

## Regresi Binomial Negatif untuk Memodelkan Kematian Bayi di Kalimantan Timur

### *Negative Binomial Regression for Modeling of Infant Mortality in East Kalimantan*

M. Fathurahman

Program Studi Statistika, Jurusan Matematika FMIPA Universitas Mulawarman

E-mail: fathur@fmipa.unmul.ac.id

#### ABSTRACT

*Negative Binomial Regression (NBR) is an alternative regression model to model the relationship between the dependent variable in overdispersion count data and one or more independent variables. Overdispersion is a problem in Poisson regression modeling. Namely, the variance of the dependent variable is more than the mean. If there is overdispersion, then the parameter estimator of the Poisson regression model has a standard error value that is not under-estimated. The NBR model was applied to modeling infant mortality in East Kalimantan in 2019. Data on infant mortality in East Kalimantan in 2019 indicated overdispersion. Infant mortality is an indicator that can measure the progress of development outcomes in the health sector in a region. In the last three years, from 2017 to 2019, infant mortality data in East Kalimantan has increased. Therefore, it is necessary to do modeling to get the factors that cause it. The modeling results with NBR show that the percentage of the complete neonatal visit of KN3, the percentage of infant health services, and the percentage of visits by pregnant women K4 significantly affect infant mortality in East Kalimantan in 2019.*

**Keywords:** Count Data, Poisson Regression, Overdispersion, NBR, Infant Mortality

#### Pendahuluan

Regresi Poisson adalah suatu pendekatan regresi yang populer untuk memodelkan data cacah. Model regresi Poisson dikembangkan melalui *Generalized Linear Models* (GLM) (Agresti, 2015). Dalam pemodelan regresi Poisson, asumsi ekuidispersi harus dipenuhi. Asumsi ini menunjukkan bahwa ragam dan rerata dari variabel dependen mempunyai nilai yang sama. Pada penerapannya, asumsi ekuidispersi terkadang tidak dapat dipenuhi karena adanya overdispersi, yaitu ragam lebih dari rerata. Jika variabel dependen terindikasi overdispersi tetap dimodelkan dengan regresi Poisson, maka akan didapatkan *estimator* parameter model yang memiliki nilai *standard error* yang bukan nilai sebenarnya (*under-estimate*). Akibatnya, hasil uji signifikansi terhadap parameter model menghasilkan variabel independen memengaruhi variabel dependen padahal kenyataannya tidak berpengaruh signifikan (Hilbe, 2014). Oleh karena itu, diperlukan suatu pendekatan untuk menangani overdispersi.

RBN merupakan suatu pendekatan yang dapat menangani overdispersi pada variabel dependen model regresi Poisson. Model RBN dibangun berdasarkan campuran (*mixed*) antara distribusi Poisson dan distribusi gamma (Hilbe, 2014). Penelitian yang mengembangkan dan menerapkan model RBN telah banyak dilakukan. Afri, dkk (2012) dan Gomes, dkk (2017) mengembangkan model RBN dan model RBN untuk data spasial heterogenitas dan menerapkannya pada pemodelan kematian bayi dan kecelakaan

transportasi. Estimasi terhadap model RBN menggunakan metode *Maximum Likelihood* (ML) dan Newton-Raphson. Sedangkan, pengujian hipotesis signifikansi parameter model menggunakan metode *Likelihood Ratio Test* (LRT) dan Wald. Sementara itu, Dai, dkk (2013) melakukan estimasi terhadap parameter dispersi pada distribusi binomial negatif yang merupakan distribusi dari variabel dependen model RBN. Ardiles, dkk (2018) mengembangkan dan menerapkan RBN pada pemodelan polusi udara. Wardani dan Wulandari (2020) melakukan pemodelan kematian bayi menggunakan RBN dan Islam, dkk (2020) menerapkan RBN untuk memodelkan faktor-faktor yang berhubungan dan berpengaruh terhadap cakupan pelayanan *Antenatal Care* (ANC). Iqbal, dkk (2021) dan Ibarra-Espinosa, dkk (2022) menerapkan RBN pada pemodelan kasus COVID-19.

Penerapan model RBN pada penelitian ini mengacu pada hasil publikasi Dinas Kesehatan Provinsi Kalimantan Timur (2020) yang menyatakan bahwa terjadi peningkatan jumlah kasus kematian bayi di Kalimantan Timur dalam tiga tahun terakhir, yaitu dari tahun 2017 sampai dengan 2019. Pada tahun 2017 jumlah kasus kematian bayi sebanyak 605 kasus. Terjadi peningkatan jumlah kasus kematian bayi di tahun 2018 sebanyak 671 kasus dan kembali meningkat sebanyak 688 kasus pada tahun 2019.

**Tinjauan Pustaka**

**Regresi Poisson**

Regresi Poisson merupakan suatu model regresi parametrik yang nonlinier. Variabel dependen ( $Y$ ) model regresi Poisson berdistribusi Poisson dengan fungsi massa probabilitas didefinisikan sebagai berikut (Hilbe, 2014):

$$P(Y = y|\mu_1) = \frac{e^{-\mu_1} \mu_1^y}{y!}, y = 0, 1, 2, \dots \quad (1)$$

dimana  $\mu$  adalah parameter dan  $\mu > 0$ . Rerata dan ragam dari  $Y$  berturut-turut dinyatakan dengan  $E(Y) = \mu_1$  dan  $Var(Y) = \mu_1$ .

