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,	Pembimbing Pertama,
Anisa, S.Si., M.Si.	**Drs. M. Saleh AF, M.Si.**
NIP. 19730227 199802 2 001	**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.	Sekertaris	: 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

بسم الله الرحمن الرحيم

Alhamdulillaahi Rabbil Aalamiin, tiada kata paling indah yang dipanjatkan oleh penulis melainkan rasa syukur kehadirat 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. **Ibu 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 Sahriman, 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**

emua pihak yang telah banyak berpartisipasi, baik secara langsung aupun tidak langsung dalam penyusunan tugas akhir ini yang tidak mpat 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 ekslusif, 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

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

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 [illegible] 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 Tribshirani (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 penelitannya 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 permasalah 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).

$$VIF_j = \frac{1}{1 - R_i^2} \qquad i = 1, 2, \dots, n \tag{2.1}$$

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 iksi 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) :

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

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 :

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

dengan $x_{ik} = \left(x_{i1}, x_{i2}, \dots x_{ip}\right)$ 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 $git\ (Y \leq j|x_i)$:

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

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

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

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

dengan nilai β_k untuk setiap $k = (1, 2, \ldots, p)$ 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 :

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

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

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

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

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

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, diasumsiskan 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 sebegai berikut :

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

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

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

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

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 ruh 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 kuadratik (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 $\beta_1, \beta_2, \ldots, \beta_p = 0$.
2. Mencari variabel prediktor x_j yang paling berkorelasi dengan r.
3. Mengubah nilai β_j dari 0 kearah 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 kearah 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 ikasi *LAR* merupakan cara yang efisien dalam komputasi solusi masalah terutama ketika jumlah variabel prediktor yang digunakan jauh lebih ari pada 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 *stap*.

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:

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

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 ...n pangan yang diwakili oleh pemberian ASI Eksklusif bayi, pola asuh ...akili oleh rumah tangga berperilaku hidup bersih dan sehat, pelayanan ... dasar 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.

BAB III

METODE PENELITIAN

3.1 Jenis dan Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder yaitu data bayi gizi buruk di provinsi Sulawesi Selatan tahun 2015 yang diperoleh dari publikasi Dinas Kesehatan Prov. Sulawesi Selatan (https://dinkes.sulselprov.go.id).

3.2 Identifikasi Variabel

Adapun variabel yang digunakan pada peneltian ini terdiri dari satu varibel respon dan enam variabel prediktor. Variabel respon (Y) adalah Jumlah bayi gizi buruk di Sulawesi Selatan tahun 2015, dimana dibagi menjadi tiga kategori :

0 = Rendah (0 - 4)

1 = Sedang (5 - 8)

2 = Tinggi (y ≥ 9)

Variabel prediktor yang digunakan berdasarkan data bayi gizi buruk adalah sebagai berikut

(X_1) = Jumlah bayi (0-6 Bulan) yang diberi ASI Eksklusif

(X_2) = Jumlah bayi yang mendapat Imunisasi Hepatitis B

(X_3) = Jumlah bayi yang mendapat Imunisasi BCG

(X_4) = Jumlah bayi yang mendapat Imunisasi Polio

(X_5) = Jumlah bayi yang mendapat Imunisasi Campak

(X_6) = Jumlah rumah dengan sanitasi layak

3.3 Metode Analisis

Metode analisis yang digunakan dalam peneletian ini adalah regresi logistik ordinal dengan metode *Least Absolut Shrinkage and Selection Operator*, dengan langkah-langkah sebagai berikut :

1. Menaksir koefisien regresi logistik ordinal.
   a. Menentukan fungsi *log-likelihood* untuk vektor β.
   b. Menambahkan penalty Lagrangian pada fungsi *log-likelihood* untuk vektor β yang disebut *log-likelihood.*

c. Meminimumkan fungsi *log-likelihood* dengan menurunkan fungsi tersebut terhadap parameternya dan disamakan dengan nol.

2. Penerapan metode *Least Absolut Shrinkage and Selection Operator* pada data bayi gizi buruk provinsi Sulawesi Selatan tahun 2015.

   a. Mendetekti ada tidaknya multikolinearitas pada data dengan melihat nilai VIF menggunakan SPSS. Jika data mengandung multikolinearitas maka analisis dapat dilanjut pada metode *LASSO* menggunakan LAR.

   b. Menghitung $\hat{\beta}$ dengan algoritma LAR yaitu :

      i. Membakukan variabel prediktor sehingga memiliki nilai tengah nol dan ragam satu. Mulai dengan sisaan $r = y - \bar{y}$, dan $\beta_1, \beta_2, \ldots, \beta_p = 0$.

      ii. Mencari variabel prediktor x_j yang paling berkorelasi dengan r.

      iii. Mengubah nilai β_j dari 0 kearah koefisien kuadrat terkecil r dengan x_j sampai variabel prediktor yang lain x_k memiliki korelasi yang sama dengan galat akibat x_j.

      iv. Mengubah nilai β_j dan β_k kearah 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).

      v. Mengeluarkan koefisien yang bukan nol dari gugus variabel aktif dan hitung kembali.

