1}{n-1}(X'X)\right)\left(\frac{1}{n-1}(X'X) + DKD'\right)^{-1}$$

dengan mengambil $K = kI$, maka

$$\begin{aligned} VIF &= \left(\frac{1}{n-1}(X'X) + DkID'\right)^{-1}\left(\frac{1}{n-1}(X'X)\right)\left(\frac{1}{n-1}(X'X) + DkID'\right)^{-1} \\ &= \left(\frac{1}{n-1}(X'X) + kDID'\right)^{-1}\left(\frac{1}{n-1}(X'X)\right)\left(\frac{1}{n-1}(X'X) + kDID'\right)^{-1} \\ &= \left(\frac{1}{n-1}(X'X) + kDD'\right)^{-1}\left(\frac{1}{n-1}(X'X)\right)\left(\frac{1}{n-1}(X'X) + kDD'\right)^{-1} \\ &= \left(\frac{1}{n-1}(X'X) + kI\right)^{-1}\left(\frac{1}{n-1}(X'X)\right)\left(\frac{1}{n-1}(X'X) + kI\right)^{-1} \end{aligned}$$

Perhitungan yang dilakukan dengan bantuan program MATLAB menghasilkan VIF yang sangat kecil, yaitu sebesar 0,6036. Sehingga dari nilai

VIF ini dapat dikatakan bahwa dengan menggunakan parameter rigde $k = 0{,}19$, multikolinieritas antar variabel bebas X_1 dan X_2 dapat teratasi dengan baik.

4.3.4 Analisis Hasil Estimasi Curah Hujan

Untuk mengetahui ketepatan model yang telah didapatkan dengan data curah hujan, temperatur, dan kelembaban udara, maka dilakukan estimasi terjadinya hujan pada tahun 2007 hingga 2009 dengan menggunakan data temperatur dan kelembaban yang telah diperoleh dan digunakan untuk mendapatkan model tersebut. Adapun estimasi dari nilai y_i^*, π_i, dan y_i pada tahun 2007 hingga 2009 adalah sebagai berikut:

Tabel 4.2 Estimasi Probabilitas Terjadinya Hujan Tahun 2007

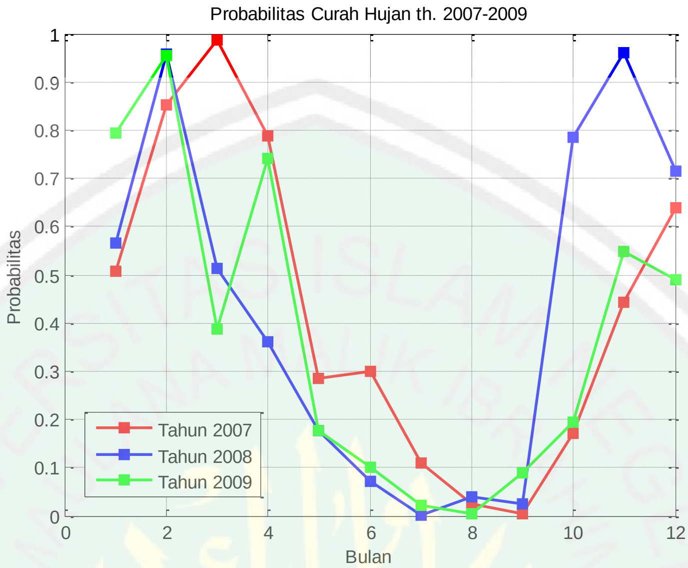
No.	Bulan dan Tahun		$\hat{y}_i^*$	y_i	$\hat{y}_i$	$\hat{\pi}_i$
1.	Januari		0,0198	0	1	0,5079
2.	Februari		1,0430	1	1	0,8515
3.	Maret		2,2126	1	1	0,9865
4.	April		0,7980	1	1	0,7876
5.	Mei	2	-0,5682	0	0	0,2849
6.	Juni	0	-0,5241	0	0	0,3001
7.	Juli	0	-1,2354	0	0	0,1083
8.	Agustus	7	-1,9543	0	0	0,0253
9.	September		-2,6324	0	0	0,0042
10.	Oktober		-0,9489	0	0	0,1713
11.	November		-0,1409	1	0	0,4440
12.	Desember		0,3587	1	1	0,6401