Secara matematis, model regresi Poisson diformulasikan sebagai berikut (Hilbe, 2014):

$$\mu_{1i} = \exp(\beta_1^T x_i), i = 1, 2, \dots, n \quad (2)$$

dimana  $\mu_{1i}$  adalah rerata variabel dependen untuk observasi ke- $i$ ,  $\beta_1^T$  adalah vektor parameter, dan  $x_i$  vektor variabel independen untuk observasi ke- $i$ . Vektor parameter dan variabel independen berturut-turut dinyatakan sebagai

$$\beta_1^T = [\beta_{10} \ \beta_{11} \ \beta_{12} \ \dots \ \beta_{1q}]$$

$$x_i = [1 \ X_{1i} \ X_{2i} \ \dots \ X_{qi}]^T.$$

Berdasarkan metode GLM [1], model regresi Poisson pada Persamaan (2) dapat ditulis menjadi

$$\eta_1(x_i) = \ln \mu_{1i} = \beta_1^T x_i. \quad (3)$$

dimana  $\eta_1(x_i)$  adalah fungsi penghubung (*link function*) yang bergantung pada variabel independen.

Untuk mendapatkan model regresi Poisson pada Persamaan (4), dilakukan estimasi terhadap parameter modelnya menggunakan metode ML (Hilbe, 2014). Estimasi parameter diawali dengan membentuk fungsi *likelihood* dan *ln-likelihood*. Misalkan  $\mathcal{L}_1 = \mathcal{L}(\beta_1)$  dan  $\ell_1 = \ell(\beta_1)$ . Berdasarkan Persamaan (1), dapat dibentuk fungsi *likelihood* dan *ln-likelihood* sebagai berikut:

$$\mathcal{L}_1 = \prod_{i=1}^n \frac{\exp[-\exp(\beta_1^T x_i)] [\exp(\beta_1^T x_i)]^{y_i}}{y_i!} \quad (4)$$

$$\ell_1 = \sum_{i=1}^n [y_i \beta_1^T x_i - \exp(\beta_1^T x_i) - \ln y_i!]. \quad (5)$$

Selanjutnya, memaksimalkan fungsi *ln-likelihood* pada Persamaan (5) dengan melakukan derivatif parsial terhadap parameter yang diestimasi, yaitu

$$\frac{\partial \ell_1}{\partial \beta_1^T} = - \sum_{i=1}^n [y_i - \exp(\beta_1^T x_i)] x_i = \mathbf{0}. \quad (6)$$

*Estimator* ML parameter model regresi Poisson pada Persamaan (6) berbentuk fungsi yang tidak eksplisit. Oleh karena itu, untuk mendapatkan *estimator* ML digunakan pendekatan numerik dengan metode *Fisher scoring* (Fathurahman, 2019). Setelah didapatkan *estimator* parameter model, dilakukan uji signifikansi parameter yang bertujuan untuk mendapatkan variabel independen yang memengaruhi variabel dependen. Pengujian signifikansi parameter meliputi uji simultan dan uji parsial (Hilbe, 2014).

Uji simultan digunakan untuk mengetahui adanya pengaruh dari variabel independen terhadap variabel dependen secara bersama atau untuk mengetahui minimal terdapat satu variabel independen yang memengaruhi variabel dependen. Uji simultan juga dapat digunakan untuk mengetahui layak atau tidaknya model yang digunakan. Sedangkan, uji parsial digunakan untuk mengetahui variabel independen yang memengaruhi variabel dependen secara individu. Hipotesis untuk uji simultan adalah:

$$H_0 : \beta_{11} = \beta_{12} = \dots = \beta_{1q} = 0$$

$$H_1 : \text{minimal satu } \beta_{1h} \neq 0, \quad h = 1, 2, \dots, q. \quad (7)$$

Statistik uji yang digunakan untuk pengujian hipotesis pada Persamaan (7) adalah statistik *Wilk's lambda* yang diperoleh dengan metode LRT dan didefinisikan sebagai berikut (Hilbe, 2014):

$$G_1 = 2 \left( L(\hat{\Omega}_1) - L(\hat{\omega}_1) \right), \quad (8)$$

dimana:

$$L(\hat{\Omega}_1) = \sum_{i=1}^n [-\exp(\hat{\beta}_1^T x_i) + y_i \hat{\beta}_1^T x_i - \ln(y_i!)]$$

$$L(\hat{\omega}_1) = \sum_{i=1}^n [-\exp(\hat{\beta}_0^T x_i) + y_i \hat{\beta}_0^T x_i - \ln(y_i!).]$$

Statistik uji *Wilk's lambda* pada Persamaan (8) secara asimtotik menghampiri distribusi *Khi-kuadrat* (Pawitan, 2001), dengan derajat bebas  $v_1$  yang didapat dari selisih antara banyaknya parameter model dibawah populasi dan hipotesis nol ( $H_0$ ), yaitu  $v_1 = (q + 1) - 1 = q$ . Oleh karena itu, daerah penolakan hipotesis nol untuk menguji hipotesis pada Persamaan (7) adalah tolak  $H_0$  bila nilai  $G_1 > \chi_{(\alpha, v_1)}^2$  atau *P-value* kurang dari  $\alpha$ , dengan  $\alpha$  adalah tingkat signifikansi dan nilai  $\chi_{(\alpha, v_1)}^2$  didapat dari tabel distribusi *Khi-kuadrat*.

