

Identifikasi Data Berpengaruh dengan Pemodelan Regresi Beta pada Kasus Angka Kematian Ibu di Indonesia Tahun 2022

Muhammad Bahry Baidhowy*, Nusar Hajarisman

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Islam Bandung, Indonesia.

*arycaw15@gmail.com, nusarhajarisman@unisba.ac.id

Abstract. Beta Regression is an analysis method to see the effect of predictor variables on response variables in the form of proportions. To include the average response along with its dispersion parameters, it is necessary to reparameterize the beta distribution density function. To estimate the model parameters, the Maximum Likelihood Estimator (MLE) method is used, with a log-likelihood function derivative where the estimation process can be solved numerically using the Fisher Scoring algorithm. In one data set, there may be data that greatly determines the regression model, which is caused by the data being outliers or influential data. To identify outliers, Standardized Residual and Leverage values can be used. Meanwhile, to identify influential data, Cook's Distance can be used. The data used in this study are secondary data sourced from the Indonesian Ministry of Health. An influential data was identified, namely the 33rd observation (West Papua Province), so this observation was excluded from the analysis. Then only the variables of maternal health services K4 (X_1) and postpartum health services KF Complete (X_5) were obtained which had a significant effect on maternal mortality. Of the 33 provinces in Indonesia in 2022 (apart from West Papua), the highest maternal mortality rate was in West Sulawesi Province, and the lowest maternal mortality rate was in Aceh Province.

Keywords: *Maternal mortality, Cook's Distance, Influential data, Beta Regression.*

Abstrak. Regresi Beta merupakan metode analisis untuk melihat pengaruh dari variabel prediktor, terhadap variabel respon yang berbentuk proporsi. Untuk menyertakan rata-rata respon bersamaan dengan parameter dispersinya, maka perlu dilakukan reparameterisasi dari fungsi densitas distribusi beta. Untuk mengestimasi parameter model digunakan metode Maximum Likelihood Estimator (MLE), dengan turunan fungsi log-likelihood dimana proses penaksirannya dapat diselesaikan secara numerik dengan menggunakan algoritma Fisher Scoring. Dalam satu set data memungkinkan terdapat data yang sangat menentukan model regresi, yang disebabkan data tersebut merupakan data pencilan atau data berpengaruh. Untuk mengidentifikasi data pencilan dapat menggunakan Standardized Residual dan nilai Leverage. Sedangkan untuk mengidentifikasi data berpengaruh dapat menggunakan Cook's Distance. Data yang digunakan dalam penelitian ini merupakan data sekunder yang bersumber dari Kementerian Kesehatan RI. Teridentifikasi sebuah data berpengaruh yaitu pengamatan ke-33 (Provinsi Papua Barat), sehingga pengamatan ini disisihkan dari analisis. Lalu didapatkan hanya variabel pelayanan kesehatan ibu hamil K4 (X_1) dan pelayanan kesehatan ibu nifas KF Lengkap (X_5) yang berpengaruh signifikan terhadap angka kematian ibu. Dari 33 Provinsi di Indonesia tahun 2022 (selain Papua Barat) angka kematian ibu tertinggi terdapat pada Provinsi Sulawesi Barat, dan angka kematian ibu terendah terdapat pada Provinsi Aceh.

Kata Kunci: *Angka Kematian Ibu, Cook's Distance, Data Berpengaruh, Regresi Beta.*

A. Pendahuluan

Analisis Regresi merupakan metode statistik yang digunakan di berbagai bidang untuk memahami bagaimana variabel prediktor mempengaruhi variabel respon. Menurut Hajarisman (2012), model regresi biasa mengasumsikan residual berdistribusi normal dan diestimasi menggunakan Metode Kuadrat Terkecil (MKT). Namun, untuk data proporsi (0 hingga 1), MKT tidak cocok digunakan karena akan menimbulkan masalah non-linieritas dan heteroskedastisitas. Penggunaan Regresi Beta lebih cocok untuk data proporsi, karena memberikan penaksiran parameter yang akurat dan efisien dengan estimasi parameter menggunakan *Maximum Likelihood Estimator* (MLE) [1].

Dalam suatu analisis data memungkinkan terdapat ketidaksesuaian atau kesalahan, hal ini mungkin dikarenakan terdapatnya *outliers* ataupun *influential data*. Data pencilan atau *outliers* dapat diidentifikasi menggunakan matriks Hat dan *Standardized residuals*. Sedangkan untuk mengidentifikasi data berpengaruh atau *influential data* dapat menggunakan statistik Cook's Distance [2].

