

Penerapan Model Regresi Hurdle Binomial Negatif Menggunakan Algoritma *Broyden-Fletcher-Goldfarb-Shanno* pada Data Jumlah Kematian Bayi di Kota Makassar Tahun 2017

Anisa Haura Salsa Fatih Yusuf¹, Andi Kresna Jaya², Sitti Sahrinan^{3*}

^{1,2,3}Departemen Statistika, Fakultas MIPA, Universitas Hasanudin, Kota Makassar, 90245, Indonesia

*Corresponding author, email: anshaura@gmail.com

Abstract

Poisson regression is a nonlinear regression method used to analyse the relationship between discrete response variables. Equidispersion is the assumption that must be met in the Poisson regression. Furthermore, there are cases in which the equidispersion assumption is invalidated when using the Poisson regression model to analyze data. One such case is overdispersion, which occurs when there is an excess of zero. As a result, the Negative Hurdle Binomial (HBN) regression is implemented to solve the overdispersion issue. Maximum Likelihood Estimation (MLE) with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm was applied in this study to perform parameter estimation. In addition, the HBN regression model was used to analyze the data on the number of infant mortality cases in Makassar in 2017 with the variables assumed to be significant with infant mortality. The percentage of infants who were exclusively breastfed was the variable that had a significant impact on the outcome of HBN regression on the data on the number of infant mortality that experienced overdispersion.

Keywords: Hurdle Binomial Negatif, Maximum Likelihood Estimation, Overdispersion, excess zero, the number of infant mortality.

Abstrak

Regresi Poisson merupakan salah satu regresi nonlinier yang digunakan untuk memodelkan hubungan antara variabel respon yang berupa data diskrit. Asumsi yang harus dipenuhi dalam regresi Poisson yaitu equidispersi. Namun, dalam menganalisis data dengan model regresi Poisson dapat terjadi pelanggaran asumsi equidispersi salah satunya yakni overdispersi yang diakibatkan dengan adanya excess zeros. Oleh karena itu, diperlukan sebuah model untuk mengatasi pelanggaran overdispersi yaitu regresi Hurdle Binomial Negatif (HBN). Pada penelitian ini, estimasi parameter dilakukan menggunakan Maximum Likelihood Estimation (MLE) dengan algoritma Broyden-Fletcher-Goldfarb-Shanno (BFGS). Selanjutnya, model regresi HBN diterapkan pada data jumlah kasus kematian bayi di Makassar tahun 2017 dengan faktor-faktor yang diduga memengaruhi terjadinya kematian bayi. Hasil yang diperoleh dengan menggunakan regresi HBN pada data jumlah kematian bayi yang mengalami overdispersi yaitu variabel yang berpengaruh signifikan adalah persentase bayi yang diberi ASI eksklusif.

Kata Kunci: Hurdle Binomial Negatif, Maximum Likelihood Estimation, Overdispersi, excess zeros jumlah kematian bayi.

1. Pendahuluan

Salah satu indikator derajat kesehatan dan kesejahteraan masyarakat suatu daerah ialah kematian bayi, indikator ini dapat menggambarkan situasional pelayanan kesehatan secara umum di suatu wilayah. Angka Kematian Bayi (AKB) juga dapat menggambarkan kondisi sosial ekonomi masyarakat setempat, karena bayi adalah kelompok usia yang paling rentan terkena dampak dari perubahan lingkungan maupun sosial ekonomi [1]. Hubungan jumlah kematian bayi dengan faktor-faktor penyebabnya dapat diketahui dengan menggunakan salah satu analisis regresi yaitu regresi Poisson. Asumsi penting pada analisis regresi Poisson adalah nilai variansi harus sama dengan nilai rata-rata yang biasa disebut dengan equidispersi [2]. Namun, asumsi tersebut seringkali tidak terpenuhi yang mana pada kasus tertentu terdapat nilai variansi pada variabel respon lebih kecil dari nilai rata-ratanya (underdispersi) atau nilai variansi data lebih besar dari nilai rata-ratanya (overdispersi).