Selanjutnya, melakukan uji parsial menggunakan *Wald test*. Hipotesis untuk uji parsial adalah:

$$\begin{aligned} H_0 : \beta_{1h} &= 0 \\ H_1 : \beta_{1h} &\neq 0, h = 1, 2, \dots, q. \end{aligned} \quad (9)$$

Statistik uji untuk menguji hipotesis pada Persamaan (9) diformulasikan sebagai berikut (Hilbe, 2014):

$$W_1 = \frac{\hat{\beta}_{1h}}{SE(\hat{\beta}_{1h})}, \quad (10)$$

dimana  $\hat{\beta}_{1h}$  adalah *estimator* ML parameter yang diperoleh dengan metode *Fisher-scoring* dan  $SE(\hat{\beta}_{1h}) = \sqrt{\widehat{var}(\hat{\beta}_{1h})}$  adalah *standard error* untuk *estimator* ML parameter yang diperoleh dari elemen-elemen diagonal utama matriks ragam koragam,  $Cov(\hat{\beta}_1) = I^{-1}(\hat{\beta}_1)$  dengan  $I(\hat{\beta}_1)$  adalah matriks informasi Fisher.

Statistik uji Wald pada Persamaan (10) secara asimtotik menghampiri distribusi normal standar (Pawitan, 2001). Oleh karena itu, daerah kritis untuk menguji hipotesis pada Persamaan (9) adalah tolak  $H_0$  bila nilai  $|W_1| > Z_{\alpha/2}$  atau *P-value* kurang dari  $\alpha$ , dengan nilai  $Z_{\alpha/2}$  didapat dari tabel distribusi normal standar.

### Regresi Binomial Negatif

Regresi Binomial Negatif (RBN) merupakan model regresi yang dibangun menggunakan perpaduan (*mixed*) antara distribusi Poisson dan distribusi gamma (Hilbe, 2011). Misalkan  $Y$  adalah variabel dependen yang merupakan variabel *random* berdistribusi binomial negatif, maka fungsi distribusi probabilitasnya didefinisikan sebagai berikut (Hilbe, 2011):

$$P(Y = y | \mu_2, \tau) = \frac{\Gamma(y + \tau^{-1})}{\Gamma(\tau^{-1})\Gamma(y + 1)} \Delta_1 \quad (11)$$

dimana

$$\Delta_1 = \left( \frac{\tau^{-1}}{\tau^{-1} + \mu_2} \right)^{\zeta^{-1}} \left( \frac{\mu_2}{\tau^{-1} + \mu_2} \right)^y,$$

untuk  $y = 0, 1, 2, \dots$  dan  $\tau \geq 0$ .  $\mu_2 = E(Y)$  adalah rerata dari  $Y$  dan  $\tau$  adalah parameter dispersi.

Model RBN didefinisikan dalam formula sebagai berikut (Hilbe, 2011):

$$\eta_2(x_i) = \ln \mu_{2i} = \beta_2^T x_i \quad (12)$$

dimana  $\eta_2(x_i)$  adalah fungsi penghubung yang bergantung pada vektor variabel independen untuk observasi ke- $i$ ,  $x_i$ ;  $\mu_{2i}$  adalah rerata dari variabel dependen untuk observasi ke- $i$ ; dan  $\beta_2^T$

adalah vektor parameter yang dinyatakan dengan  $\beta_2^T = [\beta_{20} \ \beta_{21} \ \beta_{22} \ \dots \ \beta_{2q}]$ .

Model RBN pada Persamaan (12) dapat diperoleh dengan cara mengestimasi parameter model menggunakan metode ML (Hilbe, 2011). *Estimator* ML diperoleh dengan memaksimalkan fungsi *likelihood* atau fungsi *ln-likelihood* yang diformulasikan sebagai berikut:

$$\mathcal{L}_2 = \prod_{i=1}^n \left[ \left( \prod_{j=1}^{y_i-1} j/\tau \right) \frac{1}{(y_i!)} \Delta_{2i} \right] \quad (12)$$

dimana  $\mathcal{L}_2 = \mathcal{L}(\beta_2)$ ,

$$\Delta_{2i} = \left( \frac{1}{1 + \tau\mu_{2i}} \right)^{1/\tau} \left( \frac{\tau\mu_{2i}}{1 + \tau\mu_{2i}} \right)^{y_i}.$$

$$\ell_2 = \sum_{i=1}^n \left[ \left( \sum_{j=1}^{y_i-1} \ln(j + 1/\tau) \right) + \Delta_{3i} \right] \quad (13)$$

dimana  $\ell_2 = \ell(\beta_2)$ ,

$$\begin{aligned} \Delta_{3i} = & -\ln(y_i!) + y_i \ln(\tau\mu_{2i}) + \\ & -(y_i + 1/\tau) \ln(1 + \tau\mu_{2i}). \end{aligned}$$

Derivatif parsial fungsi *ln-likelihood* pada Persamaan (13) terhadap parameter yang diestimasi menghasilkan *estimator* ML yang berbentuk tidak eksplisit seperti pada Persamaan (14).

$$\frac{\partial \ell_2}{\partial \beta_2^T} = \sum_{i=1}^n \left( \frac{\mu_{2i}}{1 + \tau\mu_{2i}} \frac{(y_i - \mu_{2i})x_i}{\mu_{2i}} \right) = 0, \quad (14)$$

dengan  $\mu_{2i} = \exp(\beta_2^T x_i)$ .