      vi. Mengulang langkah nomor 4 sampai semua p variabel prediktor dimasukkan. Setelah min $(N-1, p)$ langkah, solusi model penuh untuk kuadrat terkecil diperoleh.

   c. Membandingkan penduga koefisien regresi logistik ordinal dengan *LASSO*.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Estimasi Koefisien Regresi Logistik Ordinal dengan Metode *Least Absolute Shrinkage and Selection Operator*

Model regresi logistik ordinal :

$$P(Y \leq j|x_i) = \pi(x) = \frac{exp(\beta_{0j}+\sum_{k=1}^{P}\beta_k x_{ik})}{1+exp(\beta_{0j}+\sum_{k=1}^{P}\beta_k x_{ik})} \quad (4.1)$$

dimana:

j = kategori respon (3 kategori) ; $j = 0, 1, 2$

k = jumlah variabel prediktor

x_i = nilai pengamatan ke-i ; $i = 1, 2, \ldots, n$

Estimasi parameter pada model regresi logistik ordinal didapatkan dengan melakukan turunan parsial fungsi *log-likelihood* yang diestimasi dengan mengikuti distribusi multinomial.

Bentuk fungsi *likelihood* dengan tiga kategori :

$$L(\beta) = \prod_{i=1}^{n}\{[\pi_0(x_i)]^{y_{0i}}\,[\pi_1(x_i)]^{y_{1i}}\,[\pi_2(x_i)]^{y_{2i}}\} \quad (4.2)$$

dan fungsi *log-likelihood* pada persamaan (4.2) yaitu :

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\ln[\pi_0(x_i)] + y_{1i}\ln[\pi_1(x_i)] + y_{2i}\ln[\pi_2(x_i)]\} \quad (4.3)$$

peluang kumulatif pada persamaan (2.6) dimasukkan maka bentuk dari fungsi *log-likelihood* menjadi :

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\ln[\pi_0(x_i)] + y_{1i}\ln[\pi_1(x_i)] + y_{2i}\ln[\pi_2(x_i)]\}$$

$$l(\beta) = \sum_{i=1}^{n}\left\{y_{0i}\ln\left[\frac{\exp(\beta_{00}+\sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(\beta_{00}+\sum_{k=1}^{p}\beta_k x_{ik})}\right] + \right.$$

$$y_{1i}\ln\left[\frac{\exp(\beta_{01}+\sum_{k=1}^{p}\beta_k x_{ik}) - \exp(\beta_{00}+\sum_{k=1}^{p}\beta_k x_{ik})}{(1+\exp(\beta_{01}+\sum_{k=1}^{p}\beta_k x_{ik}))\,(1+\exp(\beta_{00}+\sum_{k=1}^{p}\beta_k x_{ik}))}\right] +$$

$$\left. y_{2i}\ln\left[\frac{1}{1+\exp(\beta_{00}+\sum_{k=1}^{p}\beta_k x_{ik})}\right]\right\}$$

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - y_{0i}\left(\ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) + y_{1i}\left(\ln\left[exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right] - \ln\left[\left(1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right)\left(1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right)\right]\right) - y_{2i}\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right)\}$$

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - y_{0i}\left(\ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) + y_{1i}\left(\ln\left[exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - y_{1i}\left(\ln\left[\left(1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right)\left(1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right)\right]\right) - y_{2i}\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right)\}$$

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - y_{0i}\left(\ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) + y_{1i}\left(\ln\left[exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - y_{1i}\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right] + \ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - y_{2i}\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right)\}$$

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - y_{0i}\left(\ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) + y_{1i}\left(\ln\left[exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - y_{1i}\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - y_{1i}\left(\ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - y_{2i}\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right)\}$$

$$l(\beta) = \sum_{i=1}^{n}\{y_{0i}\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - (y_{0i} + y_{1i})\left(\ln\left[1 + exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) + y_{1i}\left(\ln\left[exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - exp\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right) - (y_{1i} + y_{2i})\left(\ln\left[1 + exp\left(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\right)\right]\right)\} \qquad (4.4)$$

dalam metode *LASSO* fungsi *log-likelihood* pada persamaan (4.4) dengan kendala $\sum_{k=1}^{p}|\beta_k| \leq t$, t merupakan merupakan suatu besaran yang mengontrol besarnya pengecilan pada pendugaan koefisien *LASSO* dimana $t \geq 0$. Kendala tersebut ekivalen dengan penambahan *Penalty Lagrange* $\frac{1}{|\beta_k|}\lambda \sum \beta_k^2$ pada *log-likelihood* dengan $\lambda \geq 0$ yang bergantung t.