Salah satu alternatif yang dapat digunakan untuk mengatasi model regresi Poisson yang mengalami overdispersi yaitu menggunakan model regresi binomial negatif. Model regresi binomial negatif lebih fleksibel dibandingkan dengan model Poisson karena asumsi nilai variansi dan nilai rata-ratanya tidak harus sama besar. Model binomial negatif merupakan model campuran antara Poisson dan Gamma yang mana distribusi Gamma digunakan untuk menyesuaikan kehadiran overdispersi dalam model Poisson [3]. Pada kasus tertentu, data count tidak hanya mengalami overdispersi akan tetapi dapat juga mengalami kelebihan nol (excess zeros). Data kelebihan nol (excess zeros) adalah kondisi ketika jumlah nilai nol pada variabel respon lebih banyak dari jumlah nilai diskrit lainnya [4].

Berdasarkan hal tersebut, beberapa solusi yang dapat digunakan untuk mengatasi nilai nol yang berlebih yaitu dengan mengaplikasikan model zero inflated atau model hurdle [5]. Zero inflated akan memecah model menjadi 2 bagian, yaitu model pertama merupakan munculnya data count yang hanya berisi nilai nol dan model kedua yaitu nilai non-negatif. Model hurdle pada dasarnya hampir mirip dengan model zero inflated yang melakukan dua pemodelan. Namun, pada pemodelan kedua, model hurdle menggunakan zero truncated untuk data yang bernilai positif pada data diskrit [6]. Keunggulan dari model hurdle adalah kedua model didalamnya dapat dilakukan penaksiran parameter secara terpisah atau dengan kata lain dimaksimumkan secara terpisah sehingga diharapkan dapat lebih mudah dalam penginterpretasiannya [7].

2. Material dan Metode

2.1 Regresi Linier Berganda

Data Profil Dinas Kesehatan Kota Makassar yang digunakan dalam penelitian ini berupa data sekunder yang diperoleh dari *website* resmi Dinas Kesehatan Kota Makassar tahun 2018 [3]. Penelitian ini menggunakan data jumlah kasus kematian bayi di Kota Makassar dengan variabel yang digunakan dalam penelitian ini terdiri dari satu

variabel respon yaitu jumlah kematian bayi (Y) dan empat variabel prediktor diantaranya persentase bayi dengan Berat Badan Lahir Rendah (BBLR) (X_1), persentase pelayanan kesehatan bayi (X_2), persentase bayi yang diberi ASI eksklusif (X_3), dan persentase kunjungan neonatus (X_4). Data pada penelitian ini akan dianalisis menggunakan model regresi *hurdle* binomial negatif sehingga diperoleh estimasi dengan metode *Maximum Likelihood Estimation* (MLE) sebagai berikut:

$$\begin{aligned}
 (\phi, \delta_j, \beta_j; y_i) = & \sum_{y_i=0}^n \ln \left[\frac{\exp(\sum_{j=1}^p z_{ij} \delta_j)}{1 + \exp(\sum_{j=1}^p z_{ij} \delta_j)} \right] + \sum_{y_i>0}^n \ln \left[\frac{1}{1 + \exp(\sum_{j=1}^p z_{ij} \delta_j)} \right] \\
 & + \sum_{\substack{i=1 \\ y_i>0}}^n \left[\ln \left(\Gamma \left(y_i + \frac{1}{\phi} \right) \right) - \ln(\Gamma(y_i + 1)) - \ln \Gamma \frac{1}{\phi} \right] \\
 & + \sum_{\substack{i=1 \\ y_i>0}}^n \left[\ln(\phi \exp(\sum_{j=1}^p x_{ij} \beta_j))^{y_i} - \ln(1 + \phi \exp(\sum_{j=1}^p x_{ij} \beta_j))^{y_i} \right] \\
 & + \sum_{\substack{i=1 \\ y_i>0}}^n \left[\ln \left(1 + \phi \exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right)^{-\phi^{-1}} \right. \\
 & \left. - \ln \left(1 - \left(1 + \phi \exp \left(\sum_{j=1}^p x_{ij} \beta_j \right) \right)^{-\phi^{-1}} \right) \right] \tag{1}
 \end{aligned}$$

Karena turunan pertama fungsi ln-likelihood model regresi Hurdle Binomial Negatif (HBN) terhadap masing-masing parameter $\theta = \delta, \beta$, dan ϕ tidak dapat diselesaikan secara analitik maka estimasi dilanjutkan menggunakan algoritma Broyden-Fletcher-Goldfarb-Shanno (BFGS) dengan langkah-langkah berikut.