Karena *estimator* ML parameter model RBN berbentuk fungsi yang tidak eksplisit, maka untuk mendapatkan *estimator* ML membutuhkan pendekatan secara numerik. Salah satu pendekatan numerik yang dapat digunakan untuk mendapatkan *estimator* ML adalah metode *Fisher-scoring* (Fathurahman, 2019).

Selanjutnya, dilakukan pengujian hipotesis parameter model RBN yang terdiri atas uji simultan dan uji parsial. Uji simultan dilakukan dengan metode LRT dan uji parsial dengan uji Wald. Hipotesis untuk uji simultan adalah:

$$\begin{aligned} H_0 : \beta_{21} = \beta_{22} = \dots = \beta_{2q} &= 0 \\ H_1 : \text{minimal satu } \beta_{2h} &\neq 0, \\ &h = 1, 2, \dots, q. \end{aligned} \quad (15)$$

Statistik uji untuk hipotesis pada Persamaan (15) adalah statistik uji *Wilk's lambda* yang diformulasikan sebagai berikut (Hilbe, 2011):

$$G_2 = 2 \left( L(\hat{\Omega}_2) - L(\hat{\omega}_2) \right), \quad (16)$$

dimana:

$$L(\hat{\Omega}_2) = \sum_{i=1}^n [\Delta_{4i} + \Delta_{5i}]$$

dengan

$$\Delta_{4i} = \left( \sum_{j=1}^{y_i-1} \ln(j + 1/\hat{\tau}) \right)$$

$$\Delta_{5i} = -\ln(y_i!) + y_i \ln(\hat{\tau}\hat{\mu}_{2i}) +$$

$$-(y_i + 1/\hat{\tau}) \ln(1 + \hat{\tau}\hat{\mu}_{2i}).$$

$$L(\hat{\omega}_2) = \sum_{i=1}^n [\Delta_{4i} + \Delta_{6i}]$$

dengan

$$\Delta_{6i} = -\ln(y_i!) + y_i \ln(\hat{\tau}\hat{\mu}_{3i}) +$$

$$-(y_i + 1/\hat{\tau}) \ln(1 + \hat{\tau}\hat{\mu}_{3i})$$

dan  $\hat{\mu}_{3i} = \exp(\hat{\beta}_{20})$ .

Statistik uji *Wilk's lambda* pada Persamaan (16) secara asimtotik menghampiri distribusi *Khi-kuadrat* (Pawitan, 2001), dengan derajat bebas  $v_2$  yang didapat dari selisih antara banyaknya parameter model dibawah populasi dan hipotesis nol, yaitu  $v_2 = (q + 2) - 2 = q$ . Daerah penolakan hipotesis nol untuk menguji hipotesis pada Persamaan (15) adalah tolak  $H_0$  bila nilai  $G_2 > \chi^2_{(\alpha, v_2)}$  atau *P-value* kurang dari  $\alpha$ . Nilai  $\chi^2_{(\alpha, v_2)}$  didapat dari tabel distribusi *Khi-kuadrat*.

Setelah uji simultan, dilakukan uji parsial menggunakan hipotesis sebagai berikut:

$$H_0 : \beta_{2h} = 0$$

$$H_1 : \beta_{2h} \neq 0, h = 1, 2, \dots, q. \quad (17)$$

Statistik uji untuk menguji hipotesis pada Persamaan (17) adalah statistik Wald yang dinyatakan sebagai berikut (Hilbe, 2014):

$$W_2 = \frac{\hat{\beta}_{2h}}{SE(\hat{\beta}_{2h})}, \quad (18)$$

dimana  $\hat{\beta}_{2h}$  adalah *estimator* ML parameter model yang diperoleh dengan metode *Fisher scoring* dan  $SE(\hat{\beta}_{2h}) = \sqrt{\widehat{var}(\hat{\beta}_{2h})}$  adalah *standard error estimator* ML parameter yang diperoleh dari elemen-elemen diagonal utama matriks ragam koragam,  $Cov(\hat{\beta}_2) = I^{-1}(\hat{\beta}_2)$  dengan  $I(\hat{\beta}_2)$  adalah matriks informasi Fisher.

Statistik Wald pada Persamaan (18) secara asimtotik menghampiri distribusi normal standar (Pawitan. 2001). Daerah kritis untuk menguji hipotesis pada Persamaan (17) adalah tolak  $H_0$

bila nilai  $|W_2| > Z_{\alpha/2}$  atau *P-value* kurang dari  $\alpha$  dan nilai  $Z_{\alpha/2}$  didapat dari tabel distribusi normal standar.

### Multikolinieritas

Multikolinieritas merupakan suatu pelanggaran terhadap asumsi dalam pemodelan regresi, yaitu terjadinya korelasi diantara variabel independen. Jika terjadi multikolinieritas dalam model regres, maka mengakibatkan *standard error* estimasi parameter model menjadi sangat besar dan hasil pengujian signifikansi parameter menghasilkan parameter yang tidak berpengaruh signifikan terhadap model (Gujarati, 2003). Untuk mendeteksi adanya multikolinieritas dalam model regresi dapat menggunakan metode *Variance Inflation Factor* (VIF) dengan formula sebagai berikut:

$$VIF_h = \frac{1}{1 - R_h^2}, h = 1, 2, \dots, q \quad (19)$$

dimana  $R_h^2$  adalah nilai koefisien determinasi antara variabel independen ke- $h$  dan variabel independen lainnya,  $h$  adalah indeks untuk variabel independen dan  $q$  adalah banyaknya variabel independen. Suatu model regresi bebas multikolinieritas bila memiliki nilai VIF untuk semua variabel independennya kurang dari 10 (Gujarati, 2003).