Penalty Lagrange berasal dari distribusi laplace dengan mean 0. Fungsi kepadatan peluang variabel acak distribusi laplace dengan mean μ dan variansi $2/\lambda^2$

$$f(\beta_k) = \frac{\lambda}{2}\exp(-\lambda|\beta_k - \mu|) \tag{4.5}$$

Fungsi *likelihood* dari persamaan (4.5) adalah

$$\begin{aligned} L(\beta) &= \left[\frac{\lambda}{2}\exp(-\lambda|\beta_1|)\right]\left[\frac{\lambda}{2}\exp(-\lambda|\beta_2|)\right]\ldots\left[\frac{\lambda}{2}\exp\left(-\lambda\left|\beta_p\right|\right)\right] \\ &= \prod_{k=1}^{p}\frac{\lambda}{2}\exp(-\lambda|\beta_k|) \end{aligned} \tag{4.6}$$

dan fungsi *log-likelihood* dari persamaan (4.6) adalah

$$\begin{aligned} l(\beta) &= \ln\left[\prod_{k=1}^{p}\frac{\lambda}{2}\exp(-\lambda|\beta_k|)\right] \\ &= p\ln\lambda - p\ln 2 + \ln\left[\exp -\lambda\sum_{k=1}^{p}|\beta_k|\right] \\ &= p\ln\lambda - p\ln 2 - \lambda\sum_{k=1}^{p}|\beta_k| \end{aligned} \tag{4.7}$$

Fungsi *log-likelihood* pada persamaan (4.4) ditambahkan fungsi *Penalty Lagrange* pada persamaan (4.7) maka diperoleh sebagai berikut

$$\begin{aligned} l(\beta)_c = \sum_{i=1}^{n}\Big(& y_{0i}\left(\beta_{00} + \sum_{k=1}^{p}\beta_k x_{ik}\right) - (y_{0i} + y_{1i})\Big(\ln\Big[1 + exp\Big(\beta_{00} + \\ & \sum_{k=1}^{p}\beta_k x_{ik}\Big)\Big]\Big) + y_{1i}\Big(\ln\Big[exp\Big(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\Big) - exp\Big(\beta_{00} + \\ & \sum_{k=1}^{p}\beta_k x_{ik}\Big)\Big]\Big) - (y_{1i} + y_{2i})\Big(\ln\Big[1 + exp\Big(\beta_{01} + \sum_{k=1}^{p}\beta_k x_{ik}\Big)\Big]\Big)\Big) + \\ & p\ln\lambda - p\ln 2 - \lambda\sum_{k=1}^{p}|\beta_k| \end{aligned}$$

Untuk mendapatkan estimasi $\hat{\beta}$ diperoleh dengan

$$= \arg\min l(\beta)_c$$

$$= \arg\min \Big\{ \sum_{i=1}^{n} \Big(y_{0i} \left(\beta_{00} + \sum_{k=1}^{p} \beta_k x_{ik} \right) - (y_{0i} + y_{1i}) \big(\ln \big[1 + exp\big(\beta_{00} + \sum_{k=1}^{p} \beta_k x_{ik} \big) \big] \big) + y_{1i} \big(\ln \big[exp\big(\beta_{01} + \sum_{k=1}^{p} \beta_k x_{ik} \big) - exp\big(\beta_{00} + \sum_{k=1}^{p} \beta_k x_{ik} \big) \big] \big) - (y_{1i} + y_{2i}) \big(\ln \big[1 + exp\big(\beta_{01} + \sum_{k=1}^{p} \beta_k x_{ik} \big) \big] \big) \Big) + p \ln \lambda - p \ln 2 - \lambda \sum_{k=1}^{p} |\beta_k| \Big\} \quad (4.8)$$

selanjutnya pada persamaan (4.8) diturunkan terhadap parameter β dan hasil turunannya sama dengan nol untuk mencapai titik kritis seperti berikut

Turunan $l(\beta)_c$ terhadap β_{0j} $\quad ; j = 0,1$

$$\frac{\partial l(\beta)_c}{\partial \beta_{00}} = \sum_{i=1}^{n} \left(y_{0i} - (y_{0i} + y_{1i}) \left(\frac{exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{1+exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})} \right) - y_{1i} \left(\frac{exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{\left(exp(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})\right) - \left(exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})\right)} \right) \right) = 0$$

$$\frac{\partial l(\beta)_c}{\partial \beta_{01}} = \sum_{i=1}^{n} \left(y_{1i} \left(\frac{exp(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{\left(exp(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})\right) - \left(exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})\right)} \right) - (y_{1i} + y_{2i}) \left(\frac{exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{1+exp(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})} \right) \right) = 0$$

Turunan $l(\beta)_c$ terhadap β_k $\quad ; k = 1,2,\dots,p$

$$\frac{\partial l(\beta)_c}{\partial \beta_1} = \sum_{i=1}^{n} \left(y_{0i}(x_{i1}) + (y_{0i} + y_{1i}) \left(\frac{exp\ (\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\ (\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})} (x_{i1}) \right) + y_{1i} \left(\frac{exp\ (\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\ (\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})} (x_{i1}) - \left(\frac{exp\ (\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\ (\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})} (x_{i1}) \right) \right) (y_{1i} + y_{2i}) \left(\frac{exp\ (\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{1+exp\ (\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})} (x_{i1}) \right) \right) - \lambda \frac{|\beta_1|}{\beta_1} = 0$$

$$\frac{\partial l(\beta)_c}{\partial \beta_2} = \sum_{i=1}^{n}\left(y_{0i}(x_{i2}) + (y_{0i} + y_{1i})\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{i2})\right) + \right.$$