1. Tentukan nilai awal dari parameter yang akan ditaksir ($\theta^{(0)}$) dan $\mathbf{H}^{(0)}$, $\mathbf{H}^{(0)} = \mathbf{I}$.
2. Membentuk vektor gradien (\mathbf{g}) dengan menggunakan turunan pertama fungsi ln-likelihood.

$$\mathbf{g}^{(k)} = \nabla f(\theta^{(k)})$$

3. Melakukan iterasi dengan $k = 0, 1, 2, \dots$ pada persamaan

$$\theta^{(k+1)} = \theta^{(k)} - \lambda^{(k)} (\mathbf{H}^{(k)})^{-1} \mathbf{g}^{(k)}$$

Jika nilai estimasi yang diperoleh tidak memenuhi $\|\theta^{(k+1)} - \theta^{(k)}\| \leq 10^{-3}$, maka iterasi dilanjutkan dengan menghitung matriks pembaruan BFGS

$$\mathbf{H}^{(k+1)} = \left(\mathbf{I} - \frac{\mathbf{s}^{(k)} (\mathbf{y}^{(k)})^T}{(\mathbf{y}^{(k)})^T \mathbf{s}^{(k)}} \right) \mathbf{H}^{(k)} \left(\mathbf{I} - \frac{\mathbf{y}^{(k)} (\mathbf{s}^{(k)})^T}{(\mathbf{y}^{(k)})^T \mathbf{s}^{(k)}} \right) + \frac{\mathbf{s}^{(k)} (\mathbf{s}^{(k)})^T}{(\mathbf{y}^{(k)})^T \mathbf{s}^{(k)}}$$

dengan

$$\mathbf{s}^{(k)} = \theta^{(k+1)} - \theta^{(k)}$$

$$\mathbf{y}^{(k)} = \mathbf{g}^{(k+1)} - \mathbf{g}^{(k)}$$

$H^{(0)}$ = matriks identitas $n \times n$

Adapun langkah-langkah penerapan model regresi HBN menggunakan metode MLE pada data jumlah kasus kematian bayi di Kota Makassar tahun 2017 adalah sebagai berikut:

1. Melakukan pengujian kecocokan distribusi Poisson.
2. Melakukan pengujian dispersi
3. Melakukan pengujian kelebihan nol pada variabel respon.
4. Mengestimasi parameter model regresi HBN dengan MLE menggunakan Algoritma BFGS.
5. Melakukan uji signifikansi parameter secara simultan dan parsial satu-satu.
6. Membuat interpretasi dan kesimpulan

3. Hasil dan Diskusi

3.1 Pengujian Kecocokan Distribusi Poisson

Pengujian kecocokan distribusi poisson dilakukan dengan uji *Kolmogorov-Smirnov* dengan hipotesis sebagai berikut:

H_0 : jumlah kasus kematian bayi di Kota Makassar tahun 2017 berdistribusi Poisson

H_1 : jumlah kasus kematian bayi di Kota Makassar tahun 2017 tidak berdistribusi Poisson

Statistik Uji

$$D_{hitung} = maks |F_n(Y) - F_0(Y)|, F_n(Y) = \frac{m}{n} \text{ dan } 1 \leq m \leq n \quad (2)$$

dengan

$F_n(Y)$: fungsi distribusi kumulatif yang diamati

$F_0(Y)$: fungsi distribusi kumulatif yang dihipotesiskan

m : banyaknya pengamatan yang sama atau kurang dari variabel Y

n : banyaknya seluruh pengamatan

Berdasarkan persamaan (2) diperoleh nilai $D_{hitung} = 0,112$ lebih kecil daripada nilai $D_{tabel} = 0,199$ sehingga H_0 diterima. Artinya, variabel jumlah kasus kematian bayi di Kota Makassar tahun 2017 mengikuti distribusi Poisson. Karena variabel jumlah kasus kematian bayi berdistribusi Poisson maka dapat dilakukan analisis regresi Poisson.

