

---

## Penerapan Metode *Linearized Ridge Regression* pada Data yang Mengandung Multikolinearitas

Mukrimin Adam<sup>1</sup>, Sitti Sahriman<sup>2</sup>, Nasrah Sirajang<sup>3\*</sup>

<sup>1,2,3</sup>Departemen Statistika, Fakultas MIPA, Universitas Hasanudin, Kota Makassar, 90245, Indonesia

\*Corresponding author, email: [adammukrimin@gmail.com](mailto:adammukrimin@gmail.com)

### Abstract

One of the assumptions that must be met in the multiple linear regression model is that there is no multicollinearity problem among the independent variables. However, if there is a multicollinearity problem, then parameter estimation can be done using the linearized ridge regression (LRR) method. The LRR method has the advantage of choosing an optimal constant that is easy to determine and also has a minimum PRESS value. In this study, the infant mortality rate in South Sulawesi Province will be modeled using the LRR method based on the variables of the amount of vitamin A given, the number of health services, the number of babies born with low weight, the number of mothers who give birth assisted by medical personnel, and the number of babies who are breastfed. exclusive. One measure to see the goodness of the regression model is the Prediction Error Sum of Squares (PRESS). Based on the t-test at a significance level of 5%, the total coverage of vitamin A administration and the number of babies born with low weight gave a significant effect on infant mortality with a PRESS value of 0.6846.

**Keywords:** Infant Mortality Rate, Multicollinearity, Linearized Ridge Regression, Prediction Error Sum of Squares.

### Abstrak

Salah satu asumsi yang harus dipenuhi dalam model regresi linier berganda adalah tidak terjadinya masalah multikolinearitas diantara variabel bebas. Namun jika terjadi masalah multikolinearitas, maka estimasi parameter dapat dilakukan dengan menggunakan metode linearized ridge regression (LRR). Metode LRR memiliki kelebihan pemilihan konstanta optimal yang mudah ditentukan dan juga memiliki nilai PRESS yang minimum. Pada penelitian ini, angka kematian bayi di Provinsi Sulawesi Selatan akan dimodelkan dengan menggunakan metode LRR berdasarkan variabel jumlah pemberian vitamin A, jumlah pelayanan kesehatan, jumlah bayi lahir dengan berat badan rendah, jumlah ibu bersalin yang ditolong tenaga medis, dan jumlah bayi yang diberi ASI eksklusif. Salah satu ukuran untuk melihat kebaikan model regresi adalah Prediction Error Sum of Squares (PRESS). Berdasarkan uji t pada tingkat signifikansi 5% diperoleh jumlah cakupan pemberian vitamin A dan jumlah bayi lahir dengan berat badan rendah memberikan pengaruh yang signifikan terhadap angka kematian bayi dengan nilai PRESS sebesar 0.6846.

**Kata Kunci:** Angka Kematian Bayi, Multikolinearitas, Linearized Ridge Regression, Prediction Error Sum of Squares.

## **1. Pendahuluan**

Aspek kesehatan merupakan salah satu aspek kualitas sumber daya manusia yang penting untuk diperhatikan di seluruh dunia dan juga sebagai pencapaian komitmen internasional yang tertuang pada tujuan Sustainable Development Goals (SDGs) [1]. Salah satu indikator penting untuk mengetahui derajat kesehatan suatu negara dan bahkan untuk mengukur tingkat kemajuan suatu bangsa adalah angka kematian bayi (AKB) [2]. Masih tingginya AKB menjadi salah satu masalah kesehatan utama di Indonesia.

Berdasarkan hasil Survei Demografi dan Kesehatan Indonesia (SDKI) tahun 2019 yang menunjukkan bahwa AKB sebesar 6.02 per 1000 kelahiran hidup telah mengalami penurunan dari tahun-tahun sebelumnya, seperti pada tahun 2007 yang masih mencapai 34 per 1000 kelahiran hidup dan tahun 2012 sebesar 32 per 1000 kelahiran hidup [2]. Walaupun AKB berdasarkan hasil SDKI beberapa kurun waktu telah menunjukkan penurunan, namun bila dibandingkan dengan dengan negara lain AKB di Indonesia masih tergolong tinggi.

Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis hubungan antara variabel terikat dan variabel bebas. Salah satu metode analisis regresi yang sering digunakan adalah metode kuadrat terkecil. Metode kuadrat terkecil akan dapat mengestimasi parameter dengan baik apabila antara variabel bebasnya tidak terjadi masalah multikolinearitas. Multikolinearitas adalah suatu keadaan yang mana variabel-variabel bebas dalam model regresi linier berganda saling bergantung linier [3]. Ketika variabel bebas saling bergantung linier, estimasi parameter dengan menggunakan metode kuadrat terkecil menjadi kurang tepat karena akan menghasilkan hasil estimasi dengan variansi yang besar sehingga estimasi dengan metode kuadrat terkecil menjadi tidak stabil [4].

Pada tahun 2011, Liu dan Gao memperkenalkan metode Linearized Ridge Regression (LRR) yang merupakan perbaikan dari metode-metode sebelumnya yang dapat mengatasi multikolinearitas. Dalam membentuk estimator LRR, Liu dan Gao menggabungkan estimator generalized ridge dengan estimator liu sehingga menghasilkan estimator yang memiliki kelebihan yaitu pemilihan konstanta optimal yang mudah ditentukan dan juga nilai Prediction Error Sum of Squares (PRESS) yang minimum [5]. Nilai PRESS merupakan salah satu ukuran yang dapat digunakan untuk melihat kebaikan dari model regresi [6]. Berdasarkan uraian tersebut, penulis akan melakukan penelitian dengan judul “Pemodelan *Linearized Ridge Regression* Pada Data yang Mengandung Multikolinearitas”.

## **2. Material dan Metode**

Data yang digunakan dalam penelitian ini merupakan data sekunder dari 24 kabupaten/kota di Provinsi Sulawesi Selatan mengenai faktor-faktor yang mempengaruhi Angka Kematian Bayi (AKB) di Provinsi Sulawesi Selatan tahun 2019. Data diambil dari

publikasi Dinas Kesehatan Provinsi Sulawesi Selatan pada situs <http://dinkes.sulselprov.go.id> berupa buku yang berjudul “Profil Kesehatan Provinsi Sulawesi Selatan Tahun 2020”. Variabel yang digunakan dalam penelitian ini terdiri dari satu variabel dependen yaitu angka kematian bayi ( $y$ ), serta lima variabel independen yaitu jumlah pemberian vitamin A ( $x_1$ ), jumlah pelayanan kesehatan bayi ( $x_2$ ), jumlah bayi lahir dengan berat badan rendah ( $x_3$ ), jumlah ibu bersalin ditolong tenaga medis ( $x_4$ ) dan jumlah bayi yang diberi ASI eksklusif ( $x_5$ ).

Data dianalisis dengan menggunakan metode *Linearized Ridge Regression* yang merupakan pengembangan dari metode *generalized ridge* dan metode Liu, yaitu:

$$\hat{\beta}_{LRR}^* = (\mathbf{Z}'\mathbf{Z} + \mathbf{I})^{-1}[(\mathbf{Z}'\mathbf{Z}) + \mathbf{QDQ}']\hat{\beta}^* \quad (1)$$

dengan,

- $\mathbf{Z}$  = Matriks dari variabel bebas untuk data yang telah distandarisasi ( $n \times p$ )
- $\mathbf{I}$  = Matriks identitas ( $p \times p$ )
- $\mathbf{Q}$  = Matriks ortogonal yang dibentuk oleh vektor eigen dari matriks  $\mathbf{Z}$
- $\mathbf{D}$  = Konstanta bias dari metode *Linearized Ridge Regression*
- $\hat{\beta}^*$  = Parameter yang diperoleh dengan menggunakan metode kuadrat terkecil

Penerapan model regresi linier berganda menggunakan metode *linearized ridge regression* pada data angka kematian bayi di Provinsi Sulawesi Selatan tahun 2019 dapat dilakukan dengan langkah-langkah sebagai berikut:

- a. Melakukan uji multikolinieritas pada data.
- b. Melakukan standarisasi pada data.
- c. Mengestimasi parameter model regresi linier berganda dengan menggunakan estimator *linearized ridge regression* yang telah diperoleh pada langkah sebelumnya.
- d. Membentuk model regresi linier berganda berdasarkan estimasi parameter yang telah diperoleh sebelumnya menggunakan estimator *linearized ridge regression*.
- e. Melakukan pengujian signifikansi parameter model regresi linier berganda dengan estimator *linearized ridge regression* secara simultan dan parsial.
- f. Mengidentifikasi variabel bebas yang berpengaruh terhadap variabel terikat.
- g. Menginterpretasi model yang telah diperoleh dan menarik kesimpulan.