$$y_{1i}\left(\frac{exp\,(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{i2}) - \right.$$

$$\left.\left.\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{i2})\right)\right)(y_{1i} + \right.$$

$$\left. y_{2i})\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{1+exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{i2})\right)\right) - \lambda\frac{|\beta_2|}{\beta_2} = 0$$

$$\vdots$$

$$\frac{\partial l(\beta)_c}{\partial \beta_p} = \sum_{i=1}^{n}\left(y_{0i}(x_{ip}) + (y_{0i} + y_{1i})\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ip})\right) + \right.$$

$$y_{1i}\left(\frac{exp\,(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ip}) - \right.$$

$$\left.\left.\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ip})\right)\right)(y_{1i} + \right.$$

$$\left. y_{2i})\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{1+exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ip})\right)\right) - \lambda\frac{|\beta_p|}{\beta_p} = 0 \qquad (4.9)$$

Untuk turunan terhadap β_k $;k = 1,2,\dots,p$ diperoleh bentuk secara umum sebagai berikut :

$$\frac{\partial l(\beta)_c}{\partial \beta_k} = \sum_{i=1}^{n}\left(y_{0i}(x_{ik}) + (y_{0i} + y_{1i})\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ik})\right) + \right.$$

$$y_{1i}\left(\frac{exp\,(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{01}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ik}) - \right.$$

$$\left.\left.\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}(x_{ik})\right)\right)(y_{1i} + \right.$$

$$\left. y_{2i})\left(\frac{exp\,(\beta_{00}+\beta_1 x_{i1}+\cdots+\beta_p x_{ip})}{1+exp\,(\beta_{00}+\beta_1+\cdots+\beta_p x_{ip})}(x_{ik})\right)\right) - \lambda\frac{|\beta_k|}{\beta_k} = 0 \qquad (4.10)$$

Persamaan (4.10) dapat dilihat bahwa nilai turunan pertama dari fungsi *log likelihood* tidak memberikan penyelesaian karena parameter-parameternya masih saling terkait satu sama lain. Pada estimasi parameter-parameter ini akan dilakukan pendekatan lain untuk mendapatkan hasil estimasinya. Pendekatan yang digunakan yaitu menggunakan algoritma *Least Angle Regression (LAR).*

4.2 Penerapan Metode *Least Absolute Shrinkage and Selection Operator* pada Data Bayi Gizi Buruk di Provinsi Sulawesi Selatan Tahun 2015

Contoh penerapan model regresi logistik ordinal adalah data gizi buruk di provinsi Sulawesi Selatan yang bersumber dari Dinas Kesehatan Provinsi Sulawesi Selatan dengan memperhitungkan variabel prediktor jumlah bayi (0-6 Bulan) yang diberi ASI Eksklusif (X_1), jumlah bayi yang mendapat Imunisasi Hepatitis B (X_2), jumlah bayi yang mendapat Imunisasi BCG (X_3), jumlah bayi yang mendapat Imunisasi Polio (X_4), jumlah bayi yang mendapat Imunisasi Campa (X_5), jumlah rumah dengan sanitasi layak (X_6), dan jumlah rumah dengan air minum sesuai syarat kesehatan (X_7). Selengkapnya data yang digunakan ditunjukan pada Lampiran 1.

Pengujian multikolinearitas dilakukan dengan melihat nilai *Variance Inflation Factors* (VIF). Berdasarkan Lampiran 2, ringkasan hasil pengujian multikolinearitas variabel-variabel prediktor dapat dilihat pada tabel berikut :

Tabel 4.1. Nilai *Variance Inflation Factors* variabel prediktor

Variabel Prediktor	VIF	Keterangan
X_1	12,075	Multikolinearitas
X_2	133,279	Multikolinearitas
X_3	1943,334	Multikolinearitas
X_4	1401,609	Multikolinearitas
X_5	845,468	Multikolinearitas
X_6	15,351	Multikolinearitas

Sumber : Data diolah, 2018

bel 4.1. menjelaskan nilai VIF dari variabel (X_1), (X_2), (X_3), (X_4), (X_5),) lebih dari 10, sehingga dapat disimpulkan bahwa terdapat multikolinearitas pada variabel prediktor tersebut. Artinya variabel (X_1), (X_2), (X_3), (X_4), (X_5), dan (X_6) masing-masing telah mempengaruhi nilai dari variabel lain. Hal ini menunjukan bahwa metode *LASSO* digunakan pada data.

Pengujian awal dilakukan dengan menduga paramater β_k menggunakan regresi logistik ordinal. Dilihat pada Lampiran 3 maka estimasi parameter untuk setiap variabel pada data bayi gizi buruk sebagai berikut :

Tabel 4.2. Hasil estimasi parameter pada bayi gizi buruk provinsi Sulawesi Selatan dengan regresi logistik ordinal

Variabel Prediktor	Koefisien Estimasi	Standart Error	*p-value*
X_1	1,021e-03	7,294e-04	0,1615
X_2	-1,053e-03	1,005e-03	0,295
X_3	-9,999e-05	3,746e-03	0,9787
X_4	3,292e-03	3,310e-03	0,32
X_5	-3,342e-03	2,635e03	0,2047
X_6	7,9343-05	3,972e-05	0,0458

Sumber : Data diolah, 2018

Berdasarkan Tabel 4.2., diperoleh bahwa hanya ada satu variabel prediktor yang signifikan mempengaruhi variabel respon yaitu variabel jumlah rumah dengan sanitasi layak (X_6). Hal ini ditunjukan dengan nilai *p-value* X_6 = 0,0458 kurang dari α = 0,05. Selanjutnya data dianalisis menggunakan regresi logistik ordinal dengan metode *LASSO* menggunakan algoritma *LAR* melalui bantuan *software* R 3.3.2.

