

**METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION*
OPERATOR UNTUK MENGATASI MULTIKOLINEARITAS
PADA REGRESI LOGISTIK ORDINAL**

SKRIPSI



Oleh :

NUR ARIF BAHMID

H121 12 273

**PROGRAM STUDI STATISTIKA DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS HASANUDDIN**

MAKASSAR

NOVEMBER 2018



**METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION*
OPERATOR UNTUK MENGATASI MULTIKOLINEARITAS
PADA REGRESI LOGISTIK ORDINAL**

SKRIPSI

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

NUR ARIF BAHMID

H121 12 273

**PROGRAM STUDI STATISTIKA DEPARTEMEN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

UNIVERSITAS HASANUDDIN

MAKASSAR

NOVEMBER 2018



LEMBAR PERNYATAAN KEOTENTIKAN

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

**Metode *Least Absolute Shrinkage and Selection Operator* untuk Mengatasi
Multikolinearitas pada Regresi Logistik Ordinal**

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

Makassar, 30 November 2018



NUR ARIF BAHMID
NIM. H121 12 273



METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR* UNTUK MENGATASI MULTIKOLINEARITAS PADA REGRESI LOGISTIK ORDINAL

Disetujui Oleh:

Pembimbing Utama,



Anisa, S.Si., M.Si.
NIP. 19730227 199802 2 001

Pembimbing Pertama,



Drs. M. Saleh AF, M.Si.
NIP. 19540804 197802 1 001



Pada Tanggal : 30 November 2018

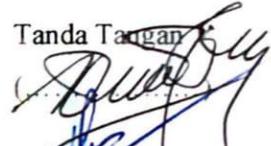
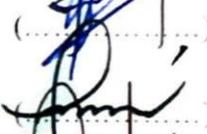
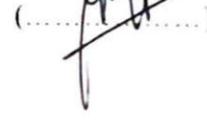
HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : NUR ARIF BAHMID
NIM : H121 12 273
Program Studi : STATISTIKA
Judul : Metode *Least Absolute Shrinkage and Selection Operator*
untuk Mengatasi Multikolinearitas pada Regresi Logistik
Ordinal

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

DEWAN PENGUJI

- | | | Tanda Tangan |
|----|--|--|
| 1. | Ketua : Dr. La Podje Talangko, M.Si. | () |
| 2. | Sekretaris : Sitti Sahriman, S.Si. M.Si. | () |
| 3. | Anggota : Dr. Nirwan Ilyas, M.Si. | () |
| 4. | Anggota : Anisa, S.Si., M.Si. | () |
| 5. | Anggota : Drs. M. Saleh AF, M.Si. | () |

Ditetapkan di : Makassar

Tanggal : 30 November 2018



KATA PENGANTAR

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِیْمِ

Alhamdulillah Rabbil Aalamiin, tiada kata paling indah yang dipanjatkan oleh penulis melainkan rasa syukur kehadiran Allah Subhanahu Wata'ala atas berkat rahmat, karunia dan hidayah-Nya. Shalawat dan salam tak lupa penulis kirimkan untuk Rasulullah Shallallahu 'alaihi wasallam yang telah membawa umatnya dari alam jahiliyah menuju alam yang berilmu seperti sekarang ini. Penulis sangat bersyukur atas segala limpahan nikmat dan berkah, terutama kesehatan, kemudahan, serta kemampuan untuk menyelesaikan tugas akhir ini.