Pada penelitian ini menerapkan metode yang dipaparkan sebelumnya pada data Angka Kematian Ibu (AKI) di Indonesia tahun 2022. Kematian ibu didefinisikan sebagai kematian selama kehamilan, persalinan, dan nifas yang diakibatkan komplikasi terkait diluar penyebab kecelakaan. Menurut Kementerian Kesehatan RI (2023), AKI merupakan indikator kesehatan yang sensitif, karena mencerminkan interaksi berbagai aspek. Pemerintah telah melakukan upaya untuk menurunkan AKI dengan mengadakan pelayanan kesehatan kepada ibu, baik ketika masa kehamilan, persalinan, maupun nifas [3].

Berdasarkan latar belakang yang telah diuraikan, maka perumusan masalah dalam penelitian ini adalah : “Apakah terdapat data pencilan yang berpotensi sebagai data berpengaruh?, lalu bagaimana pengaruhnya terhadap model regresi beta?” dan “Bagaimana pengaruh dari pelayanan kesehatan ibu terhadap AKI di Indonesia tahun 2022?”.

B. Metodologi Penelitian

Dalam penelitian digunakan data sekunder, dimana data ini bersumber dari Publikasi Kementerian Kesehatan Indonesia yang berjudul Profil Kesehatan Indonesia 2022. Data yang peneliti gunakan mencakup Angka Kematian Ibu dan Persentase Pelayanan Ibu di setiap Profinsi di Indonesia pada tahun 2022. Data yang peneliti gunakan terdiri dari satu variabel respons, enam variabel prediktor, dan 34 pengamatan.

Distribusi Beta

Menurut Hajarisman (2012), distribusi beta dikatakan cukup fleksibel untuk memodelkan data dengan variabel respon yang berbentuk proporsi, dengan selang interval antara 0 hingga 1. Fungsi densitas variabel acak Y yang mengikuti distribusi betasebagai berikut.

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}, \quad 0 < y < 1 \quad (1)$$

Dimana $a > 0, b > 0$, dan $\Gamma(\cdot)$ merupakan simbol fungsi gamma. Dengan $E(Y) = a/a+b$ dan $var(Y) = ab/(a+b)^2(a+b+1)$. [1]

Model Regresi Beta

Dalam model ini terdapat dua parameter, yaitu parameter rata-rata dan parameter dispersi yang keduanya mencirikan parameter berdistribusi beta. Untuk membentuk model regresi beta dengan menyertakan rata-rata respon dengan parameter dispersinya secara bersamaan perlu dilakukan reparameterisasi dari fungsi densitas beta. Dimisalkan $\mu = a/(a+b)$ dan $a+b = \kappa$, sehingga diperoleh $a = \mu\kappa$ dan $b = (1-\mu)\kappa$. Sehingga fungsi densitasnya menjadi.

$$f(y; \mu, \kappa) = \frac{\Gamma(\kappa)}{\Gamma(\mu\kappa)\Gamma((1-\mu)\kappa)} y^{\mu\kappa-1} (1-y)^{(1-\mu)\kappa-1}, \quad 0 < y < 1 \quad (2)$$

dimana μ = rata-rata dari variabel respon, dan κ = parameter dispersi. reparameterisasi ini menyatakan bahwa $0 < \mu < 1$ dan $\kappa > 0$, dan juga ditunjukkan bahwa $a = \mu\kappa > 0$ dan $b =$

$$\kappa(1 - \mu) > 0 \quad [1].$$

Bentuk Model Regresi Beta

Model regresi beta dibentuk dengan pendekatan *Generalized Linear Model* (GLM) dengan dua fungsi hubung, yaitu untuk parameter lokasi μ dan parameter dispersi κ . Misalkan X dan W adalah matriks kovariat, dengan x_i dan w_i adalah vektor baris ke- i dari kedua matriks tersebut. Selain itu, β dan δ diasumsikan sebagai vektor koefisien regresi. Maka model linier umum untuk parameter lokasinya

$$g(\mu_i) = x_i^T \beta \quad (3)$$

$g(\cdot)$ merupakan fungsi hubung yang mempunyai turunan [1].