Pada metode *LASSO*, diawali dengan menetapkan semua koefisien dengan angka nol. Selanjutnya, variabel yang memiliki korelasi tinggi dengan galat akan masuk ke dalam model.

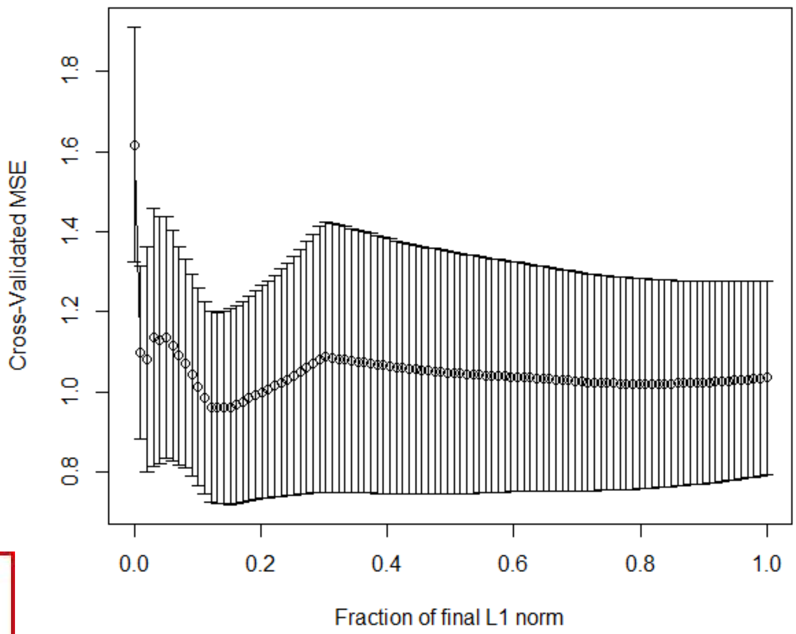
Variabel pertama yang masuk ke dalam model adalah X_2, artinya X_2 ...an variabel yang paling berkorelasi dengan sisaan. Variabel selanjutnya ...ng berkorelasi dengan sisaan setelah X_2 adalah X_6, sehingga X_6 adalah ...kedua yang masuk dalam model. Variabel yang terakhir masuk ke dalam model adalah X_3. Variabel-variabel yang masuk ke dalam model pada setiap tahapan dengan melihat nilai korelasi variabel prediktor terhadap sisaan dapat dilihat pada Tabel 4.3. Koefisien variabel yang masuk ke dalam model dapat dilihat pada Lampiran 4.

Tabel 4.3. Nilai korelasi variabel prediktor terhadap sisaan

Tahap	Variabel prediktor	Nilai korelasi
1	X_2	0,101
2	X_6	0,034
3	X_1	0,025
4	X_5	0,0097
5	X_4	0,0095
6	X_3	0,0094

Sumber : Data diolah, 2018

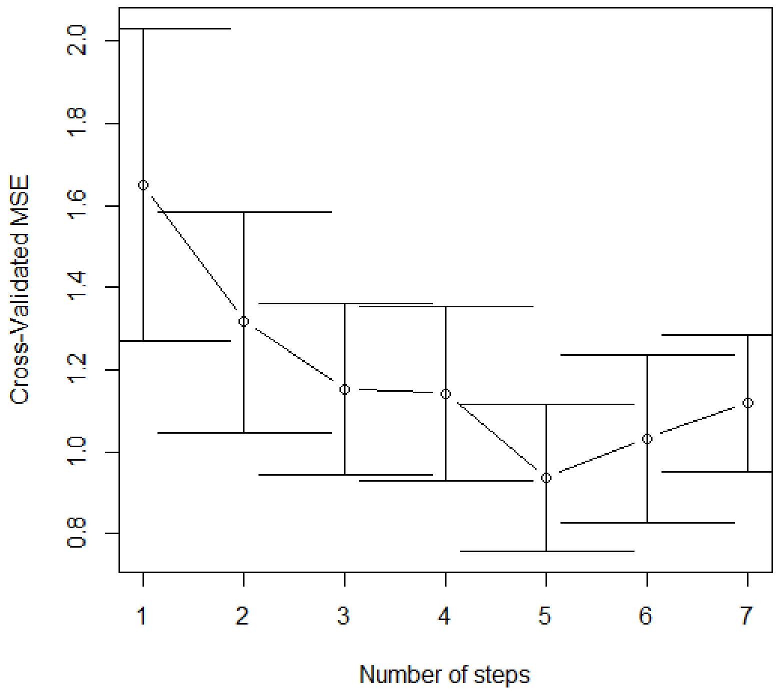
Pemilihan model terbaik dalam metode *LASSO* dilakukan dengan menggunakan kriteria validasi silang, yaitu dengan menggunakan mode *fraction* dan mode *step*. Pada mode *fraction*, nilai validasi silang dihitung berdasarkan $\sum|\hat{\beta}_j| \,/\, \max \sum|\hat{\beta}_j|$.