Penyusunan tugas akhir ini tidak luput dari dukungan dan bantuan yang penulis dapatkan dari berbagai pihak, baik itu berupa dukungan moril maupun materil. Oleh karena itu, penulis haturkan ucapan terima kasih yang setulus-tulusnya kepada orang tuaku tercinta dan terkasih, yang sangat kuhormati, Ayahanda **Ir. Basir Halim, MMA.**, dan Ibunda **Hj. Nursiah** yang telah merawat, mendidik dan membesarkan penulis dengan bertabur cinta, kasih sayang, serta dengan ikhlas telah mengiringi juga membimbing setiap langkah ini dengan do'a, restu, dan nasihatnya sehingga terbentuk pribadi penulis seperti sekarang ini. Untuk saudara saudariku **Nur Alim Bahmid, Nur Alif Bahmid, Nur Aliah Bahmid, Nur Afif Bahmid** dan **Ayu Yustika** terimakasih telah menjadi kakak dan adik yang baik bagi penulis. Suatu kesyukuran penuh bagi penulis kepada Allah Subhanahu Wata'ala karena memiliki orangtua dan saudara terhebat seperti ini. Untuk semua keluarga besar penulis, terima kasih atas doa dan dukungannya selama ini

Penghargaan dan ucapan terima kasih dengan penuh ketulusan juga penulis ucapkan kepada :

1. **Hu Prof. Dr. Dwia Aries Tina Palubuhu, MA**, selaku **Rektor** **Universitas Hasanuddin** beserta seluruh jajarannya.



2. **Bapak Dr. Eng. Amiruddin**, selaku **Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin**.
3. **Bapak Prof. Dr. Amir Kamal Amir, M.Sc.**, selaku **Ketua Departemen Matematika**, **Ibu Dr. Erna Tri Herdiani, S.Si., M.Si.**, selaku **Ketua Program Studi Statistika**, dan **segenap dosen pengajar** serta **staf Departemen Matematika** yang membekali ilmu dan kemudahan-kemudahan kepada penulis dalam berbagai hal selama ini menjadi mahasiswa di Departemen Matematika.
4. **Ibu Anisa, S.Si., M.Si.** selaku pembimbing utama dan **Bapak Drs. M. Saleh AF, M.Si.** selaku pembimbing pertama, yang telah meluangkan waktunya bahkan waktu liburnya untuk membimbing, mengajarkan, dan memberikan koreksi serta motivasi kepada penulis dalam penyusunan tugas akhir ini.
5. **Bapak Dr. La Podje Talangko, M.Si.** selaku ketua Tim Penguji, **Ibu Sitti Sahrinan, S.Si. M.Si.** selaku Sekretaris Tim Penguji dan **Bapak Dr. Nirwan Ilyas, M.Si.** selaku Tim Penguji yang telah memberikan koreksi dan saran yang sangat membangun penyusunan tugas akhir ini.
6. Teman-teman seperjuangan menyusun skripsi **Rahmat, Khamid, Adnan, Dekana, Iqbal, Rheo, Eki, Rihul, Tisa, Indah**, dan terkhusus untuk **Boghi Kurniawan Perdana, Sigit Asrianto Sahab**, dan **Ilham Sudirman** tetap semangat kawan dan semoga ALLAH Subhanahu Wata'ala senantiasa memberikan jalan untuk kalian.
7. Teman dan sahabat seperjuangan **Dekana, Khamid, Rahmat, Iswan, Boghi, Sigit, Iqbal, Ilham, Adnan, Anggun, Selpa (Rahimahallah), Dipa**, dan **Vira**.
8. **Statistika 2012** dan **Rekurensi 2012** telah menyemangati dan mewarnai hari-hari penulis selama enam tahun terakhir ini.
9. Teman-teman KKN 90 Posko 16 Polewali **Fikar, Maxi, Ida, Nome'**, dan **Ratih**

semua pihak yang telah banyak berpartisipasi, baik secara langsung maupun tidak langsung dalam penyusunan tugas akhir ini yang tidak sempat penulis sebutkan satu per satu.



Penulis menyadari bahwa tugas akhir ini sangat jauh dari kesempurnaan, sehingga kritik dan saran yang membangun akan penulis terima dengan harapan agar tugas akhir ini dapat memberikan manfaat bagi semua pihak.

Makassar, 30 November 2018

Penulis



PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

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

Nama : Nur Arif Bahmid
NIM : H121 12 273
Program Studi : Statistika
Departemen : Matematika
Fakultas : Matematika dan Ilmu Pengetahuan Alam
Jenis Karya : Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Metode *Least Absolute Shrinkage and Selection Operator* untuk Mengatasi Multikolinearitas pada Regresi Logistik Ordinal

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

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada tanggal 30 November 2018.