### Kematian Bayi

Kematian bayi adalah kematian yang terjadi antara saat setelah bayi lahir sampai bayi belum berusia tepat satu tahun (Amiruddin dan Hasmi, 2013). Kematian bayi dapat dibagi dalam kematian neonatal (28 hari pertama) dan kematian pascaneonatal (antara 28 – 365 hari) (Rachmawati, 2014). Secara garis besar, dari sisi penyebabnya kematian bayi terdapat dua jenis, yaitu endogen dan eksogen. Kematian bayi endogen (kematian neonatal) merupakan kematian bayi yang terjadi pada bulan pertama setelah dilahirkan, dan penyebabnya secara umum adalah faktor-faktor yang dibawa bayi sejak lahir, yang diturunkan dari orang tuanya selama masa kehamilan. Sedangkan, kematian bayi eksogen (kematian post neonatal) merupakan kematian bayi yang terjadi setelah usia bayi 1 bulan hingga kurang dari 1 tahun, dengan penyebabnya adalah faktor-faktor yang berkaitan dengan lingkungan eksternal (Tarigan, dkk, 2017).

### Metode Penelitian

#### Sumber Data dan Variabel Penelitian

Data pada penelitian ini merupakan data sekunder yang didapat dari Dinas Kesehatan Provinsi Kalimantan Timur. Sedangkan, variabel penelitian terdiri atas variabel dependen dan

variabel independen. Variabel dependen adalah kematian bayi ( $Y$ ). Sedangkan, variabel independen meliputi persentase bayi Berat Badan Lahir Rendah (BBLR) ( $X_1$ ), persentase bayi yang mendapatkan Air Susu Ibu (ASI) secara eksklusif ( $X_2$ ), persentase cakupan kunjungan neonatal lengkap KN3 ( $X_3$ ), persentase cakupan pelayanan kesehatan bayi ( $X_4$ ), dan persentase cakupan kunjungan ibu hamil K4 ( $X_5$ ).

**Populasi, Sampel, dan Teknik Sampling**

Populasi pada penelitian ini adalah bayi yang meninggal pada usia 0 sampai dengan 11 bulan di Kalimantan Timur. Sampel yang digunakan adalah bayi yang meninggal pada usia 0 sampai dengan 11 bulan di Kalimantan Timur pada tahun 2019. Teknik pengambilan sampel (*sampling*) menggunakan *purposive sampling*, yaitu pengambilan sampel dengan pertimbangan. Pertimbangan peneliti dalam mengambil sampel adalah kematian bayi di Kalimantan Timur dalam tiga tahun terakhir dari tahun 2017 sampai dengan 2019 terus mengalami peningkatan.

**Teknik Analisis Data**

Teknik analisis data dalam penelitian ini adalah sebagai berikut:

1. Melakukan analisis statistik deskriptif terhadap data penelitian.
2. Mendeteksi multikolinieritas pada variabel independen menggunakan nilai VIF.
3. Memodelkan kematian bayi dengan regresi Poisson yang meliputi:
  - a. Melakukan estimasi parameter model menggunakan metode ML dan *Fisher-scoring*.
  - b. Melakukan uji signifikansi parameter model menggunakan metode LRT dan Wald.
  - c. Mendeteksi overdispersi pada regresi Poisson.
4. Memodelkan kematian bayi dengan RBN untuk menangani overdispersi pada regresi Poisson yang terdiri atas:
  - a. Melakukan estimasi parameter model dengan metode ML dan *Fisher-scoring*.
  - b. Melakukan uji signifikansi parameter model menggunakan metode LRT dan Wald.
  - c. Melakukan interpretasi terhadap model RBN.
5. Menarik kesimpulan.

**Hasil dan Pembahasan**  
**Deskripsi Data Penelitian**

Untuk mengetahui gambaran atau deskripsi dari kematian bayi di Provinsi Kalimantan Timur tahun 2019, dilakukan analisis statistik deskriptif terhadap data penelitian yang terdapat pada

variabel dependen dan independen. Hasil yang diperoleh disajikan pada Tabel 1.

**Tabel 1.** Deskripsi Data Penelitian

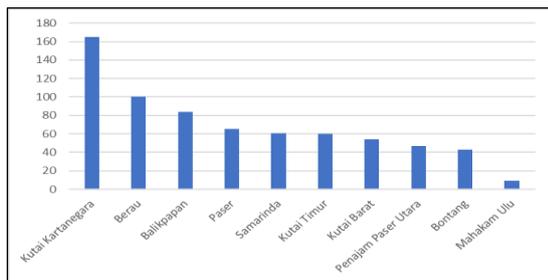
Variabel	Min	Mak	Rerata	SB
$Y$	9	165	69	42
$X_1$	2,7	10,3	6,11	2,8
$X_2$	66,5	96,8	78,97	9,3
$X_3$	76,3	100	92,09	7,31
$X_4$	72,7	100	90,69	10,48
$X_5$	69,4	99,5	88,92	9,12

Min = Minimum, Mak = Maksimum

SB = Simpangan Baku.

Pada Tabel 1 tampak bahwa rerata kematian bayi di Kalimantan Timur tahun 2019 adalah sebanyak 69 kasus, dengan simpangan baku sebanyak 42 kasus. Kematian bayi tertinggi di Kabupaten Kutai Kartanegara, yaitu sebanyak 165 kasus dan terendah di Kabupaten Mahakam Ulu, yaitu sebanyak 9 kasus. Perkembangan kematian bayi di Kalimantan Timur tahun 2019 disajikan pada Gambar 1.