Penaksiran Parameter

Fungsi hubung diperlukan untuk memetakan parameter dari selang terbatas 0 sampai 1 ke ruang bilangan nyata. Fungsi hubung logit merupakan fungsi yang akan mengembalikan penaksir parameter dalam bentuk *odds ratio*. fungsi ini didefinisikan

$$g(\mu) = \text{logit}(\mu) = \ln\left(\frac{\mu}{1 - \mu}\right) = x_i^T \beta \rightarrow \mu = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} \quad (4)$$

Selanjutnya untuk mengestimasi parameter menggunakan *maksimum likelihood* digunakan metode iterasi numerik *Fisher Scoring*, dimana digunakan turunan fungsi *log-likelihood* dalam proses perhitungannya. Fungsi *log-likelihood* yang didasarkan pada sampel n pengamatan yang independen diberikan oleh

$$l(\beta, \kappa) = \sum_{i=1}^n l(\mu_i, \kappa) ; \text{dimana} \quad (5)$$

$$l(\mu_i, \kappa) = \log \Gamma(\kappa) - \log \Gamma(\mu_i \kappa) - \log \Gamma((1 - \mu_i) \kappa) + (\mu_i \kappa - 1) \log y_i + \{(1 - \mu_i) \kappa\} \log (1 - y_i)$$

misalkan $z_i = \log \{y_i / (1 - y_i)\}$ dan $\mu_i^* = \psi(\mu_i \kappa) - \psi((1 - \mu_i) \kappa)$. Selanjutnya vektor skor adalah turunan pertama terhadap β dan κ dari fungsi *log-likelihood*, sebagai berikut.

$$U(\beta, \kappa) = \frac{\partial l(\mu_i, \kappa)}{\partial \theta} = \begin{bmatrix} \frac{\partial l(\mu_i, \kappa)}{\partial \beta} \\ \frac{\partial l(\mu_i, \kappa)}{\partial \kappa} \end{bmatrix} \quad (6)$$

$$= \begin{bmatrix} \kappa X^T W (z_i - \mu_i^*) \\ \sum_{i=1}^n \{\mu_i (z_i - \mu_i^*) + \log(1 - y_i) - \psi((1 - \mu_i) \kappa) + \psi(\kappa)\} \end{bmatrix}$$

Sedangkan matriks informasi merupakan kedua dari fungsi *log-likelihood* yang diberikan dalam persamaan (5) terhadap parameter (β, κ) diberikan oleh

$$I(\beta, \kappa) = \begin{bmatrix} \frac{\partial l(\beta, \kappa)}{\partial \beta^2} & \frac{\partial l(\beta, \kappa)}{\partial \beta \partial \kappa} \\ \frac{\partial l(\beta, \kappa)}{\partial \kappa \partial \beta} & \frac{\partial l(\beta, \kappa)}{\partial \kappa^2} \end{bmatrix} \quad (7)$$

dengan $\partial l(\beta, \kappa) / \partial \beta^2 = \kappa X^T W X$, $\partial l(\beta, \kappa) / \partial \beta \partial \kappa = X^T T c$, dan $\partial l(\beta, \kappa) / \partial \kappa^2 = \text{tr}(D)$ [1].

Uji Signifikansi Parameter Secara Simultan

Menurut Agresti (2007) [4], untuk mengetahui signifikansi parameter secara serempak dapat dilakukan pengujian dengan menggunakan *likelihood ratio* atau Uji G sebagai berikut.

$$G = -2 \ln \frac{L(\omega)}{L(\Omega)} = 2(L(\omega) - L(\Omega)) \quad (8)$$

dimana

$L(\omega)$ = *likelihood* model tanpa menyertakan variabel prediktor

$L(\Omega)$ = *likelihood* model dengan menyertakan variabel prediktor

Dimana aturan pengambilan keputusan adalah tolak H_0 jika $G > \chi^2_{(a,db)}$, dengan $db = k-1$.

Uji Signifikansi Parameter Secara Parsial

Menurut Agresti (2007) [4], Untuk mengetahui signifikansi parameter secara individual digunakan uji Wald, karena pengujian ini cocok untuk model yang parameternya diestimasi menggunakan MLE. Dimana statistik uji wald.

$$W = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}, SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}, Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum_i^n (x_{ij} - \bar{x}_j)^2} \quad (9)$$

dengan:

$\hat{\beta}_j$ = nilai estimasi parameter variabel prediktor ke- j .

$SE(\hat{\beta}_j)$ = nilai *standard error* dari estimasi variabel prediktor ke- j .

$j = 1, 2, \dots, k$.