mbar 4.1. Nilai validasi silang dengan menggunakan mode *fraction*

Pada Gambar 4.1. terlihat bahwa *fraction* $\sum|\hat{\beta}_j| / \max\sum|\hat{\beta}_j|$ sekitar 0,175679 merupakan nilai validasi silang yang minimum. Nilai validasi silang minimum tersebut dapat berbeda setiap kali melakukan pemanggilan fungsinya. Pada Lampiran 4 terlihat bahwa nilai $\sum|\hat{\beta}_j| / \max\sum|\hat{\beta}_j|$ sekitar 0, 175679 berada antara tahap 5.

Mode *step* menghitung nilai validasi silang pada setiap tahapan variabel yang masuk dalam model. Pemilihan model terbaik dengan menggunakan mode *step* dalam menghitung nilai validasi silang menunjukkan bahwa model terbaik terlihat tahap 5 Gambar 4.2. Hal ini berdasarkan plot validasi silang yang merupakan nilai minimum.



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

Berdasarkan kedua metode validasi silang tersebut, maka model terbaik *LASSO* yang terpilih pada data ini adalah model pada tahap 5. Pada lampiran 5 empat variabel prediktor yang masuk ke dalam model, yaitu X_1, X_2, X_5, Adanya kendala pada metode *LASSO* menyebabkan nilai penduga koefisien parameter mengecil sehingga variabel prediktor yang penting atau berpengaruh terhadap model dimasukkan ke dalam model, sedangkan variabel prediktor yang kurang penting akan dikecilkan sampai nol dan terseleksi dari model sehingga model menjadi lebih efisien.

Tabel 4.4. Nilai koefisien regresi logistik ordinal dengan metode *LASSO*

Variabel Prediktor	Koefisien Regresi Logistik Ordinal	Koefisien Metode *LASSO*
X_1	0,001021	-0,000263
X_2	-0,001053	0,000298
X_3	-9.999e-05	0
X_4	0,003292	0
X_5	-0,003342	1,7698E-05
X_6	7.934E-05	-2,1729E-05

Sumber : Data diolah, 2018

Pada Tabel 4.4., terlihat bahwa metode *LASSO* cenderung dan dapat mengecilkan koefisien parameter persamaan regresi ordinal sampai tepat nol atau memiliki pengaruh lebih kecil terhadap variabel respon. Nilai penduga koefisien parameter untuk variabel X_3 dan X_4 dikecilkan sampai tepat nol, sehingga variabel-variabel tersebut tidak memiliki pengaruh atau kurang penting terhadap model. Melalui pengecilan yang sampai tepat nol ini, *LASSO* dapat pula digunakan sebagai metode seleksi variabel. Melalui seleksi variabel ini, model menjadi lebih sederhana dan efisien serta dapat mengatasi masalah multikolinearitas.

Persamaaan regresi logistik ordinal yang dihasilkan dengan menggunakan metode *LASSO* :

$$\pi(X_i) = \frac{\exp\left((-0{,}000263)\,X_1 + 0{,}000298\,X_2 + 0{,}000018\,X_5 + (-0{,}000022)X_6\right)}{1+\exp\left((-0{,}000263)\,X_1 + 0{,}000298\,X_2 + 0{,}000018\,X_5 + (-0{,}000022)X_6\right)} \quad (4.11)$$

rdasarkan Persamaan (4.11) maka dapat dijelaskan jika bayi yang ASI eksklusif bertambah satu, maka terjadi penurunan jumlah bayi gizi besar 0,4999. Jika seorang bayi mendapatkan imunisasi hepatitis B bertambahnya satu, maka terjadi penurunan jumlah bayi gizi buruk sebesar 0,50007. Jika seorang bayi mendapatkan imunisasi campak bertambahnya satu, maka terjadi penurunan jumlah bayi gizi buruk sebesar 0,5. Jika rumah dengan sanitasi yang layak bertambahnya satu, maka terjadi penurunan jumlah bayi gizi buruk sebesar 0,499995.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan uraian pembahasan mengenai estimasi koefisien model regresi logistik ordinal dengan metode *Least Absolute Shrinkage and Selection Operator*, maka dapat diambil kesimpulan sebagai berikut :

1. Taksiran model terbaik data bayi gizi buruk kabupaten/kota di Provinsi Sulawesi Selatan melalui pendekatan regresi logistik ordinal dengan metode *LASSO* adalah sebagai berikut

$$\pi(X_i) = \frac{\exp\big((-0{,}000263)\,X_1 + 0{,}000298\,X_2 + 0{,}000018\,X_5 + (-0{,}000022)X_6\big)}{1 + \exp\big((-0{,}000263)\,X_1 + 0{,}000298\,X_2 + 0{,}000018\,X_5 + (-0{,}000022)X_6\big)}$$

2. Berdasarkan penerapan model dengan metode *LASSO* menggunakan algoritma *LAR* pada data bayi gizi buruk kabupaten/kota di Provinsi Sulawesi Selatan terdapat empat variabel yang mempengaruhi gizi buruk yaitu pemberian ASI ekslusif, imunisasi hepatitis B, imunisasi campak, dan rumah dengan sanitasi layak.