Sementara itu, rerata persentase bayi BBLR sebesar 6,11 persen, dengan simpangan baku sebesar 2,8 persen. Persentase bayi BBLR tertinggi di Kabupaten Kutai Barat sebesar 10,3 persen dan terendah di Kota Samarinda sebesar 2,7 persen. Rerata persentase bayi yang mendapatkan ASI eksklusif sebesar 78,97 persen, dengan simpangan baku sebesar 9,3 persen. Persentase bayi yang mendapatkan ASI eksklusif tertinggi di Kota Bontang, yaitu sebesar 96,8 persen. Sedangkan, terendah di Kabupaten Kutai Barat, yaitu sebesar 66,5 persen. Rerata cakupan kunjungan neonatal lengkap KN3 sebesar 92,09 persen, dengan simpangan baku sebesar 7,31 persen. Cakupan kunjungan neonatal lengkap KN3 tertinggi di Kabupaten Penajam Paser Utara sebesar 100 persen dan terendah di Kota Bontang sebesar 76,3 persen. Rerata cakupan pelayanan kesehatan bayi sebesar 90,69 persen, dengan simpangan baku sebesar 10,48 persen. Cakupan pelayanan kesehatan bayi tertinggi di Kabupaten Penajam Paser Utara, yaitu sebesar 100 persen. Sedangkan, terendah di Kabupaten Mahakam Ulu, yaitu sebesar 72,7 persen. Rerata cakupan kunjungan ibu hamil K4 sebesar 88,92 persen, dengan simpangan baku sebesar 9,12 persen. Cakupan kunjungan ibu hamil K4 tertinggi di Kabupaten Penajam Paser Utara sebesar 99,5 persen dan terendah di Kabupaten Berau sebesar 69,4 persen.



**Gambar 1.** Diagram perkembangan kematian bayi di Kalimantan Timur tahun 2019

**Pendeteksian Multikolinieritas**

Pendeteksian multikolinieritas terhadap variabel independen menggunakan kriteria VIF. Hasil yang didapat seperti pada Tabel 2.

**Tabel 2.** Nilai VIF Variabel Independen

Variabel	VIF
$X_1$	1,32
$X_2$	2,33
$X_3$	1,66
$X_4$	2,07
$X_5$	1,72

Berdasarkan Tabel 2, nilai VIF dari masing-masing variabel independen tidak lebih dari 10. Hasil ini mengindikasikan tidak terdapat multikolinieritas pada variabel independen. Oleh karena itu, semua variabel independen layak untuk dimodelkan dengan regresi Poisson dan RBN.

**Pemodelan Kematian Bayi dengan Regresi Poisson**

Pemodelan kematian bayi di Kalimantan Timur tahun 2019 dengan regresi Poisson diawali dengan melakukan estimasi dan uji signifikansi parameter. Hasil yang diperoleh ditampilkan pada Tabel 3.

**Tabel 3.** Hasil Estimasi dan Uji Parsial Parameter Model Regresi Poisson

Par	Est	SE	$W_1$	P
$\hat{\beta}_{10}$	8,5044	1,1004	7,7285	< 0,001*
$\hat{\beta}_{11}$	-0,0368	0,0157	-2,3439	0,0256*
$\hat{\beta}_{12}$	-0,0173	0,0066	-2,6212	0,0129*
$\hat{\beta}_{13}$	-0,0348	0,0074	-4,7027	< 0,001*
$\hat{\beta}_{14}$	0,0379	0,0063	6,0159	< 0,001*
$\hat{\beta}_{15}$	-0,0339	0,0065	-5,2154	< 0,001*

\*Signifikan pada  $\alpha = 0,1$ .

Berdasarkan Tabel 3, didapatkan model regresi Poisson yang dinyatakan sebagai berikut:

$$\eta_1(x_i) = 8,5044 - 0,0368X_{1i} - 0,0173X_{2i} + -0,0348X_{3i} + 0,0379X_{4i} - 0,0339X_{5i} \tag{20}$$

Selanjutnya, dilakukan uji signifikansi parameter untuk mendapatkan parameter yang signifikan terhadap model. Hasil uji signifikansi parameter secara serentak melalui uji simultan diperoleh nilai statistik uji *Wilk's lambda* ( $G_1$ ) sebesar 111,75 dan *P-value* sebesar  $1,75 \times 10^{-22}$ . Sementara itu, dari tabel distribusi *Khi-kuadrat* didapatkan nilai  $\chi^2_{(0,1;5)}$  sebesar 9,24. Karena nilai  $G_1$  lebih dari  $\chi^2_{(0,1;5)}$  dan nilai *P* tidak lebih dari  $\alpha$ , maka tolak  $H_0$ . Sehingga, dapat disimpulkan bahwa variabel persentase bayi BBLR, persentase bayi yang mendapatkan ASI secara eksklusif, persentase cakupan kunjungan neonatal lengkap KN3, persentase cakupan pelayanan kesehatan bayi, dan persentase cakupan kunjungan ibu hamil K4 secara simultan memengaruhi kematian bayi di Kalimantan Timur tahun 2019.