Yang menyatakan,



f Bahmid)

ABSTRAK

Multikolinearitas merupakan sebuah masalah yang sering terjadi dalam regresi logistik. Ada beberapa metode yang dapat mengatasi masalah multikolinearitas, diantaranya adalah metode *Least Absolute Shrinkage and Selection Operator*. Metode *Least Absolute Shrinkage and Selection Operator* (LASSO) merupakan teknik regresi yang melakukan penaksiran dengan meminimumkan jumlah kuadrat eror menggunakan suatu kendala. Metode LASSO dengan menggunakan algoritma *Least Angle Regression* (LAR) dapat mengatasi multikolinearitas dan mengestimasi parameter pada regresi logistik ordinal. Penerapan model regresi logistik ordinal dengan metode *LASSO* pada data bayi gizi buruk kabupaten/kota di Provinsi Sulawesi Selatan terdapat empat variabel yang mempengaruhi gizi buruk yaitu pemberian ASI eksklusif, imunisasi hepatitis B, imunisasi campak, dan rumah dengan sanitasi layak.

Kata kunci : multikolinearitas, regresi logistik ordinal, LASSO, algoritma LAR.
data bayi gizi buruk



ABSTRACT

Multicollinearity is a problem that often occurs in a logistic regression. There are several methods that can be applied to overcome multicollinearity problems, including the Least Absolute Shrinkage and Selection Operator method. The Least Absolute Shrinkage and Selection Operator (LASSO) method is a regression technique that estimation by minimizing the residual sum of squared subject to constraints. The LASSO method using the Least Angle Regression (LAR) algorithm can overcome multicollinearity and estimate parameters in ordinal logistic regression. The application of the ordinal logistic regression model with the LASSO method to the data on malnutrition in districts in South Sulawesi has four variables that affect malnutrition, namely exclusive breastfeeding, hepatitis B immunization, measles immunization, and houses with proper sanitation.

Keywords: multicollinearity, ordinal logistic regression , LASSO , LAR algorithm, malnutrition data



DAFTAR ISI

Halaman Judul	ii
Halaman Pernyataan Keotentikan	iii
Halaman Persetujuan Pembimbing	iv
Halaman Pengesahan	v
Kata Pengantar	vi
Persetujuan Publikasi Karya Ilmiah	ix
Abstrak	x
Daftar Isi	xii
Daftar Gambar	xiv
Daftar Tabel	xv
Daftar Lampiran	xvi
1. PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
2. TINJAUAN PUSTAKA	4
2.1 Multikolinearitas	4
2.2 Regresi Logistik.....	4
2.3 Regresi Logistik Ordinal.....	5
2.4 <i>Least Absolute Shrinkage and Selection Operator</i>	7
2.5 Algoritma <i>Least Angle Regression</i>	8
2.6 Validasi Silang	9
2.7 Gizi Buruk.....	9
3. METODE PENELITIAN	11
3.1 Sumber Data.....	11
Identifikasi Variabel.....	11
Metode Analisis	11



4. HASIL DAN PEMBAHASAN	13
4.1 Estimasi Koefisien Regresi Logistik Ordinal dengan Metode <i>Least Absolute Shrinkage and Selection Operator</i>	13
4.2 Penerapan Metode <i>Least Absolute Shrinkage and Selection Operator</i> pada Data Bayi Gizi Buruk di Provinsi Sulawesi Selatan Tahun 2015.....	18
5. PENUTUP	24
5.1 Kesimpulan	24
5.2 Saran	24
DAFTAR PUSTAKA	25
LAMPIRAN.....	27