Tabel 4.3 Estimasi Probabilitas Terjadinya Hujan Tahun 2008-2009

No.	Bulan dan Tahun		$\hat{y}_i^*$	y_i	$\hat{y}_i$	$\hat{\pi}_i$
1.	Januari		0,1685	1	1	0,5669
2.	Februari		1,7284	1	1	0,9580
3.	Maret		0,0297	1	1	0,5118
4.	April		-0,3558	0	0	0,3610
5.	Mei	2	-0,9277	0	0	0,1768
6.	Juni	0	-1,4684	0	0	0,0710
7.	Juli	0	-3,1804	0	0	0,0007
8.	Agustus	8	-1,7436	0	0	0,0406
9.	September		-1,9772	0	0	0,0240
10.	Oktober		0,7916	0	1	0,7857
11.	November		1,7496	1	1	0,9599
12.	Desember		0,5701	1	1	0,7157
13.	Januari		0,8192	1	1	0,7937
14.	Februari		1,6859	1	1	0,9541
15.	Maret		-0,2889	0	0	0,3863
16.	April		0,6493	0	1	0,7419
17.	Mei	2	-0,9277	0	0	0,1768
18.	Juni	0	-1,2827	0	0	0,0998
19.	Juli	0	-2,0181	0	0	0,0218
20.	Agustus	9	-2,7031	0	0	0,0034
21.	September		-1,3385	0	0	0,0904
22.	Oktober		-0,8656	0	0	0,1934
23.	November		0,1240	1	1	0,5494
24.	Desember		-0,0237	1	0	0,4906

Tabel 4.2 dan 4.3 di atas menunjukkan bahwa, jika $\hat{y}_i^* > 0$ maka $\hat{y}_i$ akan menunjukkan angka 1, yakni bulan yang sering terjadi hujan. Sedangkan jika

$\hat{y}_i^* \leq 0$ maka $\hat{y}_i$ menunjukkan angka 0, atau bulan yang jarang terjadi hujan. Dalam bentuk grafik, dapat digambarkan sebagai berikut:



Gambar 4.1 Grafik Probabilitas Curah Hujan Tahun 2007-2009

Berdasarkan grafik tersebut dapat disimpulkan bahwa pada tahun 2007, bulan yang sering terjadi hujan dengan perkiraan curah hujan yang relatif tinggi adalah bulan Januari, Februari, Maret, April, dan Desember. Curah hujan tertinggi terjadi pada bulan Maret dengan probabilitas curah hujan sebesar 0,9865. Sedangkan probabilitas curah hujan terendah terjadi pada bulan September, yakni sebesar 0,0042. Pada tahun 2008, perkiraan bulan yang sering terjadi hujan adalah bulan Januari, Februari, Maret, Oktober, November, dan Desember, dengan probabilitas curah hujan tertinggi sebesar 0,9599 terjadi pada bulan November, dan probabilitas curah hujan terendah sebesar 0,0007 terjadi pada bulan Juli. Pada tahun 2009, perkiranan bulan yang sering terjadi hujan adalah bulan Januari, Februari, April, dan November. Probabilitas curah hujan tertinggi pada bulan Februari sebesar 0,9541, dan probabilitas curah hujan terendah pada bulan Agustus sebesar 0,0034. Sehingga, secara umum dapat dikatakan bahwa perkiraan bulan yang sering terjadi hujan dalam satu tahun dengan probabilitas curah hujan yang relatif tinggi yaitu tiga bulan pertama dan tiga bulan terakhir, atau bulan Oktober hingga bulan Maret, yang biasa disebut sebagai musim penghujan.

Pada Tabel 4.2 di atas juga terlihat bahwa terdapat 5 estimasi curah hujan $(\hat{y}_i)$ yang tidak sesuai dengan curah hujan yang sebenarnya (y_i). Hal ini dapat dilihat pada bulan Januari 2007 $(\hat{y}_1)$, bulan November 2007 $(\hat{y}_{11})$, bulan Oktober 2008 $(\hat{y}_{22})$, bulan April 2009 $(\hat{y}_{28})$, dan bulan Desember 2009 $(\hat{y}_{36})$.