Berdasarkan hasil uji parsial pada Tabel 3, dapat diketahui bahwa *estimator* parameter  $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{13}, \hat{\beta}_{15}$  mempunyai nilai statistik Wald kurang dari  $-Z_{\alpha/2}$  dan untuk *estimator* parameter  $\hat{\beta}_{14}$  mempunyai nilai statistik Wald lebih dari  $Z_{\alpha/2}$ , dimana nilai  $Z_{\alpha/2}$  sebesar 1,65. Oleh karena itu, tolak  $H_0$ . Hal ini menunjukkan bahwa masing-masing dari variabel persentase bayi BBLR, persentase bayi yang mendapatkan ASI secara eksklusif, persentase cakupan kunjungan neonatal lengkap KN3, persentase cakupan pelayanan kesehatan bayi, dan persentase cakupan kunjungan ibu hamil K4 secara signifikan memengaruhi kematian bayi di Kalimantan Timur tahun 2019.

**Pendeteksian Overdispersi**

Setelah dilakukan analisis regresi Poisson, dilakukan pendeteksian overdispersi. Hasil yang diperoleh adalah ragam data kematian bayi di Kalimantan Timur tahun 2019 (1.728) lebih dari nilai reratanya (69). Sementara itu, nilai devians (9,83) dibagi dengan derajat bebasnya (5) lebih dari 1, yaitu sebesar 1,97. Hasil ini menunjukkan bahwa terdapat overdispersi pada model regresi Poisson. Oleh karena itu, untuk menangani overdispersi ini digunakan model RBN.

**Pemodelan Kematian Bayi dengan RBN**

Setelah terdeteksi adanya overdispersi pada regresi Poisson, maka dilakukan pemodelan kasus kematian bayi di Provinsi Kalimantan Timur tahun 2019 dengan RBN. Hasil estimasi dan uji parsial parameter disajikan pada Tabel 4.

Berdasarkan Tabel 4, didapatkan model RBN:  $\eta_2(x_i) = 8,3725 - 0,0644X_{1i} - 0,0112X_{2i} + -0,0413X_{3i} + 0,0503X_{4i} - 0,0419X_{5i}$ .

$$\tag{21}$$

**Tabel 4.** Hasil Estimasi dan Uji Parsial Parameter Model RBN

Par	Est	SE	$W_1$	P
$\hat{\beta}_{20}$	8,3725	3,6779	2,2764	0,0299*
$\hat{\beta}_{21}$	-0,0644	0,0526	-1,2243	0,1885
$\hat{\beta}_{22}$	-0,0112	0,0209	-0,5359	0,3456
$\hat{\beta}_{23}$	-0,0413	0,0228	-1,8114	0,0773*
$\hat{\beta}_{24}$	0,0503	0,0183	2,7486	0,0091*
$\hat{\beta}_{25}$	-0,0419	0,0188	-2,2287	0,0333*

\*Signifikan pada  $\alpha = 0,1$ .

Interpretasi dari model RBN pada Persamaan (21) adalah:

1. Jika variabel persentase cakupan kunjungan neonatal lengkap KN3 meningkat sebesar 1 persen, maka akan menurunkan ln rerata kematian bayi di Kalimantan Timur sebesar 0,0413, dimana variabel yang lain dianggap tetap.
2. Jika variabel persentase cakupan pelayanan kesehatan bayi menurun sebesar 1 persen, maka akan meningkatkan ln rerata kematian bayi di Kalimantan Timur sebesar 0,0503, dimana variabel yang lain dianggap tetap.
3. Jika variabel persentase cakupan kunjungan ibu hamil K4 meningkat sebesar 1 persen, maka akan menurunkan ln rerata kematian bayi di Kalimantan Timur sebesar 0,0419, dimana variabel yang lain dianggap tetap.

Sementara itu, diperoleh nilai statistik  $G_2$  sebesar 9,83 dengan  $P$ -value sebesar 0,08 dan nilai  $\chi^2_{(0,1;5)}$  sebesar 9,24. Hasil ini menunjukkan bahwa nilai  $G_2$  lebih dari  $\chi^2_{(0,1;5)}$  dan  $P$ -value tidak lebih dari  $\alpha$ . Oleh karena itu, diputuskan tolak  $H_0$  dan dapat disimpulkan bahwa variabel persentase bayi BBLR, persentase bayi yang mendapatkan ASI secara eksklusif, persentase cakupan kunjungan neonatal lengkap KN3, persentase cakupan pelayanan kesehatan bayi, dan persentase cakupan kunjungan ibu hamil K4 secara simultan memengaruhi kematian bayi di Kalimantan Timur tahun 2019.

Selanjutnya, pada Tabel 4 tampak bahwa nilai statistik uji parsial  $W_2$  untuk *estimator* parameter  $\hat{\beta}_{23}$  dan  $\hat{\beta}_{25}$  mempunyai nilai statistik Wald tidak lebih dari  $-Z_{\alpha/2}$  dan untuk  $\hat{\beta}_{24}$  mempunyai nilai statistik Wald lebih dari  $Z_{\alpha/2}$ , dimana nilai  $Z_{\alpha/2}$  sebesar 1,645. Sehingga, diputuskan tolak  $H_0$ . Hal ini menunjukkan bahwa persentase cakupan kunjungan neonatal lengkap KN3, persentase cakupan pelayanan kesehatan bayi, dan persentase cakupan kunjungan ibu hamil K4 secara individu memengaruhi kematian bayi di Kalimantan Timur tahun 2019.

