

Pemodelan Regresi Binomial Negatif menggunakan Estimator *Jackknife Negative Binomial Ridge Regression* pada Data Angka Kematian Bayi Provinsi Sulawesi Selatan

Kezia Agra Palinoan^{1,*}, Andi Kresna Jaya¹, Anna Islamiyati¹

¹Universitas Hasanuddin

Dikirim: Juni 2023;

Diterima: September 2024;

Dipublikasi: September 2024

Alamat Email Korespondensi: kezia.palinoan.kp@gmail.com

Abstrak

Analisis regresi Binomial Negatif adalah metode yang digunakan untuk menganalisis hubungan antara variabel prediktor terhadap variabel respon yang berdistribusi Poisson. Namun, regresi Poisson tidak dapat digunakan untuk memodelkan data dengan overdispersi maupun terdapat multikolinearitas. Untuk menyelesaikan masalah tersebut digunakan regresi Binomial Negatif dengan estimator *Jackknife Negative Binomial Ridge Regression*. Dalam penelitian ini, estimasi parameter regresi Binomial Negatif dengan estimator *Jackknife Negative Binomial Ridge Regression* diterapkan pada data tingkat kematian bayi di Sulawesi Selatan tahun 2017. Metode *Jackknife* berperan untuk mereduksi bias sehingga dapat diperoleh penaksiran parameter dengan bias yang kecil sedangkan metode *ridge* untuk menangani multikolinearitas. Metode pemilihan parameter *ridge* menggunakan nilai MSE terkecil. Model terbaik terbentuk pada model dengan parameter *ridge* 0.0081. Berdasarkan estimasi parameter yang terbentuk menunjukkan bahwa variabel jumlah bayi dengan berat badan lahir rendah (X_1), jumlah bayi yang diberi ASI eksklusif (X_2), jumlah bayi yang mendapatkan vitamin A (X_3), jumlah cakupan pelayanan K4 pada ibu hamil (X_4), jumlah ibu hamil yang menerima imunisasi TT2+ (X_5), dan jumlah kelahiran (X_6) signifikan mempengaruhi jumlah kematian bayi.

Kata Kunci:

Jackknife Negative Binomial Ridge Regression, Jumlah Kematian Bayi, Multikolinearitas, Regresi Binomial Negatif

PENDAHULUAN

Angka Kematian Bayi (AKB) adalah jumlah bayi yang meninggal sebelum berumur satu tahun pada tahun tertentu per 1000 kelahiran hidup. Karena bayi baru lahir adalah kelompok usia yang paling terpapar pengaruh perubahan lingkungan dan sosial ekonomi, ciri dari kondisi sosial ekonomi masyarakat lokal dapat dilihat dari angka kematian bayinya [1]. Namun karena pencapaian penurunan angka kematian ibu dari negara industri maju, angka kematian bayi kini dianggap sebagai indikator unggulan untuk menilai kepuasan konsumen pelayanan kesehatan [2].

Analisis regresi digunakan di bidang kesehatan untuk mengetahui variabel-variabel model regresi yang memengaruhi kematian bayi baru lahir dan menyelidiki variabel yang mempengaruhi tingkat kematian bayi. Jumlah kematian bayi baru lahir di bawah satu tahun dapat digunakan untuk menghitung angka kematian bayi. Karena jumlah kematian bayi baru lahir bersifat diskrit, maka terdapat dalam data yang berdistribusi Poisson. Salah satu model regresi yang dapat digunakan untuk mengeksplorasi hubungan yang

melibatkan faktor prediktor dan variabel target secara data diskrit adalah Model Regresi Poisson dalam [3].

Untuk melakukan analisis regresi Poisson, beberapa asumsi harus dipenuhi, salah satunya adalah ekuidispersi. Namun regresi Poisson tidak dapat digunakan untuk memodelkan data dengan overdispersi yang dimana hitung data sering menunjukkan varians yang lebih besar daripada rata-rata. Salah satu strategi untuk mengatasi overdispersi pada regresi Poisson adalah regresi Binomial Negatif [4].