DAFTAR GAMBAR

Gambar 4.1 Nilai validasi silang dengan menggunakan mode *fraction*..... 20

Gambar 4.2 Nilai validasi silang dengan menggunakan mode *step*..... 21



DAFTAR TABEL

Tabel 4.1	Nilai <i>Variance Inflation Factors</i> variabel prediktor	18
Tabel 4.2	Hasil estimasi parameter pada bayi gizi buruk provinsi Sulawesi Selatan dengan regresi logistik ordinal	19
Tabel 4.3	Nilai korelasi variabel prediktor terhadap sisaan	20
Tabel 4.4	Nilai koefisien regresi logistik ordinal dengan metode <i>LASSO</i>	22



DAFTAR LAMPIRAN

Lampiran 1	Data Bayi Gizi Buruk di Provinsi Sulawesi Selatan Tahun 2015	27
Lampiran 2	Uji Multikolinearitas dengan SPSS 23	28
Lampiran 3	Sintaks Program dan Output Regresi Logistik Ordinal dengan <i>software R 3.3.2</i>	29
Lampiran 4	Sintaks Program dan Output dengan metode <i>LASSO software R</i> 3.3.2	30
Lampiran 5	Nilai untuk Setiap Tahapan Metode <i>LASSO</i>	31



BAB I

PENDAHULUAN

1.1 Latar Belakang

Statistika adalah ilmu yang mempelajari bagaimana merencanakan, mengumpulkan, menganalisis, menginterpretasi, dan mempresentasikan data. Statistika dapat digunakan untuk memprediksi kejadian mendatang serta mengetahui faktor-faktor yang berpengaruh, salah satunya dengan menggunakan analisis regresi. Analisis regresi adalah suatu metode analisis statistik yang bertujuan untuk mengetahui hubungan antar variabel respon (Y) dengan beberapa variabel prediktor (X). Model regresi yang terbentuk bergantung pada skala pengukuran data yang dianalisis, misalnya pada model regresi klasik variabel respon minimal berskala interval, sedangkan model regresi logistik digunakan jika variabel respon merupakan data kategorik.

Regresi logistik digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau beberapa variabel prediktor yang berupa kategorik. Berdasarkan skala data, regresi logistik dibedakan menjadi tiga macam yaitu regresi logistik biner, regresi logistik ordinal, dan regresi logistik multinomial. Regresi logistik ordinal adalah hubungan antara variabel prediktor yang bersifat kategorik dan/atau kontinu dengan dua variabel atau lebih serta variabel respon memiliki lebih dari dua kategori yang berskala ordinal (Agresti, 2007).

Salah satu asumsi regresi logistik adalah tidak terdapat multikolinearitas diantara variabel prediktor yang termasuk dalam model. Multikolinearitas terjadi apabila terdapat hubungan atau korelasi yang tinggi diantara variabel prediktor. Jika multikolinearitas terdapat di dalam persamaan regresi maka mengakibatkan penggunaan *Ordinary Least Square* (OLS) dapat diperoleh tetapi nilai *standart error* akan cenderung membesar seiring dengan meningkatnya kolinearitas antar prediktor dan mendapatkan model yang kurang layak.



Masalah multikolinearitas dapat diatasi dengan menggunakan beberapa metode regresi, yaitu *ridge regression*, *principal component analysis* (PCA), *partial least squares* (PLS), dan *least absolute shrinkage and selection operator* (LASSO). Dari keempat metode tersebut, metode LASSO tergolong baru dalam regresi dan pertama kali diperkenalkan oleh Tibshirani (1996). *Least absolute shrinkage and selection operator* merupakan teknik regresi yang melakukan penaksiran dengan meminumkan jumlah kuadrat error dengan suatu kendala. Karena kendala tersebut maka LASSO mengecilkan sejumlah koefisien menuju nol dan pada waktu yang sama melakukan seleksi variabel (Tibshirani (1996), dalam Widjaja (2015)).

Nurhayati (2014) dalam penelitiannya menyatakan LASSO menghasilkan model terbaik dibanding *principal component analysis* dan *partial least squares* berdasarkan nilai *root mean square error*. Selain itu, Pusporini (2012) menyatakan LASSO lebih baik digunakan dibanding *ridge regression* karena model yang dihasilkan lebih sederhana dan pengecilan koefisien tepat nol. Pada penelitian Widjaja (2015) menyarankan untuk mengestimasi koefisien pada regresi logistik ordinal dan multinomial dengan metode LASSO.