Pada bulan Januari 2007, terlihat bahwa $\hat{y}_1$ menunjukkan bulan yang sering terjadi hujan dengan $\hat{\pi}_1$ sebesar 0,5079, sedangkan y_1 menunjukkan bulan yang jarang terjadi hujan. Untuk dapat dikatakan bulan yang jarang terjadi hujan, maka $\hat{\pi}_1$ harus lebih kecil dari 0,5. Sehingga galat yang dihasilkan oleh $\hat{y}_1$ adalah 0,008. Pada bulan November 2007, $\hat{y}_{11}$ menunjukkan bulan yang jarang terjadi hujan dengan $\hat{\pi}_{11}$ sebesar 0,4440. Sehingga galat dari $\hat{y}_{11}$ sebesar 0,056. Selanjutnya pada bulan Oktober 2008, $\hat{y}_{22}$ menunjukkan bulan yang sering terjadi hujan dengan $\hat{\pi}_{22}$ sebesar 0,7857 dan galat sebesar 0,2858. Pada bulan April 2009, $\hat{y}_{28}$ menunjukkan bulan yang sering terjadi hujan dengan $\hat{\pi}_{28}$ sebesar 0,7419 dan galat sebesar 0,2420. Dan pada bulan Desember 2009, $\hat{y}_{36}$ menunjukkan bulan yang jarang terjadi hujan dengan $\hat{\pi}_{36}$ sebesar 0,4906 dan galat sebesar 0,0094.

4.4 Konsep Multikolinieritas dalam Al-Quran

Pada bab sebelumnya, telah ditunjukkan bahwa kelak manusia akan dikatakan sebagai ahli surga jika ia beriman dan beramal shalih. Tingkat keimanan dan amal shalih ini dapat diketahui dari hubungan manusia dengan Allah Swt. (*hablum minallah*) dan dengan hubungan manusia dengan manusia yang lain (*hablum minannas*). Jika seseorang semakin taat kepada Allah Swt. dan semakin baik hubungannya dengan sesama manusia, maka semakin tinggi pula tingkat keimanan dan amal shalihnya.

Allah berfirman dalam al-Quran surat al-Nisa'/4:1 yang berbunyi:

يَـٰٓأَيُّهَا ٱلنَّاسُ ٱتَّقُوا۟ رَبَّكُمُ ٱلَّذِى خَلَقَكُم مِّن نَّفْسٍ وَٰحِدَةٍ وَخَلَقَ مِنْهَا زَوْجَهَا وَبَثَّ مِنْهُمَا رِجَالًا كَثِيرًا وَنِسَآءً ۚ وَٱتَّقُوا۟ ٱللَّهَ ٱلَّذِى تَسَآءَلُونَ بِهِۦ وَٱلْأَرْحَامَ ۚ إِنَّ ٱللَّهَ كَانَ عَلَيْكُمْ رَقِيبًا ۝١

"Hai sekalian manusia, bertakwalah kepada Tuhanmu yang telah menciptakan kamu dari seorang diri, dan dari padanya Allah menciptakan istrinya, dan dari pada keduanya Allah mengembangbiakkan laki-laki dan perempuan yang banyak. Dan bertakwalah kepada Allah yang dengan (mempergunakan) nama-Nya kamu saling meminta satu sama lain, dan (peliharalah) hubungan silaturrahim. Sesungguhnya Allah selalu menjaga dan mengawasi kamu" (QS. al-Nisa'/4:1).

Melalui ayat tersebut, Allah Swt. menyerukan kepada seluruh umat manusia untuk bertakwa kepada-Nya dan menjaga silaturrahim antar sesama manusia. Dalam hal ini, manusia sebagai variabel respon atau variabel terikat (Y), terikat atau bergantung kepada Allah Swt. sebagai variabel prediktor atau variabel bebas pertama (X_1), dan manusia lain sebagai variabel prediktor kedua (X_2).

Dalam analisis regresi, ketika terdapat hubungan yang linier (berbanding lurus) antara variabel prediktor yang satu dengan variabel prediktor lainnya, maka hal ini disebut dengan multikolinieritas. Pada hubungan manusia (Y) dengan

Allah Swt. (X_1) dan manusia lainnya (X_2) di atas, dapat dikatakan bahwa terjadi multikolinieritas di dalamnya. Hal ini dapat terjadi karena manusia lain, yang dalam hal ini sebagai X_1, selain memiliki hubungan dengan variabel respon Y, juga memiliki hubungan atau terikat kepada Allah Swt. sebagai X_2. Hal inilah yang dikatakan sebagai multikolinieritas dalam analisis regresi.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan pembahasan pada bab sebelumnya, maka dapat dibuat kesimpulan sebagai berikut:

1. Dengan menggunakan metode *maximum likelihood*, diperoleh estimator untuk parameter model regresi probit ridge sebagai berikut:

$$\hat{\beta}_{RR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\hat{\beta}_{ML}$$

dimana

$$\hat{\beta}_{ML} = \left(X'\hat{W}X\right)^{-1}\left(X'\hat{W}\hat{z}\right)$$

dengan $\hat{W} = diag\left(\frac{\left(\phi(x_i'\beta)\right)^2}{\Phi(x_i'\beta)\left(1-\Phi(x_i'\beta)\right)}\right)$ dan $\hat{z} = \hat{\eta}_i + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$.