**Kesimpulan dan Saran**

**Kesimpulan**

Model RBN merupakan suatu metode regresi alternatif untuk mengakomodasi dan menangani masalah overdispersi pada regresi Poisson. Jika terdapat masalah overdispersi, maka *estimator* parameternya memiliki nilai *standard error* yang bukan nilai sebenarnya (*under-estimate*) dan hasil pengujian hipotesis signifikansi parameter mengakibatkan variabel independen berpengaruh terhadap variabel dependen padahal sebenarnya tidak berpengaruh secara signifikan.

Pemodelan kematian bayi di Kalimantan Timur tahun 2019 dengan regresi Poisson menunjukkan adanya overdispersi. Oleh karena itu, digunakan model RBN untuk menanganinya. Berdasarkan pemodelan dengan RBN, faktor-faktor yang signifikan memengaruhi kematian bayi di Kalimantan Timur tahun 2019 adalah persentase cakupan kunjungan neonatal lengkap KN3, persentase cakupan pelayanan kesehatan bayi, dan persentase cakupan kunjungan ibu hamil K4.

**Saran**

Saran-saran yang diajukan untuk penelitian selanjutnya adalah menggunakan pendekatan alternatif regresi yang lain untuk menangani overdispersi pada pemodelan kematian bayi di Kalimantan Timur. Misalnya regresi *generalized Poisson* dan regresi *Poisson inverse Gaussian*. Selain itu, perlu dilakukan penelitian lebih lanjut menggunakan faktor-faktor lain yang diduga berpengaruh terhadap kematian bayi sebagai variabel independen. Sehingga, dapat diperoleh lebih banyak faktor-faktor yang berpengaruh terhadap kematian bayi di Kalimantan Timur.

**Daftar Pustaka**

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Hoboken, New Jersey: John Wiley & Sons.

Amiruddin, R. and Hasmi. (2014). *Determinan Kesehatan Ibu dan Anak*. Jakarta: CV. Trans Info Media.

Afri, L. E., Aunuddin., and Djuraidah, A. (2012). Model Regresi Binomial Negatif Terboboti Geografis Untuk Data Kematian Bayi (Studi Kasus 38 Kabupaten/Kota di Jawa Timur), *Forum Statistika dan Komputasi*, 17 (2), 33-39.

Ardiles, L. G., Tadano, Y. S., Costa, S., Urbina, V., Capucim, M. N., Silva, I., Braga, A., Martins, J. A., and Martins, L. D. (2018). Negative Binomial Regression Model for Analysis of the Relationship between Hospitalization and Air Pollution.

- Atmospheric Pollution Research*, 9 (2), 333-341.
- Dai, H., Bao, Y., and Bao, M. (2013). Maximum Likelihood Estimate for the Dispersion Parameter of the Negative Binomial Distribution. *Statistics and Probability Letters*, 83, 21-27.
- Dinas Kesehatan Provinsi Kalimantan Timur. (2020). *Profil Kesehatan Kalimantan Timur 2019*. Samarinda: Dinas Kesehatan Provinsi Kalimantan Timur.
- Fathurahman, M. (2019). Pemodelan Indeks Pembangunan Kesehatan Masyarakat Kabupaten/Kota di Pulau Kalimantan Menggunakan Pendekatan Regresi Probit. *Jurnal Varian*, 2 (2), 47-54.
- Gomes, M. J. T. L., Cunto, F., and Silva, A. R. (2017). Geographically Weighted Negative Binomial Regression Applied to Zona Level Safety Performance Models. *Accident Analysis and Prevention*, 106, 254-261.
- Gujarati, D. N. (2003). *Basic Econometrics*, Fourth Edition. New York: McGraw-Hill/Irwin.
- Hilbe, J. M. (2011). *Negative Binomial Regression, Second Edition*. New York: Cambridge University Press.
- Hilbe, J. M. (2014). *Modeling Count Data*. New York: Cambridge University Press.
- Ibarra-Espinosa, S., Freitas, E. D., Ropkins, K., Dominici, F., and Rehbein, A. (2022). Negative-Binomial and Quasi-Poisson between COVID-19, Mobility, and Environment in Sao Paulo, Brazil. *Environmental Research*, 204, 1-20.
- Iqbal, W., Tang, Y. M., Chau, K. Y., Irfan, M., and Mohsin, M. (2021). Nexus between Air Pollution and NCOV-2019 in China: Application of Negative Binomial Regression Analysis. *Process Safety and Environmental Protection*, 150, 557-565.
- Islam, M. A., Kabir, M. R., and Talukder, A. (2020). Triggering Factors Associated with the Utilization of Antenatal Care Visits in Bangladesh: An Application on Negative Binomial Regression Model. *Clinical Epidemiology and Global Health*, 8, 1297-1301.
- Pawitan, Y. (2001). *All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Clarendon Press.
- Rachmawati, T., Turniati, L., and Basuki, N. H. (2011). Pola Penyakit Penyebab Kematian Bayi di Pedesaan dan Perkotaan, Kondisi Sosio Ekonomi pada Kejadian Kematian Bayi di Indonesia Hasil Riskesdas 2007. *Buletin Penelitian Sistem Kesehatan*, 14 (2), 108-114.
- Tarigan, I. U., Afifah, T., and Simbolon, D. (2017). Faktor-faktor yang berhubungan dengan Pelayanan Bayi di Indonesia: Pendekatan Analisis Multilevel. *Jurnal Kesehatan Reproduksi*, 8, 103-118.
- Wardani, D. K. and 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.