Dengan pengamiran keputusan jika nilai $W > Z_{1-\alpha/2}$ atau $W < -Z_{1-\alpha/2}$, maka H_0 ditolak.

Analisis Residu

Standardized Residuals

Adanya data pencilan dapat menyebabkan pelanggaran asumsi, sehingga menyebabkan model menjadi tidak layak. Salah satu cara untuk mendeteksi data pencilan (*outliers*) dapat menggunakan *standardized residual* dimana dapat dihitung menggunakan persamaan berikut.

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{var}(y_i)}} \quad (10)$$

dimana $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$ dan $\widehat{var}(y_i) = (1 - \mu_i)\mu_i/(\hat{k} + 1)$. Nilai residua yang jauh dari kumpulan data lainnya dapat dikatakan sebagai *outliers* [5].

Matriks Hat

Nilai *leverage* merupakan unsur diagonal dari matriks Hat, untuk membentuk matriks Hat dapat menggunakan persamaan berikut.

$$H = W^{1/2} X (X^T W X)^{-1} X^T W^{1/2} \quad (11)$$

Dimana W merupakan matriks pembobot, dan X merupakan matriks desain. Penggunaan matriks pembobot dalam regresi beta diperlukan untuk menangani masalah heteroskedastisitas, dan memastikan model mempertimbangkan variabilitas yang berbeda antar pengamatan. Rata-rata nilai *leverage* adalah k/n , maka untuk setiap h_{ii} yang lebih besar dari $2k/n$ dapat dikatakan sebagai data pencilan yang berpotensi sebagai data berpengaruh [5].

Data Berpengaruh

Cook's Distance

Untuk pemeriksaan data berpengaruh dapat menggunakan statistik Cook's *Distance*, dimana perhitungannya menggunakan nilai *standardized residual* dan *leverage*. Untuk menghitung nilai

Cook's D dapat menggunakan persamaan berikut.

$$CD_i = \frac{h_{ii}r_i^2}{k(1 - h_{ii})^2} \quad (12)$$

Menurut Collet (2003) [6], dijelaskan bahwa batas dari nilai Cook's *distance* dikatakan masih baik yaitu mengikuti distribusi t. Nilai dari ambang batas ini didapat dari tabel distribusi t dengan alpha tertentu, dan nilai $df = n - k - 1$ [7].

Angka Kematian Ibu Dan Pelayanan Kesehatan Ibu

Menurut laporan Kementerian Kesehatan RI (2023), kematian ibu terjadi selama kehamilan, persalinan, atau dalam 42 hari setelah persalinan akibat kondisi terkait kehamilan, bukan karena kecelakaan. AKI digunakan untuk mengevaluasi program kesehatan ibu dan mengukur tingkat kesehatan masyarakat karena sensitivitasnya terhadap peningkatan layanan kesehatan.

Pemerintah telah mengimplementasikan berbagai program untuk menurunkan AKI, termasuk layanan antenatal dengan pemeriksaan minimal, dan memastikan setiap persalinan dilakukan oleh tenaga kesehatan di fasilitas kesehatan. Layanan pasca melahirkan juga harus dilakukan minimal empat kali, dengan evaluasi keberhasilan program melalui indikator K4, K6, KF1, KFLengkap, dan Vitamin A [3].

Data yang digunakan dalam penelitian ini merupakan data AKI pada setiap Provinsi di Indonesia tahun 2022, berikut merupakan variabel penelitian yang digunakan dalam penelitian ini.

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Keterangan
Y	AKI	Angka kematian ibu yang berhubungan dengan kehamilan dan persalinan
X ₁	Persentase K4	Persentase ibu hamil yang mendapat perawatan antenatal sesuai standar minimal empat kali
X ₂	Persentase K6	Persentase ibu hamil yang mendapat perawatan antenatal sesuai standar minimal enam kali, serta minimal dua kali pemeriksaan dokter
X ₃	Persentase Fasyankes	Persentase ibu bersalin mendapatkan pelayanan persalinan di fasilitas layanan kesehatan
X ₄	Persentase KF1	Persentase ibu nifas yang mendapat kunjungan ke-1
X ₅	Persentase KF Lengkap	Persentase ibu nifas yang mendapat kunjungan lengkap empat kali
X ₆	Persentase Vitamin A	Persentase ibu nifas yang mendapatkan kapsul vitamin A

C. Hasil Penelitian dan Pembahasan

Pemodelan Regresi Beta

Seperti yang telah diuraikan sebelumnya, dalam penelitian ini untuk memodelkan dengan bentuk regresi beta karena data variabel respons yang berbentuk proporsi. Dengan menggunakan fungsi hubung logit, dilakukan estimasi pemodelan dengan mengasumsikan tidak terdapat *outliers* dan *influential data* dapat dilihat pada Tabel 2.