Berdasarkan uraian diatas, maka dalam penelitian ini akan membahas mengenai “**Metode *Least Absolute Shrinkage and Selection Operator* untuk Mengatasi Multikolinearitas pada Regresi Logistik Ordinal**”.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang di atas, maka permasalahan yang dibahas dalam penelitian ini sebagai berikut

- 1.2.1 Bagaimana mengestimasi koefisien regresi logistik ordinal dengan metode *least absolute shrinkage and selection operator* ?
- 1.2.2 Bagaimana penerapan model regresi logistik ordinal dengan metode *least absolute shrinkage and selection operator* pada data ?



1.3 Batasan Masalah

Penelitian ini difokuskan hanya terbatas pada pembahasan model regresi logistik ordinal yang mengandung multikolinearitas pada data.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan penelitian ini sebagai berikut

- 1.4.1 Mendapatkan penduga koefisien regresi logistik ordinal dengan menggunakan metode *least absolute shrinkage and selection operator*.
- 1.4.2 Memperoleh model regresi logistik ordinal dengan menggunakan metode *least absolute shrinkage and selection operator* pada data.

1.5 Manfaat Penelitian

Adapun manfaat yang ingin dicapai dari penelitian ini adalah memperkaya dan memperluas pengetahuan tentang regresi logistik ordinal dalam mengatasi multikolinearitas.



BAB II

TINJAUAN PUSTAKA

2.1 Multikolinearitas

Analisis regresi yang melibatkan beberapa variabel prediktor memerlukan adanya pemenuhan asumsi yaitu tidak terjadinya multikolinearitas. Multikolinearitas terjadi ketika terdapat korelasi yang tinggi antara dua atau lebih variabel. *Variance Inflation Factor (VIF)* dapat digunakan sebagai kriteria untuk mendeteksi kasus multikolinieritas pada regresi linier yang memiliki lebih dari dua variabel prediktor. Nilai *VIF* untuk parameter regresi ke-*j* diformulasikan pada persamaan (Yan, 2009).

$$VIF_j = \frac{1}{1 - R_i^2} \quad i = 1, 2, \dots, n \quad (2.1)$$

dengan R_i^2 merupakan koefisien determinasi antara X_i dengan variabel prediktor lainnya pada persamaan regresi. Apabila nilai $VIF > 10$, maka dapat diindikasikan bahwa terdapat kasus multikolinearitas yang serius. Hal-hal yang akan terjadi apabila kasus multikolinearitas tidak diatasi adalah variansi estimasi menjadi besar, interval kepercayaan menjadi lebar dikarenakan variansi dan standar error besar. Kemudian pengujian signifikansi secara parsial menjadi tidak signifikan. Serta koefisien determinasi (R^2) tinggi, tetapi sedikit variabel prediktor yang signifikan (Yan, 2009).

2.2 Regresi Logistik

Regresi logistik merupakan salah satu metode statistik nonparametrik untuk menguji hipotesis. Metode regresi logistik adalah metode matematika yang menggambarkan hubungan antara satu atau lebih variabel prediktor dengan satu variabel respon yang dikotomi yang variabelnya dianggap hanya mempunyai dua nilai yang mungkin yaitu 0 dan 1, dimana kondisi ini dapat diartikan sebagai sukses atau gagal pada analisis regresi logistik tunggal dan regresi logistik berganda.



la umumnya analisis regresi membentuk suatu persamaan untuk
iksiki variabel respon berdasarkan variabel prediktornya. Model regresi

logistik ganda adalah model regresi logistik dengan variabel prediktornya lebih dari satu variabel.