### 3. Hasil dan Diskusi

#### 3.1 Identifikasi Multikolinieritas

Pengujian multikolinieritas menggunakan kriteria *Variance Inflation Factors* (VIF). Jika nilai  $VIF > 10$ , maka dapat disimpulkan bahwa terjadi masalah multikolinieritas pada data. Pendeteksian multikolinieritas menggunakan kriteria *Variance Inflation*

Factors (VIF). Nilai VIF yang lebih dari 10, disimpulkan terjadi multikolinearitas pada model.

Tabel 1 Nilai *Variance Inflation Factors* (VIF)

Variabel	VIF	Indikasi Multikolinearitas
$x_1$	1.396	Tidak Terjadi Multikolinearitas
$x_2$	320.063	Terjadi Multikolinearitas
$x_3$	7.332	Tidak Terjadi Multikolinearitas
$x_4$	356.896	Terjadi Multikolinearitas
$x_5$	4.671	Tidak Terjadi Multikolinearitas

Berdasarkan Tabel 1, dapat diketahui bahwa variabel jumlah pelayanan kesehatan ( $x_2$ ) dan jumlah ibu bersalin ditolong tenaga medis ( $x_4$ ) memiliki nilai VIF lebih dari 10 sehingga terdapat multikolinearitas atau korelasi antar variabel prediktor pada data.

### 3.2 Penerapan Metode *Linearized Ridge Regression* Pada Data

Pemodelan data akan menggunakan estimator *Linearized Ridge Regression* yang telah diperoleh pada persamaan (1). Dengan menggunakan persamaan tersebut, diperoleh hasil sebagai berikut:

$$\begin{aligned}
 (\mathbf{Z}'\mathbf{Z} + \mathbf{I})^{-1} &= \begin{bmatrix} 0.5508 & -0.0613 & -0.0418 & -0.0665 & -0.0471 \\ -0.0613 & 0.7777 & -0.1775 & -0.2188 & -0.1636 \\ -0.0418 & -0.1775 & 0.7197 & -0.1819 & -0.1157 \\ -0.0665 & -0.2188 & -0.1819 & 0.7814 & -0.1625 \\ -0.0471 & -0.1636 & -0.1157 & -0.1625 & 0.6999 \end{bmatrix} \\
 \mathbf{Q} &= \begin{bmatrix} -0.0103 & 0.0481 & 0.0057 & 0.9537 & 0.2967 \\ -0.6866 & -0.5126 & 0.0770 & -0.1350 & 0.4916 \\ -0.0331 & 0.6287 & 0.5944 & -0.1807 & 0.4666 \\ 0.7262 & -0.4530 & 0.0950 & -0.1234 & 0.4932 \\ -0.0045 & 0.3667 & -0.7948 & -0.1561 & 0.4576 \end{bmatrix} \\
 \mathbf{D} &= \begin{bmatrix} -0.5424 & 0 & 0 & 0 & 0 \\ 0 & -2.1684 & 0 & 0 & 0 \\ 0 & 0 & -2.7127 & 0 & 0 \\ 0 & 0 & 0 & 13.7807 & 0 \\ 0 & 0 & 0 & 0 & -1.0544 \end{bmatrix} \\
 \hat{\boldsymbol{\beta}}^* &= \begin{bmatrix} -0.2363 \\ 1.3785 \\ 0.6574 \\ -2.3022 \\ -0.0047 \end{bmatrix}
 \end{aligned}$$

sehingga diperoleh hasil estimasi parameter model regresi linier berganda dengan menggunakan estimator LRR pada Tabel 2.

**Tabel 2** Output Estimator LRR

$j$	$\hat{\beta}_j(LRR)$
1	-2.1058
2	-0.0779
3	-0.9690
4	1.8155
5	0.1937

Adapun model regresi linier berganda dengan menggunakan metode LRR sebagai berikut:

$$\hat{y}^* = -2.1058z_1 - 0.0779z_2 - 0.9690z_3 + 1.8155z_4 + 0.1937z_5$$

jika model tersebut dibawa ke dalam model variabel asal, maka diperoleh model regresi linier berganda dengan menggunakan estimator LRR sebagai berikut:

$$\hat{y} = 9.2580 - 0.0008x_1 - 0.00005x_2 - 0.0113x_3 + 0.0012x_4 + 0.0002x_5$$

### 3.3 Pengujian Signifikansi Parameter

Uji signifikansi digunakan untuk mengetahui pengaruh dari variabel bebas terhadap variabel terikat baik secara simultan maupun secara parsial.