Tabel 2. Estimasi Model Regresi Beta

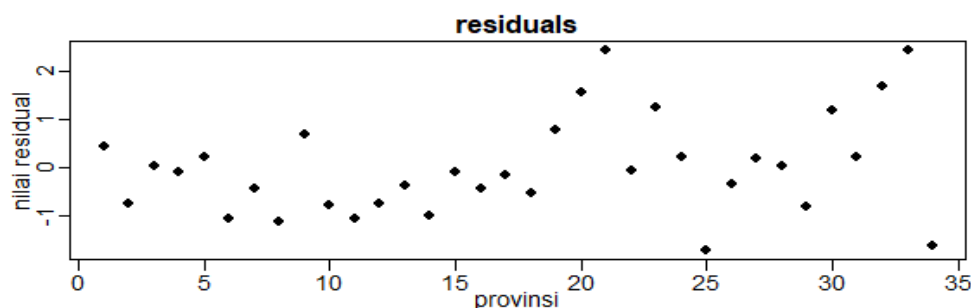
Variabel	Koefisien	Galat Baku	z-value	p-value
intercept	-6,9805	0,5135	-13,5940	<0,0001
X_1	-0,0359	0,0128	-2,8122	0,0049
X_2	-0,0058	0,0068	-0,8495	0,3956
X_3	0,0378	0,0138	2,7441	0,0061
X_4	-0,0610	0,0307	-1,9908	0,0465
X_5	0,0498	0,0220	2,2669	0,0234
X_6	0,0151	0,0113	1,3350	0,1819

Pseudo R-squared: 0.2046

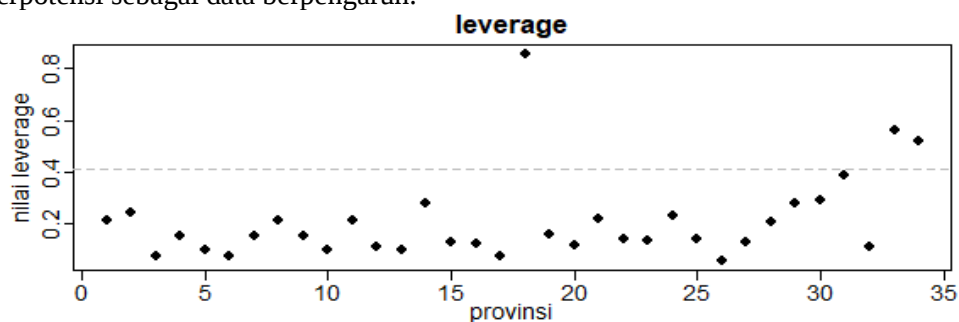
Selanjutnya dilakukan pengujian untuk parameter regresi dalam model secara simultan dengan menggunakan Uji G. diperoleh nilai dari $G = 9,9710$, sedangkan nilai dari tabel *chi-square* dengan alpha 5% sebesar 11,070. Artinya hipotesis 0 diterima, sehingga dikatakan tidak terdapat pengaruh dari variabel prediktor terhadap variabel respon. Berdasarkan hasil Uji G, bahwa hasil estimasi model mengindikasikan ketidak sesuaian, dimana berdasarkan Tabel 2 terdapat beberapa variabel prediktor yang berpengaruh signifikan terhadap variabel respon.

Identifikasi Data Berpengaruh

Langkah awal untuk mengidentifikasi data berpengaruh (*influential data*), lebih dulu mengidentifikasi data pencilan (*outliers*) dengan melihat plot dari nilai *standardized residuals* dan nilai *leverage*.