2. Hasil estimasi parameter model regresi probit ridge pada data curah hujan di Karangploso Malang dengan $k = \max\left(\frac{1}{q_j}\right)$ dimana $q_j = \frac{\lambda_{\max}}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_j^2}$, menghasilkan $\hat{\beta}_0 = -1,355$, $\hat{\beta}_1 = 1,268$, dan $\hat{\beta}_2 = 0,021$. Sehingga model regresi untuk peluang terjadinya hujan yang dipengaruhi oleh temperatur dan kelembaban udara adalah

$$y_i^* = -1,355 + 1,268X_{1i} + 0,021X_{2i}$$

dimana y_i^* merupakan variabel laten sedemikian hingga

$$\hat{y}_i = \begin{cases} 1 & \text{jika } y_i^* > 0 \\ 0 & \text{jika } y_i^* \le 0 \end{cases}$$

dengan

$\hat{y}_i$ = Estimasi terjadinya hujan

X_1 = Temperatur udara

X_2 = Kelembaban udara

Setelah menghitung nilai VIF, dapat dikatakan bahwa multikolinieritas antar variabel prediktor X_1 dengan X_2 dapat teratasi dengan nilai VIF sebesar 0,6036.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, beberapa saran yang dapat dilakukan untuk penelitian-penelitian selanjutnya antara lain:

1. Menambah banyaknya data yang digunakan untuk mengestimasi parameter model regresi probit ridge sehingga estimasi yang dihasilkan lebih akurat.
2. Menggunakan parameter ridge (k) yang berbeda-beda dalam mengestimasi parameter dan menganalisis perbedaannya.

DAFTAR PUSTAKA

Aziz, A. 2010. *Ekonometrika Teori & Praktik Eksperimen dengan MATLAB*. Malang: UIN-Maliki Press.

Cameron, A.C. dan Trivedi, P.K. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

Candra, Y. 2009. *Pembentukan Model Probit Bivariat*. Skripsi tidak dipublikasikan. Semarang: Universitas Diponegoro.

Draper, N. dan Smith, H. 1992. *Analisis Regresi Terapan*. Jakarta: PT. Gramedia Pustaka Umum.

Dudewicz, E.J. dan Mishra, S.N. 1995. *Statistika Matematika Modern*. Bandung: ITB.

Gujarati, D. N. 2006a. *Dasar Dasar Ekonometrika Jilid I*. Jakarta: Erlangga.

Gujarati, D. N. 2006b. *Dasar Dasar Ekonometrika Jilid II*. Jakarta: Erlangga.

Hoerl, A.E. dan Kennard, R.W. 1970. Ridge Regression: Biased Estimation For Nonorthogonal Problems. *Technometrics*, 12: 55-67.

Katsir, I. 2000. *Tafsir Ibnu Katsir Juz 1*. Bandung: Sinar Baru Algesindo.

Kibria, B.M.G. dan Saleh, A.K.Md.E. 2012. Improving the Estimators of the Parameters of a Probit Regression Model: A Ridge Regression Approach. *Journal of Statistical Planning and Inference*, 142: 1421-1435.

Kutner, M.H., Nachtsheim, C.J., Netter, J., dan Li, W. 2005. *Applied Linear Statistical Model*. New York: Mc Graw Hill.

Locking, H., Mansson, K., dan Shukur, G. 2011. Performance of Some Ridge Parameters for Probit Regression: with Application on Swedish Job Search Data. *Journal of Computational Econometrics*, 40: 415-433.

Myers, R.H., Montgomery, D.C., Vining, G.G., dan Robinson T.J. 2010. *Generalized Linear Models With Applications in Engineering and the Sciences, Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Nisa', H.M. 2014. *Metode Estimasi Jackknifed Riedge Regression pada Model Regresi Linier Berganda*. Skripsi tidak dipublikasikan. Malang: UIN Maulana Malik Ibrahim Malang.

Sa'adah, F.Z. 2011. *Analisis Regresi Dummy Variable Model Probit (Kasus pada Estimasi Hujan di Karangploso Malang)*. Skripsi tidak dipublikasikan. Malang: UIN Maulana Malik Ibrahim Malang.