Model regresi logistik dapat dituliskan sebagai berikut (Hosmer dan Lemeshow, 2000) :

$$\pi(x) = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)} \quad (2.2)$$

dengan β_k = parameter-parameter regresi logistik
 x_k = pengamatan variabel prediktor

2.3 Regresi Logistik Ordinal

Regresi logistik ordinal merupakan salah satu analisis regresi yang digunakan untuk menganalisa hubungan antara variabel respon dengan variabel prediktor, dimana variabel respon bersifat *polikotomus* dengan skala ordinal. Model yang dapat digunakan untuk regresi ordinal adalah model logit. Model logit tersebut adalah *cumulative logit models*. Pada model logit ini bersifat ordinal dalam respon Y dituangkan dalam peluang kumulatif sehingga model peluang kumulatif yaitu kurang dari atau sama dengan kategori respon ke- j pada p variabel prediktor yang dinyatakan dalam vektor x adalah $P(Y \leq j|x)$, dengan peluang lebih besar dari kategori respon ke- j , $P(Y > j|x)$ (Hosmer dan Lemeshow, 2000).

Peluang kumulatif, $P(Y \leq j|x)$, didefinisikan sebagai berikut :

$$P(Y \leq j|x) = \pi(x) = \frac{\exp(\beta_{0j} + \sum_{k=1}^p \beta_k x_{ik})}{1 + \exp(\beta_{0j} + \sum_{k=1}^p \beta_k x_{ik})} \quad (2.3)$$

dengan $x_{ik} = (x_{i1}, x_{i2}, \dots, x_{ip})$ merupakan nilai pengamatan ke- i , dimana $i = (1, 2, \dots, n)$ dari setiap variabel p variabel prediktor.

Pendugaan parameter regresi dilakukan dengan cara menguraikannya menggunakan transformasi logit dari *git* ($Y \leq j|x_i$) :

$$\text{logit } P(Y \leq j|x_i) = \ln \left(\frac{P(Y \leq j|x_i)}{1 - P(Y \leq j|x_i)} \right) \quad (2.4)$$



$P(Y \leq j|x_i)$ adalah peluang kumulatif pada p variabel prediktor yang dinyatakan dalam vektor x_i , dimana $j = 1, 2, \dots, J - 1$ dan J adalah banyak kategori pada variabel respon.

Dengan mensubstitusikan persamaan (2.3) pada persamaan (2.4) maka didapatkan

$$\text{logit } P(Y \leq j|x_i) = \beta_{0j} + \sum_{k=1}^p \beta_k x_{ik} \quad (2.5)$$

dengan nilai β_k untuk setiap $k = (1, 2, \dots, p)$ merupakan faktor koefisien yang tidak diketahui yang bersesuaian dengan x_i . $P(Y \leq j|x_i)$ merupakan peluang kumulatif dari kejadian $Y \leq j$. Jika terdapat tiga kategori respon maka peluang kumulatif dari respon ke- j adalah :

$$P(Y \leq 0|x_i) = \frac{\exp(\beta_{00} + \sum_{k=1}^p \beta_k x_{ik})}{1 + \exp(\beta_{00} + \sum_{k=1}^p \beta_k x_{ik})}$$

$$P(Y \leq 1|x_i) = \frac{\exp(\beta_{01} + \sum_{k=1}^p \beta_k x_{ik})}{1 + \exp(\beta_{01} + \sum_{k=1}^p \beta_k x_{ik})} \quad (2.6)$$

Maka model regresi logistik yang terbentuk adalah jika terdapat tiga kategori respon adalah :

$$\text{logit } \pi_0(x_i) = \ln \left(\frac{P(Y \leq 0|x_i)}{1 - P(Y \leq 0|x_i)} \right) = \beta_{00} + \sum_{k=1}^p \beta_k x_{ik}$$

$$\text{logit } \pi_1(x_i) = \ln \left(\frac{P(Y \leq 1|x_i)}{1 - P(Y \leq 1|x_i)} \right) = \beta_{01} + \sum_{k=1}^p \beta_k x_{ik} \quad (2.7)$$