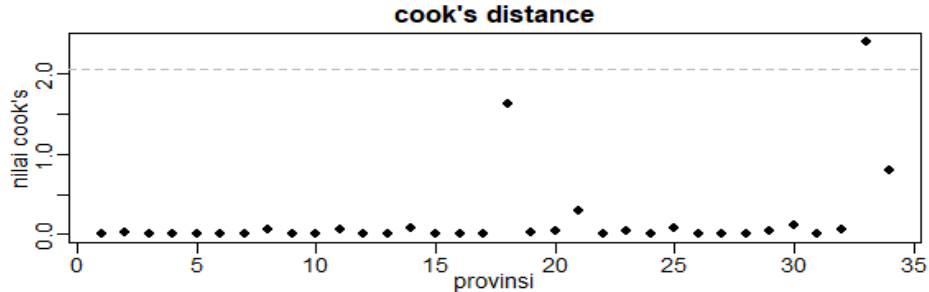
**Gambar 1.** Diagram Pencar Nilai *Standardized Residuals*

Pada Gambar 1 terlihat bahwa terdapat pengamatan yang nilainya jauh dari kumpulan pengamatan lainnya. Pengamatan ke-33 mengindikasikan bahwa ia merupakan data pencilan yang berpotensi sebagai data berpengaruh.

**Gambar 2.** Diagram Pencar Nilai *Leverage*

Berdasarkan Gambar 2 terindikasi data pencilan yang berpotensi sebagai data berpengaruh diantaranya pengamatan ke-18, 33, dan ke-34 yang memiliki nilai lebih besar dari $2k/n = 0,4118$.

Perlu diperhatikan bahwa tidak semua data pencilan merupakan data berpengaruh, dan tidak semua data berpengaruh merupakan data pencilan. Maka selanjutnya dilakukan analisis untuk mengidentifikasi data berpengaruh dengan menggunakan Cook's D.



Gambar 3. Diagram Pencar Nilai Cook's Distance

Berdasarkan Gambar 3. menggunakan batas nilai wajar Cook's D mengikuti distribusi t dengan alpha 5%, didapatkan nilai sebesar 2,052. Terlihat bahwa pengamatan ke-33 memiliki nilai yang berada diatas nilai batas wajar tersebut. Maka langkah selanjutnya, pengamatan ke-33 disisihkan dari model, untuk melihat pengaruhnya terhadap model regresi.

Revisi Model Regresi Beta

Tabel 3. Hasil Revisi Estimasi Model Regresi Beta

Variabel	Koefisien	Galat Baku	z-value	p-value
intercept	-7,2970	0,4701	-15,5232	<0,0001
X ₁	-0,0291	0,0110	-2,6649	0,0077
X ₂	-0,0013	0,0061	-0,2086	0,8348
X ₃	0,0121	0,0128	0,9469	0,3437
X ₄	-0,0245	0,0273	-0,8970	0,3698
X ₅	0,0433	0,0190	2,2804	0,0226
X ₆	0,0038	0,0100	0,3795	0,7043

Pseudo R-squared: 0,2921

Tabel 3. Datas merupakan hasil pembentukan model regresi beta tanpa pengamatan ke-33. Maka estimasi model dapat dituliskan sebagai berikut.

$$\log \left(\frac{\mu}{1 - \mu} \right) = -7,2970 - 0,0291X_1 - 0,0013X_2 + 0,0121X_3 - 0,0245X_4 + 0,0433X_5 + 0,0038X_6$$

Berdasarkan Tabel 3. Terdapat peningkatan pada nilai R-square yang menjadi sebesar 0,2921 atau dapat dikatakan variabel respon dapat dijelaskan oleh variabel prediktor sebesar 29,21%, sedangkan 70,79% lainnya dijelaskan oleh variabel lain.

Selanjutnya dilakukan pengujian parameter secara simultan dan secara parsial untuk melihat pengaruh dari variabel prediktor terhadap variabel respon. Untuk uji simultan dengan uji G didapatkan nilai G hitung sebesar 13,6608 yang artinya nilai G hitung lebih besar dari nilai chi-square tabel dengan alpha 5% sebesar 11,070. Maka dapat diputuskan hipotesis 0 ditolak, dengan kata lain minimal terdapat satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Lalu pengujian selanjutnya menggunakan uji Wald, nilai Z tabel dengan alpha sebesar 5% sebesar $\pm 1,96$. maka variabel prediktor yang berpengaruh signifikan terhadap variabel

respon yaitu variabel X_1 (persentase K4) dan X_5 (persentase KFLengkap). Maka langkah selanjutnya dilakukan pemodelan kembali dengan hanya menyertakan variabel X_1 dan X_5 . Hasil dari pemodelan disajikan dalam tabel berikut ini.