Setiawan dan Kusrini, D.E. 2010. *Ekonometrika*. Yogyakarta: Andi.

Suharyadi dan Purwanto. 2009. *Statistika Untuk Ekonomi dan Keuangan Modern*. Jakarta: Salemba Empat.

Supranto, J. 2005a. *Ekonometri Buku I*. Bogor: Ghalia Indonesia.

Supranto, J. 2005b. *Ekonometri Buku II*. Bogor: Ghalia Indonesia.

Widhiarso, W. 2012. *Berkenalan dengan Regresi Probit*. Yogyakarta: Universitas Gadjah Mada.

LAMPIRAN-LAMPIRAN

Lampiran 1 Data Curah Hujan, Temperatur, dan Kelembaban Udara di Karangploso Malang Tahun 2007-2008

Bulan dan Tahun		Curah Hujan	Temperatur Udara	Kelembaban Udara
Januari		129	20,4	48
Februari		182	20,9	52
Maret		173	21,2	42
April		235	20,8	53
Mei	2	6	20,1	43
Juni	0	15	19,7	60
Juli	0	7	18,8	48
Agustus	7	1	17,8	44
September		10	17,7	27
Oktober		61	19,7	40
November		272	20,3	47
Desember		423	20,6	55
Januari		206	20,4	55
Februari		315	21,1	58
Maret		460	20,2	61
April		66	20,1	53
Mei	2	61	19,7	41
Juni	0	2	18,4	43
Juli	0	0	17,2	40
Agustus	8	47	18,1	39
September		8	18,1	28
Oktober		92	21	33
November		174	21,1	59
Desember		241	20,7	56

Lampiran 2 Data Curah Hujan, Temperatur, dan Kelembaban Udara di Karangploso Malang Tahun 2009

Bulan dan Tahun		Curah Hujan	Temperatur Udara	Kelembaban Udara
Januari		258	20,8	54
Februari		435	21,1	56
Maret		81	20,2	46
April		67	20,8	46
Mei	2	62	19,7	41
Juni	0	70	18,9	41
Juli	0	39	17,8	41
Agustus	9	0	17,6	38
September		4	19,4	33
Oktober		35	20,1	29
November		200	20,7	35
Desember		224	20,6	37

Lampiran 3 Hasil Normalisasi Data Temperatur Udara

No.	Bulan dan Tahun		Temperatur Setelah Dinormalkan
1.	Januari	2007	0,28
2.	Februari		1,02
3.	Maret		2,11
4.	April		0,81
5.	Mei		-0,10
6.	Juni		-0,35
7.	Juli		-0,71
8.	Agustus		-1,21
9.	September		-1,46
10.	Oktober		-0,35
11.	November		0,17
12.	Desember		0,43
13.	Januari	2008	0,28
14.	Februari		1,46
15.	Maret		0,07
16.	April		-0,10
17.	Mei		-0,35
18.	Juni		-0,81
19.	Juli		-2,11
20.	Agustus		-0,96
21.	September		-0,96
22.	Oktober		1,14
23.	November		1,46
24.	Desember		0,58
25.	Januari	2009	0,81
26.	Februari		1,46
27.	Maret		0,07
28.	April		0,81
29.	Mei		-0,35
30.	Juni		-0,63
31.	Juli		-1,21
32.	Agustus		-1,70
33.	September		-0,54
34.	Oktober		-0,10
35.	November		0,58
36.	Desember		0,43