Berdasarkan kedua peluang kumulatif pada persamaan (2.6) maka didapatkan peluang untuk masing-masing kategori respon sebagai berikut :

$$P(Y = 0|x_i) = \pi_0(x_i) = P(Y \leq 0|x_i)$$

$$P(Y = 1|x_i) = \pi_1(x_i) = P(Y \leq 1|x_i) - P(Y \leq 0|x_i)$$

$$P(Y = 2|x_i) = \pi_2(x_i) = 1 - P(Y \leq 1|x_i)$$



Nilai $\pi_j(x_i)$ pada persamaan (2.7) akan dijadikan pedoman dalam proses pengklasifikasian. Suatu pengamatan masuk dalam respon kategori ke- j berdasarkan nilai $\pi_j(x_i)$ yang terbesar (Agresti, 1996).

2.4 *Least Absolute Shrinkage and Selection Operator*

Metode *Least Absolute Shrinkage and Selection Operator (LASSO)* diperkenalkan pertama kali oleh Tibshirani pada tahun 1996. *Least absolute shrinkage and selection operator* mengecilkan koefisien regresi (β_k) yang berkorelasi menjadi nol atau mendekati nol sehingga menghasilkan penduga dengan varian yang lebih kecil dan model akhir yang lebih representatif.

Misalkan $(X, y_i), i = 1, 2, \dots, n$ dimana $X = x_{i1}, x_{i2}, \dots, x_{ip}$ adalah variabel prediktor, y_i adalah variabel respon dan n banyaknya pengamatan. Pada regresi biasa, diasumsikan semua observasi adalah variabel prediktor atau y_i adalah dengan syarat x_{ij} . Asumsikan bahwa x_{ij} terstandarisasi sehingga $\sum_i x_{ij}/n = 0$ dan $\sum_i x_{ij}^2/n = 1$. Penduga parameter pada metode *LASSO* adalah sebagai berikut :

$$\hat{\beta}^{lasso} = \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \sum_{k=1}^p x_{ik} \hat{\beta}_k \right)^2 \quad (2.8)$$

dengan kendala $\sum_{k=1}^p |\hat{\beta}_k| \leq t$.

Dalam bentuk persamaan *Lagrange*, dugaan koefisien dengan menggunakan *LASSO* menjadi :

$$\hat{\beta}^{lasso} = \sum_{i=1}^N \left(y_i - \hat{\beta}_0 - \sum_{k=1}^p x_{ik} \hat{\beta}_k \right)^2 + \lambda \sum_{k=1}^p |\hat{\beta}_k| \quad (2.9)$$

λ adalah parameter yang mengontrol koefisien *LASSO* yang diatur dengan kendala $\sum_{k=1}^p |\hat{\beta}_k| \leq t$. Nilai t merupakan suatu besaran yang mengontrol besarnya pengecilan pada pendugaan koefisien *LASSO* dengan $t > 0$ (Hesti *et al.* 2008).

variabel yang berpengaruh besar dalam model terpilih dan variabel yang berpengaruh kecil terseleksi menjadi nol.



2.5 Algoritma *Least Angle Regression*

Least Angle Regression (LAR) merupakan suatu metode regresi yang algoritmanya dapat dimodifikasi menjadi algoritma komputasi untuk metode *LASSO*. Modifikasi dari *LAR* untuk *LASSO* menghasilkan efisiensi algoritma dalam menduga koefisien *LASSO* dengan komputasi yang lebih cepat dibandingkan pemrograman kuadrat (Wahdaniah, 2015).

Algoritma *LAR* adalah sebagai berikut (Hesti *et al.* 2008):