5.2 Saran

Penulis menyarankan untuk mengestimasi regresi dengan metode *Least Absolute Shrinkage and Selection Operator* pada model *GLM* (*Genaralized Linear Model*) yang lain seperti regresi poisson.

DAFTAR PUSTAKA

[Dinkes] Dinas Kesehatan. 2014. *Profil Kesehatan Prov. Sulawesi Selatan Tahun 2015*. Makassar: Dinas Kesehatan Provinsi Sulawesi Selatan.

[Dinkes] Dinas Kesehatan. 2015. *Profil Kesehatan Prov. Sulawesi Selatan Tahun 2016*. Makassar: Dinas Kesehatan Provinsi Sulawesi Selatan. (https://dinkes.sulselprov.go.id/file/publik/Data%20ProfilL%202015.pdf), diakses 4 Januari 2018

Agresti, A., (1990), *Categorical Data Analysis*, John Wiley & Sons, Inc., New York

Agresti, A., 2007, *An Introduction To Categorical Data Analysis*., John Willey & Sons, Inc. New York.

Hastie T, Tibshirani R, Friedman J. 2008. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Ed ke-2. New York: Springer.

Hosmer, D.W. dan Lemeshow, S. 2000. *Applied Logistic Regression*. John Wiley & Sons, Inc. USA.

Izenman AJ. 2008. *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*. New York: Springer.

Nurhayati, Muhammad NA, Agus MS. 2014. *Metode Regresi Komponen Utama, Regresi Kuadrat Terkecil Parsial Dan Lasso Pada Data Kemiskinan Hasil Olahan Susenas 2012*. Jurnal Statistik. IPB.

Pusporini Arum, Aunuddin, La Ode AR. 2012. *Penerapan Regresi Gulud dan Least Absolute Shrinkage And Selection Operator (LASSO) dalam penyusutan Koefisien Regresi*. Jurnal Statistik. IPB.

Tibshirani R. 1996. *Regression Shrinkage and Selection via the Lasso*. *Journal of the Royal Statistical Society Series B* 58(1): 267-288.

Imelda. 2015. Skripsi. *Estimasi Koefisien Regresi Logistik Biner dengan tode Least Absolute Shrinkage and Selection Operator*. FMIPA iversitas Hasanuddin: Makassar.

Wahdaniah. 2015. Skripsi. *Penerapan Algoritma Least Angle Regression pada Least Absorute Shrinkage and Selection Operator untuk Model Cox.* FMIPA Universitas Hasanuddin: Makassar.

X. Yan, dan X. G. Su. 2009. *Linear Regression Analysis* : *Theory and Computing*. Singapore: World Scientific.

Lampiran 1. Data Bayi Gizi Buruk di Provinsi Sulawesi Selatan Tahun 2015

Kabupaten/Kota	Y	X1	X2	X3	X4	X5	X6
Selayar	1	712	2088	1970	1922	1945	15712
Bulukumba	0	1685	6422	6640	6506	6379	71063
Bantaeng	0	976	3286	3373	3182	3464	33002
Jeneponto	1	1357	6041	5847	5865	5402	39295
Takalar	2	3052	6430	5571	5140	5465	51445
Gowa	0	3034	12124	13206	13732	13676	109331
Sinjai	0	1584	4020	4297	4117	4138	35750
Maros	2	1782	4836	6592	6994	6587	23165
Pangkep	0	1847	5585	5701	5842	5616	27543
Barru	1	1579	2948	3188	3265	3321	18381
Bone	2	5822	13183	13540	13204	13355	116303
Soppeng	2	1054	3090	3195	3120	3125	43343
Wajo	2	1944	6889	7253	7124	7041	71171
Sidrap	0	2347	4563	5078	4843	4884	23599
Pinrang	1	3568	8491	7893	7725	7532	52383
Enrekang	2	2236	2955	3611	3707	3579	31600
Luwu	1	1915	7139	7011	6886	6690	40418
Tana Toraja	1	1073	4210	4276	4275	4102	21512
Luwu Utara	0	3914	4579	5494	5064	5082	1925
Luwu Timur	0	1929	5527	5733	5623	5811	37229
Toraja Utara	2	1829	3219	4301	4234	4617	24967
Makassar	1	10723	24969	25229	25376	25221	228041
Pare-pare	2	2024	2556	2967	2635	2486	21343
Palopo	2	683	2772	3051	3078	3020	24950