Estimasi parameter regresi Poisson di sisi lain, menggunakan *Maximum Likelihood* (ML). Namun, prosedur ini sering menghasilkan temuan yang salah. Multikolinearitas adalah salah satunya. Akibatnya, hasil estimasi parameter memiliki varians yang tidak nol. Karena data memiliki sedikit pengamatan dan banyak faktor yang mempengaruhinya, terdapat indikator kekhawatiran terjadi multikolinearitas dalam data kematian bayi. Pendekatan metode *ridge* yang dikembangkan oleh Hoerl dan Kennard merupakan salah satu metode untuk menangani multikolinearitas. Kemudian metode *Jackknife* berperan untuk mereduksi bias sehingga dapat diperoleh penaksiran parameter dengan bias yang kecil. [5]. Oleh karena itu, peneliti terdorong untuk melakukan penelitian menggunakan model regresi Binomial Negatif untuk mengatasi multikolinearitas yang terjadi pada data angka kematian bayi dengan menggunakan estimator *Jackknife Negative Binomial Ridge Regression*.

METODE PENELITIAN

Jenis penelitian ini adalah penelitian kuantitatif dengan menggunakan data sekunder tahun 2017 yang diperoleh dari sumber yang telah dipublikasikan Dinas Kesehatan Provinsi Sulawesi Selatan yaitu Profil Kesehatan 2018 Provinsi Sulawesi Selatan. Data penelitian ini mencakup 7 variabel yang masing-masing terdiri dari 24 Kabupaten/Kota di Provinsi Sulawesi Selatan.

Variabel pada penelitian ini terdiri dari variabel respon yaitu jumlah kematian bayi (Y) dan variabel prediktor yaitu jumlah bayi dengan berat badan lahir rendah (X_1), jumlah bayi yang diberi ASI eksklusif (X_2), jumlah bayi yang mendapatkan vitamin A (X_3), jumlah cakupan pelayanan K4 pada ibu hamil (X_4), jumlah ibu hamil yang menerima imunisasi TT2+ (X_5), dan jumlah kelahiran (X_6).

Analisis data di penelitian ini dilakukan seperti berikut:

1. Melakukan analisis data pada data penelitian.
 - a. Statistik deskriptif pada data.
 - b. Melakukan uji distribusi poisson menggunakan statistik uji Kolmogorov-Smirnov.
2. Model regresi Binomial Negatif meliputi:
 - a. Estimasi parameter pada model regresi Binomial Negatif.
 - b. Uji signifikansi secara individu dengan uji Wald.

$$W_j = \left[\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right]^2 ; j = 1, 2, \dots, p \quad (1)$$

- c. Uji signifikansi secara serentak dengan uji rasio *likelihood*.

$$LR = -2(L_{null} - L_{full}) \quad (2)$$

3. Melakukan uji multikolinieritas.

$$VIF_j = \frac{1}{1 - R_j^2} \quad (3)$$

4. Memodelkan angka kematian bayi dengan menggunakan estimator *Jackknife Negative Binomial Ridge Regression* untuk mengatasi multikolinearitas meliputi:

- a. Melakukan *centering* dan *scaling* (transformasi korelasi) pada variabel prediktor.

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{n-1}S_j} \quad (4)$$

- b. Menganalisis model regresi Binomial Negatif dengan estimator *maximum likelihood* dari hasil data *centering* dan *scaling* (β_{ML}^*).

- c. Membentuk matriks $X^T \widehat{W} X$ dari data hasil transformasi, dimana \widehat{W} adalah matriks diagonal dari $\left(\frac{\mu}{1+a\mu}\right)$.

- d. Membentuk matriks G (vektor eigen) dari hasil matriks $X^T \widehat{W} X$

- e. Membentuk matriks $Z = XG$.