1. Membakukan variabel prediktor sehingga memiliki nilai tengah nol dan ragam satu. Semua nilai awal diberikan nilai nol. Pembakuan ini dimaksudkan agar dapat membandingkan dugaan koefisien regresi yang memiliki ragam yang berbeda dalam suatu model. Mulai dengan sisaan $r = y - \bar{y}$, dan $\beta_1, \beta_2, \dots, \beta_p = 0$.
2. Mencari variabel prediktor x_j yang paling berkorelasi dengan r .
3. Mengubah nilai β_j dari 0 ke arah koefisien kuadrat terkecil r dengan x_j sampai variabel prediktor yang lain x_k memiliki korelasi yang sama dengan galat akibat x_j .
4. Mengubah nilai β_j dan β_k ke arah yang ditentukan oleh koefisien kuadrat terkecil r dengan (x_j, x_k) sampai variabel prediktor yang lain x_l memiliki korelasi yang sama dengan galat akibat (x_j, x_k) .
5. Modifikasi algoritma *LAR* untuk mendapatkan solusi *LASSO* adalah dengan memodifikasi langkah ke-4 yaitu dengan cara mengeluarkan koefisien yang bukan nol dari gugus variabel aktif dan hitung kembali.
6. Mengulang langkah nomor 4 sampai semua p variabel prediktor dimasukkan. Setelah min $(N - 1, p)$ langkah, solusi model penuh untuk kuadrat terkecil diperoleh.

LAR selalu mengambil p langkah untuk mendapatkan penduga kuadrat terkecil secara penuh, sedangkan modifikasi *LAR* untuk *LASSO* dapat memiliki lebih dari p langkah untuk mendapatkannya. Algoritma *LASSO* dengan modifikasi *LAR* merupakan cara yang efisien dalam komputasi solusi masalah regresi terutama ketika jumlah variabel prediktor yang digunakan jauh lebih banyak daripada data amatannya (Hesti *et al.* 2008).



Dalam mencari model terbaik dalam metode *LASSO* dengan menggunakan kriteria validasi silang, yaitu dengan menggunakan mode *fraction* dan mode *step*.

2.6 Validasi Silang

Validasi silang membagi data menjadi dua bagian, yaitu data *training* dan data *test*. Data *training* digunakan untuk menentukan nilai $\hat{\beta}$ atau untuk menyusun model, sedangkan data *test* digunakan untuk menguji kebaikan prediksi $X\hat{\beta}$. Nilai validasi silang yang diperoleh merupakan penduga bagi galat prediksi (Izenman, 2008).

Salah satu metode tipe validasi silang adalah *k-fold*. Metode ini memiliki kelebihan ketika jumlah data amatan yang digunakan sedikit. Nilai galat prediksi *k-fold CV* apabila data dipartisi menjadi *c* bagian diperoleh dengan persamaan berikut:

$$CV = \frac{1}{c} \sum_{h=1}^c \sum_{(x_i, y_i) \in S} (y_i - \hat{y}_{-c}(x_i))^2 \quad (2.10)$$

dengan $\hat{y}_{-c}(x_i)$ adalah dugaan *y* dan *x_i* pada data *test S* pada saat *fold* ke-*c* tidak digunakan dalam menduga model dan *y_i* adalah variabel respon ke-*i* pada data *test S*. Menurut Izenman (2008), validasi silang yang sebaiknya digunakan adalah validasi silang *5-fold* atau *10-fold* karena menghasilkan nilai CV dengan bias tinggi tetapi ragam rendah.

2.7 Gizi Buruk

Menurut Standar WHO 2005 status gizi balita dinilai berdasarkan parameter antropometri yang terdiri dari berat badan dan panjang/tinggi badan. Status gizi balita merupakan salah satu indikator yang menggambarkan tingkat kesejahteraan masyarakat. Salah satu cara yang diukur melalui indeks Berat Badan menurut umur (BB/U) atau berat badan terhadap tinggi badan (BB/TB).

Menurut *UNICEF* (1998), penyebab tidak langsung dari gizi buruk adalah pola makan yang diwakili oleh pemberian ASI Eksklusif bayi, pola asuh yang diwakili oleh rumah tangga berperilaku hidup bersih dan sehat, pelayanan kesehatan yang diwakili oleh lima imunisasi dasar, serta sanitasi dan air bersih.



Menurut hasil Riskesdas tahun 2013 prevalensi gizi buruk kurang pada anak balita sebesar 25,6 persen, yang berarti masalah gizi berat-kurang di Sulawesi Selatan masih merupakan masalah kesehatan masyarakat dengan prevalensi tinggi.