Lampiran 4 Program MATLAB untuk Estimasi Parameter Model Regresi Probit Ridge dengan Metode *Maximum Likelihood*

```
clc,clear
format long
Y=[0                      X=[1    .28 48
   1                          1  1.02 52
   1                          1  2.11 42
   1                          1   .81 53
   0                          1  -.10 43
   0                          1  -.35 60
   0                          1  -.71 48
   0                          1 -1.21 44
   0                          1 -1.46 27
   0                          1  -.35 40
   1                          1   .17 47
   1                          1   .43 55
   1                          1   .28 55
   1                          1  1.46 58
   1                          1   .07 61
   0                          1  -.10 53
   0                          1  -.35 41
   0                          1  -.81 43
   0                          1 -2.11 40
   0                          1  -.96 39
   0                          1  -.96 28
   0                          1  1.14 33
   1                          1  1.46 59
   1                          1   .58 56
   1                          1   .81 54
   1                          1  1.46 56
   0                          1   .07 46
   0                          1   .81 46
   0                          1  -.35 41
   0                          1  -.63 41
   0                          1 -1.21 41
   0                          1 -1.70 38
   0                          1  -.54 33
   0                          1  -.10 29
   1                          1   .58 35
   1];                        1   .43 37];

n=length(Y);
b0=[0.00001;0.00001;0.00001];
iterasi=100;
for m=1:iterasi
    for i=1:n
        mu(i)=normcdf(X(i,:)*b0);
        z(i)=norminv(mu(i))+((Y(i)-mu(i))/(mu(i)*(1-mu(i))));
        W(i,i)=1/((1/(mu(i)*(1-mu(i))))^2)*mu(i)*(1-mu(i)));
    end
    Bml=(inv(X'*W*X))*(X'*W*z');
    b0=Bml;
end
[G,D]=eig(X'*W*X);
G'*D*G;
X'*W*X;
S=G*Bml;
```

```
Yml=(X*Bml);
Zml=normcdf(Yml);
for i=1:n
    mu(i)=normcdf(X(i,:)*Bml);
    if Y==0
        d(i)=-1*sqrt(-2*(Y(i)*log(mu(i))+(1-Y(i))*log(1-(mu(i)))));
    else
        d(i)=sqrt(-2*(Y(i)*log(mu(i))+(1-Y(i))*log(1-(mu(i)))));
    end
end
S2=sum(d)/32;
lambda=max(diag(D));
for i=1:length(S)
    q(i)=lambda/((33*S2)+(lambda*S(i)));
    k(i)=1/q(i);
end
k11=max(k);
I=eye(3);
Brr=inv(X'*W*X+k11*I)*(X'*W*X*Bml);
Yrr=(X*Brr);
Zrr=normcdf(Yrr);
format short
for i=1:n
    if Zml(i)<0.5
        YhatML(i)=0;
    else
        YhatML(i)=1;
    end
    if Zrr(i)<0.5
        YhatRR(i)=0;
    else
        YhatRR(i)=1;
    end
end
Bml
Brr
Estimasi=[Yrr Zrr YhatRR']

figure(1)
plot((1:12),Zrr(1:12),'s-r',...
'MarkerSize',5,'LineWidth',2,'MarkerFaceColor','r')
hold on
plot((1:12),Zrr(13:24),'s-b',...
'MarkerSize',5,'LineWidth',2,'MarkerFaceColor','b')
hold on
plot((1:12),Zrr(25:36),'s-g',...
'MarkerSize',5,'LineWidth',2,'MarkerFaceColor','g'), grid on
title('Probabilitas Curah Hujan th. 2007-2009')
xlabel('Bulan')
ylabel('Probabilitas')
legend('Tahun 2007','Tahun 2008','Tahun 2009')
```

Lampiran 5 Program MATLAB untuk Menentukan Nilai VIF

```
clc,clear
y=[     0               x1=[  .28               x2=[  48
        1                   1.02                      52
        1                   2.11                      42
        1                    .81                      53
        0                   -.10                      43
        0                   -.35                      60
        0                   -.71                      48
        0                  -1.21                      44
        0                  -1.46                      27
        0                   -.35                      40
        1                    .17                      47
        1                    .43                      55
        1                    .28                      55
        1                   1.46                      58
        1                    .07                      61
        0                   -.10                      53
        0                   -.35                      41
        0                   -.81                      43
        0                  -2.11                      40
        0                   -.96                      39
        0                   -.96                      28
        0                   1.14                      33
        1                   1.46                      59
        1                    .58                      56
        1                    .81                      54
        1                   1.46                      56
        0                    .07                      46
        0                    .81                      46
        0                   -.35                      41
        0                   -.63                      41
        0                  -1.21                      41
        0                  -1.70                      38
        0                   -.54                      33
        0                   -.10                      29
        1                    .58                      35
        1];                  .43];                    37];

n=36;
for i=1:n
    x1bar=mean(x1);
    x2bar=mean(x2);
    selisih1(i)=(x1(i)-x1bar)^2;
    selisih2(i)=(x2(i)-x2bar)^2;
end
s11=sum(selisih1);
s22=sum(selisih2);
for i=1:n
    x(i,1)=(x1(i)-x1bar)/sqrt(s11);
    x(i,2)=(x2(i)-x2bar)/sqrt(s22);
end
x   %Hasil Pemusatan dan Penskalaan Data
C=x'*x;
K=[0.19 0;0 0.19];
VIFRR=inv((1/35).*C+K)*(1/35).*C*inv((1/35)*C+K)
VIFLS=inv((1/35)*C)
```