#### 3.3.1 Uji F

Pengujian parameter dengan menggunakan uji F bertujuan untuk melihat pengaruh semua variabel bebas terhadap variabel terikat. Adapun hipotesis yang digunakan untuk uji F yaitu:

$$H_0: \beta_{LRR_1} = \beta_{LRR_2} = \dots = \beta_{LRR_5} = 0$$

$$H_1: \text{Ada } \beta_{LRR_j} \neq 0, j = 1, \dots, 5$$

Hasil perhitungan  $F_{hitung}$  ditunjukkan pada Tabel 3 dengan nilai sebesar 3.8626, nilai ini lebih besar dari  $F_{tabel}$  sebesar 2.77. Oleh karena itu, dapat disimpulkan bahwa  $H_0$  ditolak yang artinya jumlah pemberian vitamin A ( $x_1$ ), jumlah pelayanan kesehatan ( $x_2$ ), jumlah bayi lahir dengan berat badan rendah ( $x_3$ ), jumlah ibu bersalin yang ditolong tenaga kesehatan ( $x_4$ ) dan jumlah bayi yang diberi ASI eksklusif ( $x_5$ ) secara bersama-sama mempengaruhi angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019.

Tabel 2 Analisis Ragam Untuk Pengujian Signifikansi Parameter

Sumber	Derajat Bebas	Jumlah Kuadrat	Kuadrat Tengah	$F_{hitung}$
Regresi	5	1.9510	0.3902	
Sisaan	18	1.8184	0.1010	3.8626
Total	23	3.7693		

### 3.3.2 Uji Parsial

Uji t digunakan untuk menguji pengaruh setiap variabel bebas secara satu persatu terhadap variabel terikatnya. Adapun hipotesis yang digunakan untuk uji t yaitu:

$$H_0: \beta_{RRL_j} = 0, j = 1, 2, \dots, 5$$

$$H_1: \beta_{RRL_j} \neq 0$$

Hasil perhitungan nilai  $t_{hitung}$  ditunjukkan oleh Tabel 4, pada Tabel 4 dapat dilihat bahwa jumlah pemberian vitamin A ( $x_1$ ) dan jumlah bayi lahir dengan berat badan rendah ( $x_3$ ) mempunyai nilai  $|t_{hitung}|$  yang lebih besar daripada nilai  $t_{tabel}$  (2.1009) sehingga  $H_0$  ditolak. Sementara itu, untuk variabel jumlah pelayanan kesehatan ( $x_2$ ), jumlah ibu bersalin yang ditolong tenaga kesehatan ( $x_4$ ), dan jumlah bayi yang diberi ASI eksklusif ( $x_5$ ) mempunyai nilai  $|t_{hitung}|$  yang lebih kecil daripada nilai  $t_{tabel}$ , sehingga  $H_0$  diterima. Berdasarkan hal tersebut, maka dapat disimpulkan bahwa variabel jumlah pelayanan kesehatan ( $x_2$ ), jumlah ibu bersalin yang ditolong tenaga kesehatan ( $x_4$ ) dan jumlah bayi yang diberi ASI eksklusif ( $x_5$ ) memberikan pengaruh yang sangat kecil terhadap angka kematian bayi di Sulawesi Selatan pada tahun 2019 sedangkan variabel jumlah pemberian vitamin A ( $x_1$ ) dan jumlah bayi lahir dengan berat badan rendah ( $x_3$ ) berpengaruh signifikan terhadap angka kematian bayi di Sulawesi Selatan pada tahun 2019.