3.2 Pengujian Dispersi

Pengujian dispersi dilakukan menggunakan nilai *deviance* yang dibagi dengan derajat bebasnya dengan hipotesis sebagai berikut:

H_0 : $\phi = 1$ (jumlah kasus kematian bayi di Kota Makassar Tahun 2017 tidak mengalami overdispersi)

H_1 : $\phi > 1$ (jumlah kasus kematian bayi di Kota Makassar Tahun 2017 mengalami overdispersi)

Statistik Uji

$$\phi = \frac{dev}{db}$$

Berdasarkan persamaan (3) diperoleh nilai $\phi = 1,585$ lebih besar daripada 1 sehingga H_0 ditolak. Artinya, variabel jumlah kasus kematian bayi mengindikasikan terjadinya overdispersi atau dapat dikatakan asumsi equidispersi pada regresi Poisson tidak terpenuhi.

3.3 Kelebihan Nol pada Variabel Respon

Pengujian kelebihan nol pada variabel respon yaitu banyaknya jumlah kasus kematian bayi di Kota Makassar Tahun 2017 dapat dilakukan dengan melihat frekuensi nilai nol pada jumlah kasus kematian bayi. Hasil uji kelebihan nol ditunjukkan pada Tabel 1 berikut:

Table 1. Frekuensi Data Jumlah Kasus Kematian Bayi

Jumlah Kasus	Frekuensi	Persentase	Persentase Kumulatif
0	24	52,17%	52,27%
1	9	19,57%	71,74%
2	8	17,39%	89,13%
3	4	8,70%	97,83%
4	1	2,17%	100,00%

Persentase banyaknya nilai nol pada variabel jumlah kematian bayi adalah sebesar 52,17% yang merupakan persentase terbesar dari pada persentase nilai lain sehingga dapat disimpulkan bahwa data tersebut mengalami kelebihan nilai nol atau *excess zeros*.

3.4 Pemodelan Regresi Hurdle Binomial Negatif

Model regresi HBN akan diestimasi menggunakan metode MLE yang diperoleh dengan bantuan algoritma BFGS melalui proses iterasi dengan penentuan $\hat{\beta}_j$ pada model menggunakan OLS sebagai nilai awal. Iterasi yang dilakukan bertujuan untuk mendapatkan dugaan parameter yang konvergen, dengan tingkat konvergen yang digunakan adalah 0,001. Hasil perhitungan pendugaan parameter model HBN menggunakan metode MLE yang nilainya ditunjukkan pada tabel berikut:

Table 2. Estimasi Parameter Regresi Hurdle Binomial Negatif

Penduga model <i>zero hurdle</i>		Penduga model <i>truncated</i> binomial negatif	
Penduga parameter	Estimasi	Penduga parameter	Estimasi
$\hat{\delta}_0$	-3,635	$\hat{\beta}_0$	-6,522
$\hat{\delta}_1$	0,078	$\hat{\beta}_1$	0,021

$\hat{\delta}_2$	0,132	$\hat{\beta}_2$	-0,088
$\hat{\delta}_3$	0,067	$\hat{\beta}_3$	-0,071
$\hat{\delta}_4$	-0,154	$\hat{\beta}_4$	0,220
		$\hat{\phi}$	5,754

Berdasarkan Tabel 2 diperoleh model HBN pada data jumlah kasus kematian bayi di Kota Makassar Tahun 2017 adalah sebagai berikut:

Penduga model *zero hurdle*

$$\ln\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -3,635 + 0,078z_{i1} + 0,132z_{i2} + 0,067z_{i3} - 0,154z_{i4}$$

Penduga model *truncated* binomial negatif

$$\ln(\hat{\mu}_i) = -6,522 + 0,021x_{i1} - 0,088x_{i2} - 0,071x_{i3} + 0,220x_{i4}$$

3.5 Pengujian Simultan

Pengujian simultan dilakukan menggunakan uji *likelihood ratio* dengan hipotesis sebagai berikut:

Penduga model *zero hurdle*

$$H_0 : \delta_1 = \delta_2 = \delta_3 = \delta_4 = 0.$$

$$H_1 : \text{Ada } \delta_j \neq 0, j = 1, 2, 3, \text{ dan } 4.$$

Penduga model *truncated* binomial negatif

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$

$$H_1 : \text{Ada } \beta_j \neq 0, j = 1, 2, 3, \text{ dan } 4$$

Table 3. Uji likelihood ratio

Penduga Model	G_{hit}	$\chi^2_{0,05;4}$
Zero Hurdle	106,947	9,488
Truncated Binomial Negatif	223,561	9,488

Nilai statistik uji likelihood ratio pada model *zero hurdle* dan *truncated* binomial negatif secara berturut-turut yaitu 106,947 dan 223,561. Hal tersebut menunjukkan bahwa nilai $G_{hit} > \chi^2_{(0,05,4)} = 9,488$ yang artinya H_0 ditolak. Penolakan terhadap H_0 menunjukkan bahwa terdapat minimal satu variabel prediktor (variabel persentase berat badan lahir rendah, persentase pelayanan kesehatan bayi, persentase bayi yang diberi ASI eksklusif, dan persentase kunjungan neonatus) yang berpengaruh signifikan terhadap variabel respon (variabel jumlah kematian bayi).

3.6 Pengujian Simultan

Pengujian simultan dilakukan menggunakan uji *Wald* dengan hipotesis sebagai berikut:

Penduga model *zero hurdle*

$$H_0: \delta_j = 0$$

$$H_1: \delta_j \neq 0 \text{ dengan } j = 1,2,3, \text{ dan } 4.$$

Penduga model *truncated* binomial negatif

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0 \text{ dengan } j = 1,2,3, \text{ dan } 4.$$

Table 4. Pengujian Parsial Satu-Satu

Penduga model <i>zero hurdle</i>				Penduga model <i>truncated</i> binomial negatif			
Parameter	$SE(\hat{\delta}_j)$	$W(\hat{\delta}_j)$	Keterangan	Parameter	$SE(\hat{\beta}_j)$	$W(\hat{\beta}_j)$	Keterangan
$\hat{\delta}_1$	0,109	0,516	Tidak Signifikan	$\hat{\beta}_1$	0,091	0,053	Tidak Signifikan
$\hat{\delta}_2$	0,206	0,407	Tidak Signifikan	$\hat{\beta}_2$	0,144	0,372	Tidak Signifikan
$\hat{\delta}_3$	0,032	4,552	Signifikan	$\hat{\beta}_3$	0,030	5,605	Signifikan
$\hat{\delta}_4$	0,221	0,486	Tidak Signifikan	$\hat{\beta}_4$	0,180	1,494	Tidak Signifikan

Hasil pengujian signifikansi estimasi parameter model regresi HBN secara parsial berdasarkan nilai uji *Wald* untuk masing-masing model *zero hurdle* = 4,552 dan model *truncated* binomial negatif 5,605 diperoleh nilai $W > \chi^2_{(0,05,1)} = 3,841$ pada variabel persentase bayi yang diberi ASI eksklusif yang artinya yang H_0 ditolak. Penolakan terhadap H_0 menunjukkan bahwa variabel persentase bayi yang diberi ASI eksklusif memiliki pengaruh terhadap variabel jumlah kematian bayi di Kota Makassar untuk model *zero hurdle* dan model *truncated* binomial negatif.

3.7 Pemodelan Regresi Hurdle Binomial Negatif terhadap Parameter yang Signifikan

Hasil estimasi parameter regresi hurdle binomial negatif terhadap variabel yang signifikan yaitu persentase bayi yang diberi ASI eksklusif ditunjukkan pada **Tabel 5**.

Table 5. Estimasi Regresi HBN terhadap Parameter Signifikan

Penduga Parameter	Estimasi	<i>Std. Error</i>	Wald
$\hat{\delta}_0$	-4,905	2,732	4,278

$\hat{\delta}_3$	0,063	0,030	4,352
$\hat{\beta}_0$	5,081	2,185	5,410
$\hat{\beta}_3$	-0,060	0,028	4,560
$\hat{\phi}$	4,636	4,322	1,150

Penduga model regresi *hurdle* binomial negatif yang terbentuk berdasarkan Tabel 5 adalah sebagai berikut:

Penduga model *zero hurdle*

$$\ln\left(\frac{\hat{\pi}_i}{1 - \hat{\pi}_i}\right) = -4,905 + 0,063z_{i3}$$

Interpretasi :

Variabel signifikan pada pendugaan model *zero hurdle* adalah variabel bayi yang diberi ASI eksklusif dan menghasilkan:

$$P(Y_i = 0) = \frac{\exp(0,063)}{1 + \exp(0,063)} = 0,516$$

Hal ini menunjukkan bahwa peluang tidak terjadinya kematian terhadap bayi yang diberi ASI eksklusif sebesar 0,516 kali lebih besar dibanding bayi yang tidak diberi ASI eksklusif. Artinya kecenderungan ditemukannya kasus kematian bayi lebih besar dipengaruhi oleh variabel bayi yang tidak diberi ASI eksklusif dengan menganggap variabel lainnya konstan.

Penduga model *truncated* binomial negatif

$$\ln(\hat{\mu}_i) = 5,081 - 0,060x_{i3}$$

Interpretasi:

Setiap penambahan satu-satuan bayi yang diberi ASI eksklusif maka akan mengurangi rata-rata jumlah kematian bayi sebesar $\exp(-0,060) = 0,942$.

3.8 Pemodelan Regresi Hurdle Binomial Negatif terhadap Parameter yang

Signifikan

Pemilihan model terbaik regresi HBN diuji dengan memilih nilai Akaike Information Criterion (AIC) minimum.

Model Regresi HBN	AIC
Model dengan seluruh variabel	113,48
Model dengan variabel yang signifikan	112,02

Tabel 6 menunjukkan bahwa model regresi HBN dengan variabel yang signifikan mempunyai nilai AIC yang lebih kecil daripada model regresi HBN dengan seluruh

variabel sehingga model regresi HBN dengan variabel yang signifikan yaitu variabel persentase bayi yang diberi ASI eksklusif lebih baik digunakan pada data jumlah kematian bayi yang mengalami overdispersi.

4. Kesimpulan

Penduga model regresi hurdle binomial negatif pada data jumlah kematian bayi di Kota Makassar tahun 2017 diperoleh bahwa variabel yang berpengaruh signifikan yaitu bayi yang diberi ASI eksklusif. Penduga model zero hurdle menunjukkan bahwa peluang tidak terjadinya kematian terhadap bayi yang diberi ASI eksklusif sebesar 0,516 kali lebih besar dibanding bayi yang tidak diberi ASI eksklusif dan berdasarkan penduga model truncated binomial negatif menunjukkan bahwa setiap penambahan satu-satuan bayi yang diberi ASI eksklusif maka akan mengurangi rata-rata jumlah kematian bayi sebesar 0,942.

Daftar Pustaka

- [1] Abdiana. Determinan Kematian Bayi di Kota Payakumbuh. *Jurnal Kesehatan Masyarakat Andalas*, 9, 2015.
- [2] Famoye, F., John T. Wulu, J., & Singh, K. P. On the Generalized Poisson Regression Model with an Application to Accident Data. *Journal of Data Science*, 287-295, 2004.
- [3] Cox, S., West, S. G., & Aiken, L. S. The Analysis of Count Data: A Gentle Introduction to Poisson Regression and Its Alternatives. *Journal of Personality Assessment*, 121-136, 2009.
- [4] Pontoh, R. S., & Faidah, D. Y. Penerapan Hurdle Negative Binomial pada Data Tersensor. *Seminar Nasional Matematika dan Pendidikan Matematika UNY*, 2015.
- [5] Rahayu, R. L., Asrof, A., & Rustiana, S. Perbandingan Regresi Zero Inflated Negative Binomial dan Regresi Hurdle Negative Binomial pada Data Overdispersi (Studi Kasus: Kejadian Difteri di Indonesia). *Jurnal Sains Matematika dan Statistika*, 2018.
- [6] Bhakta, N. *Properties of Hurdle Negative Binomial Models for Zero-Inflated and Overdispersed Count Data. Dalam Dissertation*. The Ohio State University. 2018.
- [7] Cantoni, E., & Zedini, A. A robust version of the Hurdle model. *Journal of Statistical Planning and Inference*, 1214-1223, 2010.