- f. Menghitung estimator γ_{ML} .

$$\widehat{\gamma}_{ML} = G^{-1} \widehat{\beta}_{ML} \quad (5)$$

- g. Mencari nilai k untuk model Binomial Negatif.

$$k_1 = \frac{1}{\hat{\alpha}_{max}^2} \quad (6)$$

$$k_2 = \text{median}(s_j) \quad (7)$$

$$k_3 = \text{median}(q_j) \quad (8)$$

$$k_4 = \max(q_j) \quad (9)$$

dimana:

$$\hat{\alpha}_{max}^2 = \text{nilai maksimum dari } (G\beta_{ML}^*)^2$$

$$s_j = \frac{t_j \sigma^2}{(n-p)\sigma^2 + t_j \hat{\alpha}_j^2}$$

t_j = nilai eigen dari $X^T X$

$$\sigma^2 = \frac{(\mathbf{y} - \mu)^2}{(n-p-1)}$$

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

λ_{max} = nilai eigen maksimum dari G

- h. Mendistribusikan nilai k untuk mendapatkan estimator *Jackknife Negative Binomial Ridge Regression*.

$$\widehat{\gamma}_{JNBR} = (I + k^2 B^{-2}) \widehat{\gamma}_{ML} \quad (10)$$

$$\text{Var}(\widehat{\gamma}_{JNBR}) = (I + k^2 B^{-2}) \Lambda_{NBRR}^{-1} (I + k^2 B^{-2}) \quad (11)$$

$$\text{Bias}(\widehat{\gamma}_{JNBR}) = -k^2 B^{-2} \widehat{\gamma}_{ML} \quad (12)$$

$$\text{MSE}(\widehat{\gamma}_{JNBR}) = (I + k B^{-1}) \Lambda_{NBRR}^{-1} (I + k B^{-1}) + k^4 B^{-2} \widehat{\gamma}_{ML} \widehat{\gamma}_{ML}^T B^{-2} \quad (13)$$

- i. Membentuk model *Jackknife Negative Binomial Ridge Regression* berdasarkan parameter ridge k yang memiliki nilai MSE terkecil.

- j. Mengonversikan parameter $\widehat{\gamma}_{JNBR}$ ke dalam parameter β_j^* .

$$\beta_j^* = G \widehat{\gamma}_{JNBR} \quad (14)$$

- k. Mentransformasikan ke bentuk awal dan membentuk persamaan model regresi Binomial Negatif yang baru.

$$\beta_j = \left(\frac{1}{\sqrt{n-1}S_j} \right) \beta_j^* \quad (15)$$

$$\beta_0 = \ln(\bar{Y}) - \beta_1\bar{X}_1 - \beta_2\bar{X}_2 - \beta_3\bar{X}_3 - \beta_4\bar{X}_4 - \beta_5\bar{X}_5 \quad (16)$$

HASIL DAN PEMBAHASAN

1. Analisis Data

a. Deskriptif Data

Deskriptif data digunakan untuk mengetahui rata-rata (mean), minimum, maksimum dan standar deviasi pada penelitian ini. deskriptif data variabel yang digunakan ditunjukkan pada Tabel 1.

Tabel 1. Deskriptif Data

Variabel	Minimum	Maksimum	Rata-rata	Std. Deviasi
Y	1	91	44	23.2852
X ₁	46	846	225	154.4227
X ₂	268	15960	2831	3690.3532
X ₃	167	24094	3868	4617.4662
X ₄	1951	27525	6299	5411.2431
X ₅	0	20860	5226	4809.5931
X ₆	1941	26043	6297	5042.4342

b. Uji Distribusi Poisson

Pengujian distribusi Poisson pada variabel dependen atau terikat (Y) menggunakan uji Kolmogorov-Smirnov. Berdasarkan hasil *output RStudio* diperoleh $D_{hit} = 0.3093 < D_{(0.01,24)} = 0.32$ dan $p - value = 0.0202 > \alpha = 0.01$, maka H_0 diterima. Sehingga dapat disimpulkan bahwa data berdistribusi Poisson.