Tabel 4. Revisi Estimasi Model X_1 dan X_5

Variabel	Koefisien	Galat Baku	z-value	p-value
intercept	-7,3843	0,3946	-18,7125	<0,0001
X_1	-0,0234	0,0082	-2,8749	0,0040
X_5	0,0286	0,0073	3,8970	<0,0001

Pseudo R-squared: 0,2357

Sehingga model regresi yang dapat dibentuk berdasarkan hasil pada tabel dapat dituliskan sebagai berikut.

$$\log\left(\frac{\mu}{1-\mu}\right) = -7,3843 - 0,0234X_1 + 0,0286X_5$$

Berdasarkan hasil yang diperoleh dari pengujian secara parsial, parameter yang signifikan yaitu parameter β_1 dan β_5 dan tanpa melibatkan pengamatan ke-33, maka kedua parameter tersebut dapat diinterpretasikan. Untuk parameter β_0 yaitu sebesar $\exp(-7,3843)$ dapat diinterpretasikan bahwa apabila persentase K4 (X_1) dan persentase KFLengkap (X_5) bernilai konstan (nol), maka AKI dari 33 Provinsi di Indonesia pada tahun 2022 berkurang sebesar $\exp(-7,2517)$.

Koefisien regresi β_1 sebesar $\exp(-0,0234)$ menyatakan terdapat pengaruh negatif dari persentase K4 terhadap angka kematian ibu. Dengan kata lain jika persentase K4 (X_1) meningkat sebesar satu-satuan, maka AKI akan berkurang sebesar $\exp(-0,0234)$ dengan diasumsikan bahwa persentase KFLengkap (X_5) bernilai konstan (nol). Sedangkan koefisien regresi β_5 sebesar $\exp(0,0286)$ menyatakan terdapat pengaruh positif dari persentase KFLengkap terhadap AKI. Jika persentase KFLengkap meningkat sebesar satu-satuan, maka AKI akan meningkat sebesar $\exp(0,0286)$ dengan diasumsikan bahwa persentase K4 (X_1) bernilai konstan (nol).

Selanjutnya setelah didapatkan model yang layak, penulis ingin melihat AKI di Indonesia pada tahun 2022. Berikut ini merupakan penjabaran untuk mengestimasi AKI di setiap Provinsi di Indonesia pada tahun 2022.

$$\mu_i = \frac{\exp(\beta_0 + \beta_1 X_1 + \beta_5 X_5)}{1 + \exp(\beta_0 + \beta_1 X_1 + \beta_5 X_5)} \Rightarrow \mu_i = \frac{\exp(-7,3843 - 0,0234X_1 + 0,0286X_5)}{1 + \exp(-7,3843 - 0,0234X_1 + 0,0286X_5)} \text{ dengan } i = 1, 2, \dots, 33$$

Didapatkan hasil untuk AKI terendah terdapat pada Provinsi Aceh dengan AKI sebesar 0,0004. Sedangkan AKI tertinggi terdapat pada Provinsi Sulawesi Barat dengan AKI sebesar 0,0013.

D. Kesimpulan

Secara umum dalam penelitian ini telah ditunjukkan mengenai hubungan antara variabel prediktor dan variabel respon yang berbentuk proporsi, yang dimodelkan dengan model regresi beta, dengan studi kasus lebih lanjut untuk mengidentifikasi data berpengaruh menggunakan statistik Cook's Distance. Pada kasus penelitian ini terdapat satu pengamatan yang teridentifikasi merupakan data berpengaruh yaitu pengamatan ke-33 (Provinsi Papua Barat), sehingga pengamatan ini disisihkan dari analisis. Ketika pengamatan tersebut tidak dilibatkan dalam analisis, model menjadi lebih baik dan dapat dilanjutkan untuk menganalisis menggunakan regresi beta. Selanjutnya didapatkan faktor yang berpengaruh secara signifikan terhadap AKI di Indonesia pada tahun 2022 yaitu faktor persentase pelayanan K4 (X_1) dan persentase pelayanan KF Lengkap (X_5).

Nilai peluang dari model regresi beta dimana hanya melibatkan variabel X_1 dan X_5 saja tanpa melibatkan pengamatan ke-33, Provinsi yang memiliki AKI tertinggi yaitu Provinsi

Sulawesi Barat. Provinsi ini memiliki nilai dari variabel X_1 sebesar 76,8194% dan variabel X_5 sebesar 89,6920% dengan AKI sebesar 0,0013 atau 1 sampai 2 kematian per 1000 kelahiran hidup. Sedangkan untuk Provinsi dengan AKI terendah terdapat pada Provinsi Aceh. Provinsi ini memiliki nilai variabel X_1 sebesar 82,3155% dan nilai variabel X_5 sebesar 55,3116% dengan AKI sebesar 0,0001 atau 0 sampai 1 kematian per 1000 kelahiran hidup.