Lampiran 2. Uji Multikolinearitas dengan SPSS 23

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1,412	,418		3,376	,004		
	Asi_X1	,000	,000	1,038	1,396	,181	,083	12,075
	HepatitisB_X2	,000	,000	-2,126	-,860	,402	,008	133,279
	BCG_X3	,000	,002	-2,013	-,213	,834	,001	1943,334
	Polio_X4	,001	,001	7,239	,904	,379	,001	1401,609
	Campak_X5	-,001	,001	-5,651	-,908	,376	,001	845,468
	Sanitasi_X6	2,921E-5	,000	1,602	1,911	,073	,065	15,351

a. Dependent Variable: GiziBuruk_Y

Lampiran 3. Sintaks Program dan Output Regresi Logistik Ordinal dengan *software* R 3.3.2

```
> library(foreign)
> library(ordinal)
> data=file.choose("Gizi Buruk 2015.sav")
> gb=read.spss(data,to.data.frame=TRUE)
> giziFit<-
clm(GiziBuruk_Y~Asi_X1+HepatitisB_X2+BCG_X3+Polio_X4+Campak_X5+Sanitasi
_X6,link="logit",data=gb)
> summary(giziFit)
formula:
GiziBuruk_Y ~ Asi_X1 + HepatitisB_X2 + BCG_X3 + Polio_X4 + Campak_X5 +
Sanitasi_X6
data:    gb

 link   threshold  nobs  logLik   AIC    niter   max.grad   cond.H
 logit  flexible   24    -23.18   62.36  4(0)    8.88e-08   6.5e+10

Coefficients:
                Estimate    Std. Error  z value  Pr(>|z|)
Asi_X1          1.021e-03   7.294e-04   1.400    0.1615
HepatitisB_X2  -1.053e-03   1.005e-03  -1.047    0.2950
BCG_X3         -9.999e-05   3.746e-03  -0.027    0.9787
Polio_X4        3.292e-03   3.310e-03   0.994    0.3200
Campak_X5      -3.342e-03   2.635e-03  -1.268    0.2047
Sanitasi_X6     7.934e-05   3.972e-05   1.997    0.0458 *

Threshold coefficients:
                               Estimate  Std. Error  z value
Tinggi(9-50)|Sedang(5-8)       -1.8929   1.0136      -1.867
Sedang(5-8)|Rendah(0-4)        -0.4342   0.9516      -0.456
```

Lampiran 4. Sintaks Program dan Output dengan metode *LASSO software* R 3.3.2

```
> library(lars)
> data<-read.table("D:/Gizi Buruk 2015.txt",header=TRUE)
> y=as.matrix(data[,1])
> x=as.matrix(data[,2:7])
> lasso=lars(x,y,type="lasso")
> lasso
Call:
lars(x = x, y = y, type = "lasso")
R-squared: 0.222
Sequence of LASSO moves:
     HepatitisB_X2 Sanitasi_X6 Asi_X1 Campak_X5 Polio_X4 BCG_X3
Var              2           6      1         5        4      3
Step             1           2      3         4        5      6
> coef(lasso)
         Asi_X1 HepatitisB_X2       BCG_X3      Polio_X4    Campak_X5
[1,]  0.0000000000  0.000000e+00 0.0000000000  0.0000000000 0.0000000000
[2,]  0.0000000000  3.923205e-06 0.0000000000  0.0000000000 0.0000000000
[3,]  0.0000000000  1.531111e-05 0.0000000000  0.0000000000 0.0000000000
[4,] -0.0001480298  1.842700e-04 0.0000000000  0.0000000000 0.0000000000
[5,] -0.0002625332  2.975676e-04 0.0000000000  0.0000000000 0.0000176982
[6,] -0.0003591128  4.031246e-04 0.0000000000 -0.0008193752 0.0008306234
[7,] -0.0004241784  3.733822e-04 0.0003525302 -0.0012509336 0.0009823889
       Sanitasi_X6
[1,]  0.000000e+00
[2,]  0.000000e+00
[3,] -1.182236e-06
[4,] -1.274689e-05
[5,] -2.172876e-05
[6,] -2.776849e-05
[7,] -2.921148e-05
```

Lampiran 4. Nilai untuk Setiap Tahapan Metode *LASSO*

Tahapan	$\lvert X_1 \rvert$	$\lvert X_2 \rvert$	$\lvert X_3 \rvert$	$\lvert X_4 \rvert$	$\lvert X_5 \rvert$
1	0	0	0	0	0
2	0	3,9232E-06	0	0	0
3	0	1,5311E-05	0	0	0
4	0,000148	0,000184	0	0	0
5	0,000263	0,000298	0	0	1,7698E-05
6	0,000359	0,000403	0	0,000819	0,000831
7	0,000424	0,000373	0,000353	0,001251	0,000982

Tahapan	$\lvert X_6 \rvert$	$\sum\lvert\hat{\beta}_j\rvert$	$\sum\lvert\hat{\beta}_j\rvert \,/\, max \sum\lvert\hat{\beta}_j\rvert$
1	0	0	0
2	0	3,9232E-06	0,00115
3	1,1822E-06	1,6493E-05	0,004833
4	1,2747E-05	0,000345	0,101109
5	2,1729E-05	0,0006	0,175679
6	2,7769E-05	0,00244	0,714993
7	2,9212E-05	0,003413	1