2. Model Regresi Binomial Negatif

a. Pengujian Estimasi Parameter Model

Pengujian estimasi parameter pada model regresi Binomial Negatif yang diduga menggunakan estimator *Maksimum Likelihood*. menggunakan *software Rstudio* ditunjukkan pada Tabel 2.

Tabel 2. Estimasi Model Regresi Binomial Negatif

Variabel	Parameter	Estimasi	Std. Error
Konstanta	β_0	3.2165	0.2300
X ₁	β_1	0.0017	0.0018
X ₂	β_2	2.2594×10^{-4}	1.1339×10^{-4}
X ₃	β_3	-2.6161×10^{-4}	1.1216×10^{-4}
X ₄	β_4	-1.6363×10^{-4}	3.2580×10^{-4}
X ₅	β_5	-1.1112×10^{-4}	7.3187×10^{-5}
X ₆	β_6	3.3492×10^{-4}	3.3943×10^{-4}
	a	4.9351	

Jadi, model regresi Binomial Negatif yang terbentuk, yaitu:

$$\ln(\mu_i) = 3.2165 + 0.0017x_{i1} + 2.2594 \times 10^{-4}x_{i2} - 2.6161 \times 10^{-4}x_{i3} - 1.6363 \times 10^{-4}x_{i4} - 1.1112 \times 10^{-4}x_{i5} + 3.3492 \times 10^{-4}x_{i6}$$

b. Pengujian Signifikansi Parameter Model

- i. Uji signifikansi secara individu
Penolakan H_0 jika $W_j > \chi^2_{(0.2,1)}$ atau $p - value < 0.2$. Hasil *output* dari *software Rstudio* menggunakan persamaan (1) ditunjukkan pada Tabel 3.

Tabel 3. Nilai Uji Wald

Parameter	W	$p - value$
β_1	0.9167	1.9574
β_2	1.9926	3.5930
β_3	-2.3325	4.6309
β_4	-0.5023	1.9672
β_5	-1.5184	1.2892
β_6	0.9867	3.2378

Dengan $\alpha = 0.2$, dapat dilihat pada Tabel 3 bahwa $p - value < \alpha$ maka jumlah bayi dengan berat badan lahir rendah (X_1), jumlah bayi yang diberi ASI eksklusif (X_2), jumlah bayi yang mendapatkan vitamin A (X_3), jumlah cakupan pelayanan K4 pada ibu hamil (X_4), jumlah ibu hamil yang menerima imunisasi TT2+ (X_5), dan jumlah kelahiran (X_6) memiliki pengaruh yang signifikan terhadap jumlah kematian bayi. Sehingga semua variabel prediktor layak dimasukkan ke dalam model regresi Binomial Negatif.

- ii. Uji signifikansi secara simultan
Penolakan H_0 jika $LR > \chi^2_{(0.2,5)}$ atau $p - value < 0.2$. Hasil *output* dari *software Rstudio* menggunakan persamaan (2) diperoleh $LR = 12.1710 > \chi^2_{(0.2,5)} = 7.2890$ sehingga terdapat minimal satu variabel prediktor yang berpengaruh secara signifikan terhadap model.

3. Pengujian Multikolinearitas

Uji multikolinearitas dilakukan dengan melihat nilai VIF. Hasil *output* dari *software Rstudio* menggunakan persamaan (3) ditunjukkan pada Tabel 4.

Tabel 4. Nilai VIF

Variabel	Nilai VIF
X_1	8.2756
X_2	17.6578
X_3	27.3281
X_4	321.2357
X_5	12.5086
X_6	304.1263

Pada Tabel 4 terdapat nilai $VIF > 10$ yaitu variabel X_1, X_2, X_3, X_4, X_5 dan X_6 yang menunjukkan bahwa terjadi multikolinearitas.

4. Analisis Jackknife Negative Binomial Ridge Regression

a. Estimasi Parameter Model Data Transformasi

Sebelum melakukan penanganan multikolinearitas, data terlebih dahulu dilakukan transformasi terhadap variabel prediktor menggunakan persamaan (4). Selanjutnya dilakukan pendugaan estimasi parameter model regresi Binomial Negatif pada data yang telah ditransformasi menggunakan estimator *Maksimum Likelihood*. Hasil penduga parameter dengan menggunakan *software RStudio* ditunjukkan pada Tabel 5.