**Tabel 4** Nilai t hitung

Penduga parameter	$t_{hitung}$
$\beta_1$	-3.9906
$\beta_2$	0.9481
$\beta_3$	-2.2069
$\beta_4$	-0.7182
$\beta_5$	-0.6100

Selanjutnya akan dilakukan pemodelan berdasarkan variabel yang signifikan yaitu variabel jumlah pemberian vitamin A ( $x_1$ ) dan jumlah bayi lahir dengan berat badan rendah ( $x_3$ ) terhadap angka kematian bayi dan diperoleh model regresi *linearized ridge regression* sebagai berikut:

$$\hat{y} = 8.8467 - 0.00008x_1 - 0.0032x_3$$

adapun informasi yang dapat diperoleh berdasarkan model regresi yang telah diperoleh adalah sebagai berikut:

a.  $\beta_0 = 8.8467$

Nilai 8.8467 menunjukkan angka kematian bayi di Provinsi Sulawesi Selatan tahun 2019 variabel lain dianggap konstan atau bernilai nol.

b.  $\beta_1 = -0.00008$

Nilai -0.0008 merupakan koefisien dari variabel jumlah pemberian vitamin A pada bayi. Nilai -0.00008 menunjukkan nilai yang negatif yang berarti bahwa, apabila jumlah pemberian vitamin A pada bayi bertambah satu satuan sedangkan variabel lainnya dianggap konstan, maka angka kematian bayi akan berkurang sebesar 0.00008 dikali dengan jumlah pemberian vitamin A pada bayi. Walaupun pengaruh yang diberikan tidak begitu besar, tetapi dengan bertambahnya jumlah pemberian vitamin A pada bayi secara merata disetiap wilayah dapat mengurangi peningkatan angka kematian bayi di wilayah tersebut.

c.  $\beta_3 = -0.0032$

Nilai -0.0032 merupakan koefisien dari variabel jumlah bayi lahir dengan berat badan rendah. Nilai -0.0032 menunjukkan nilai yang negatif yang berarti bahwa, apabila jumlah bayi lahir dengan berat badan rendah berkurang satu satuan sedangkan variabel lainnya dianggap konstan, maka angka kematian bayi akan berkurang sebesar 0.0032 dikali dengan jumlah bayi lahir dengan berat badan rendah. Bayi lahir dengan berat badan rendah dapat disebabkan oleh beberapa faktor yaitu ibu yang memiliki gaya hidup yang tidak sehat seperti merokok, mengonsumsi alkohol dan menggunakan narkoba. Walaupun pengaruh yang diberikan tidak begitu besar, tetapi dengan menerapkan pola hidup sehat akan dapat meminimalisir bayi lahir dengan berat badan rendah sehingga peningkatan angka kematian bayi di wilayah tersebut akan berkurang.

#### **4. Kesimpulan**

Berdasarkan analisis dan pembahasan yang telah dilakukan, diperoleh kesimpulan dugaan model yang diperoleh untuk data Angka Kematian Bayi di Kabupaten/kota Provinsi Selatan pada tahun 2019 adalah:

$$\hat{y} = 8.8467 - 0.00008x_1 - 0.0032x_3$$

dari persamaan tersebut, dapat diambil kesimpulan bahwa variabel jumlah pemberian vitamin A pada bayi dan jumlah bayi lahir dengan berat badan rendah di tiap kabupaten/kota di Provinsi Sulawesi Selatan berpengaruh negatif terhadap angka kematian bayi di Provinsi Sulawesi Selatan pada tahun 2019.

#### **Daftar Pustaka**

- [1] Angelin, I. (2021). *Penggunaan Metode Regresi Akar Laten Robust Pada Data Yang Mengandung Multikolinearitas dan Outlier*. Matematika dan Ilmu Pengetahuan Alam, Universitas Hasanuddin.
- [2] Dinas Kesehatan (2020). *Profil Kesehatan Provinsi Sulawesi Selatan Tahun 2020*. Pusat Data dan Informasi, Makassar.
- [3] Montgomery, D. C., Peck, E. A., & Vining, G. (2021). *Introduction to Linier Regression Analysis. Fifth Edition*. United States: John Wiley & Sons, Inc.

- [4] Apriani, W. R. (2014). *Taksiran Linearized Ridge regression Sebagai penaksir Parameter Regresi Linier Berganda Pada Kasus Multikolinearitas*. Matematika dan Ilmu Pengetahuan Alam, Univerisitas Indonesia.
- [5] Liu, Q. X., & Gao, F. (2011). Linierized Ridge Regression Estimator in Linier Regression. *Communications in Statistics-Theory and Methods*, 40(12), 2182-2192
- [6] Allen, D. M. (1974). Mean Square Error of Prediction as a Criterion for Selection Variables. *Technometrics*, 13(3), 469-475.