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran dari penulis dalam penelitian ini, yaitu:

1. Pada penelitian dengan variabel prediktor yang berbentuk proporsi dimana ruang sampelnya berukuran antara 0 sampai dengan 1 untuk dapat menggunakan pemodelan regresi dengan distribusi beta.
2. Bagi penelitian selanjutnya dengan kasus Angka Kematian Ibu, perlu dilakukan pengkajian lebih lanjut mengenai hasil dari penelitian ini, dimana variabel Pelayanan Kesehatan Ibu Nifas KFLengkap (X_5) yang bertolak belakang dengan harapan dan logika.
3. Bagi penelitian selanjutnya dengan kasus yang sama mengenai angka kematian ibu, sebaiknya menambah ataupun menggunakan variabel lainnya yang mungkin bisa menjadi faktor yang diduga berpengaruh terhadap kasus angka kematian ibu sehingga didapatkan model yang lebih baik lagi.
4. Bagi penelitian selanjutnya dapat menambah analisis untuk mengidentifikasi data berpengaruh selain menggunakan statistik Cook's Distance, sehingga mendapatkan hasil yang lebih baik lagi.
5. Bagi penelitian selanjutnya dapat menggunakan metode yang lebih tepat, dimana metode tersebut tidak terpengaruh dengan adanya data berpengaruh didalam data.

Acknowledge

Puji dan syukur penulis panjatkan kehadirat Allah SWT karena berkat rahmat dan karunianya penulis dapat menyelesaikan penelitian ini. Tidak lupa sholawat serta salan senantiasa tercurah limpah kepada Nabi Muhammad SAW. Terimakasih kepada alm. Ayah, ibu, kakak, adik, dan orang terdekat penulis yang senantiasa memberikan dukungan baik materi maupun moral kepada penulis. Terimakasih kepada Bapak Dr. Nusar Hajarisman, M.Si yang telah membimbing dan meotivasi penulis hingga penelitian ini dapat diselesaikan. Terimakasih kepada sahabat penulis, serta semua pihak yang terlibat dalam penelitian ini.

Daftar Pustaka

- [1] N. Hajarisman, "Penaksiran Parameter Model Regresi Beta untuk Memodelkan Data Proporsi," *Jurnal Statistika*, vol. 12, no. 1, pp. 9–18, Mei 2012.
- [2] N. Hajarisman, "Pemeriksaan Data Berpengaruh Dalam Model Regresi Gamma," *Media Statistika*, vol. 3, no. 2, Desember 2010.
- [3] Kementerian Kesehatan R.I., *Profil Kesehatan Indonesia Tahun 2022*, Jakarta: Kementerian Kesehatan R.I., 2023.
- [4] A. Agresti, *An Introduction To Categorical Data Analysis*, 2nd ed., New Jersey: John Wiley and Sons, 2007.
- [5] S. L. P. Ferrari and F. Cribari-Neto, "Beta Regression for Modeling Rates and Proportions," *Journal of Applied Statistics*, 2004.
- [6] D. Collet, *Model Binary Data*, 2nd ed., London: Chapman and Hall, 2003.
- [7] F. Cribari-Neto and A. Zeileis, "Beta Regression in R," *Journal of Statistical Software*, 2010.
- [8] D. A. F. Hilman and A. K. Mutaqin, "Penerapan Regresi Double Poisson untuk Memprediksi Pertandingan dan Klasemen Liga 1 Indonesia," *Jurnal Riset Statistika*, vol. 3, no. 2, pp. 97–106, Dec. 2023, doi: 10.29313/jrs.v3i2.2784.
- [9] S. Pratiwi and M. Herlina, "Pengaruh Harga Pangan terhadap Inflasi dengan Metode Vector Autoregressive Integrated Moving Average," *Jurnal Riset Statistika*, vol. 3, no. 2, pp. 87–96, Dec. 2023, doi: 10.29313/jrs.v3i2.2690.

- [10] [10] D. Oktoriandi, "Penerapan Uji Q Cochran terhadap Atribut Produk Laptop Menggunakan Multiple Response Analysis (MRA)," *Jurnal Riset Statistika*, vol. 1, no. 2, pp. 127–134, Feb. 2022, doi: 10.29313/jrs.v1i2.521.