Tabel 5. Estimasi Model Regresi Binomial Negatif Data yang Ditransformasi

Parameter	Penduga Parameter
β_1	2.0784
β_2	4.3209
β_3	-6.7862
β_4	3.8592
β_5	-2.6761
β_6	4.6817
a	2.0784

Kemudian dilakukan menghitung penduga (γ_{ML}) menggunakan persamaan (5). Hasil penduga parameter menggunakan *software RStudio* ditunjukkan pada Tabel 6.

Tabel 6. Penduga Parameter γ_{ML}

Parameter	Penduga Parameter
γ_1	4.9353
γ_2	-0.3661
γ_3	0.1942
γ_4	0.4013
γ_5	-2.2860
γ_6	-8.7083
a	-7.7412

b. Estimasi Parameter Ridge k

Nilai tetapan bias k digunakan untuk menghitung nilai $\mathbf{K} = k\mathbf{I}$ dengan \mathbf{K} adalah matriks diagonal, yang akan ditambahkan pada diagonal matriks Λ_{NBRR}^{-1} . Berdasarkan metode yang digunakan, terdapat 4 parameter yang akan digunakan untuk menghitung nilai k pada persamaan (6), (7), (8), dan (9). Hasil perhitungan penduga parameter k ditunjukkan pada Tabel 7.

Tabel 7. Penduga Parameter k

Parameter	Penduga Parameter
k_1	0.0081
k_2	0.0073
k_3	1.4124×10^{-5}
k_4	1.4125×10^{-5}

c. Estimasi Parameter Jackknife Negative Binomial Ridge Regression

Dari hasil estimasi k pada Tabel 7 menggunakan γ_{ML} akan dihitung penduga parameter, variansi, bias, serta nilai MSE dari metode *Jackknife Negative Binomial Ridge Regression* menggunakan persamaan (10), (11), (12) dan (13). Berikut hasil *output* dari *software RStudio* pada Tabel 8.

Tabel 7. Penduga Parameter k

Parameter	Penduga Parameter			
	k_1	k_2	k_3	k_4
γ_1	4.9353	4.9353	4.9353	4.9353
γ_2	-0.3661	-0.3661	-0.3661	-0.3661
γ_3	0.1927	0.1929	0.1942	0.1942
γ_4	0.3795	0.3827	0.4013	0.4013
γ_5	-2.1295	-2.1516	-2.2860	-2.2860
γ_6	-5.0732	-5.3334	-8.7082	-8.7082
a	-0.5846	-0.6432	-7.7276	-7.7276
Bias	64.4555	61.7908	1.8422×10^{-4}	1.8428×10^{-4}

Variansi	179.3034	192.4818	3404.3090	3404.3074
MSE	243.7589	254.2726	3404.3092	3404.3075

Berdasarkan Tabel 8 dapat dilihat nilai MSE terkecil yaitu sebesar 243.7589 dengan parameter *ridge* $k_1 = 0.0081$. Sehingga diperoleh persamaan regresi dengan parameter dispersi $a = -0.5846$ sebagai berikut.

$$\ln(\mu_i) = 4.9353x_{i1}^* - 0.3661x_{i2}^* + 0.1927x_{i3}^* + 0.3795x_{i4}^* - 2.1295x_{i5}^* - 5.0732x_{i6}^*$$

Kemudian mengonversikan penduga parameter *Jackknife Negative Binomial Ridge Regression* (γ_{JNBR}) menjadi peduga parameter regresi awal (β_j^*) menggunakan Persamaan (14). Sehingga didapatkan model regresi dengan parameter dispersi $a = 4.9352$ sebagai berikut.

$$\ln(\mu_i) = 1.5393x_{i1}^* + 2.3411x_{i2}^* - 4.1886x_{i3}^* + 0.2672x_{i4}^* - 1.1263x_{i5}^* + 2.0484x_{i6}^*$$

d. Model Regresi Binomial Negatif Baru

Model regresi baru diperoleh dengan melakukan transformasi pada peduga parameter regresi (β_j) dengan persamaan (15) dan (16) ditunjukkan pada Tabel 9.

Tabel 9. Penduga Parameter β_j

Parameter	Penduga Parameter
β_0	3.3329
β_1	0.0021
β_2	1.3228×10^{-4}
β_3	-1.8915×10^{-4}
β_4	1.0295×10^{-5}
β_5	-4.8830×10^{-5}
β_6	8.4707×10^{-5}
a	4.9352

Sehingga bentuk model regresi Binomial Negatif dengan estimator *Jackknife Negative Binomial Ridge Regression* sebagai berikut.

$$\ln(\mu_i) = 3.3329 + 0.0021x_{i1} + 1.3228 \times 10^{-4}x_{i2} - 1.8915 \times 10^{-4}x_{i3} + 1.0295 \times 10^{-5}x_{i4} - 4.8830 \times 10^{-5}x_{i5} + 8.4707 \times 10^{-5}x_{i6}$$

PENUTUP

Hasil analisis dan pembahasan yang diperoleh berdasarkan model umum regresi Binomial Negatif yaitu

$$\eta_i = \ln(\mu_i) = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \beta_4x_{i4} + \beta_5x_{i5} + \beta_6x_{i6}$$

kemudian diperoleh kesimpulan model terbaik dari regresi Binomial Negatif yang menggunakan estimator *Jackknife Negative Binomial Ridge Regression* pada Angka Kematian Bayi di Provinsi Sulawesi Selatan tahun 2017 dengan parameter *ridge* $k_1 = 0.0081$ dan parameter dispersi $a = 4.9352$ yaitu

$$\ln(\mu_i) = 3.3329 + 0.0021x_{i1} + 1.3228 \times 10^{-4}x_{i2} - 1.8915 \times 10^{-4}x_{i3} + 1.0295 \times 10^{-5}x_{i4} - 4.8830 \times 10^{-5}x_{i5} + 8.4707 \times 10^{-5}x_{i6}$$

DAFTAR PUSTAKA

- [1] Wardani, D. K., & Wulandari, A. (2020). Pemodelan Negative Binomial Regression Pada Data Jumlah Kematian Bayi Di Kabupaten Jombang. *Transformasi: Jurnal Pendidikan Matematika Dan Matematika*, 4 (2), 311-320.

- [2] Cabral, A. H., Udus, M. Y., Jamlean, S. F., Pramesti, W., & Anuraga, G. (2019). Pemodelan Faktor yang Mempengaruhi Angka Kematian Bayi di Jawa Timur dengan Menggunakan Geographically Weighted Regression. *SNHRP*, 37-49.
- [3] Jao, N., Islamiyati, A., & Sunusi, N. (2022). Pemodelan Regresi Nonparametrik Spline Poisson Pada Tingkat Kematian Bayi di Sulawesi Selatan. *ESTIMASI: Journal of Statistics and Its Application*, 3 (1), 14-22.
- [4] Islami, Y. N., Ispriyanti, D., & Kartikasari, P. (2021). Perbandingan Model Regresi Binomial Negatif Bivariat Dengan Model Geographically Weighted Negative Binomial Bivariat Regression (GWNBBR) Pada Kasus Angka Kematian Bayi Dan Kematian Ibu Di Jawa Tengah. *Jurnal Gaussian*, 10 (4), 488-498.
- [5] Munawaroh, A. (2018). Estimator Baru Modified Jackknifed Untuk Mengatasi Multikolinieritas Pada Regresi Poisson Studi Kasus: Angka Kematian Bayi Provinsi Jawa Timur Tahun 2016 (Skripsi, Fakultas Sains dan Teknologi UIN Syarif Hidayatullah Jakarta).