

PEMODELAN REGRESI HURDLE NEGATIVE BINOMIAL TERSENSOR PADA KASUS PENYAKIT DIFTERI

Lusi Eka Afri¹, Arcat²,

^{1,2} Pendidikan Matematika Universitas Pasir Pengaraian

lusiekaafri13@gmail.com

ABSTRACT The censored Hurdle Negative Binomial regression model is a method which can be used for response variables of count data with excess zero values. This model is an alternative solution to overcome over dispersion in Poisson regression due to many zero-value observations. This model uses a two-part approach, the first part is modeling the response variable which a zero value and the second part is modeling the response variable with a non-negative integer value. The type of censored used in this research is the right censored. This research will applied on the theory and empirical studies on the censored Hurdle Negative Binomial regression model. the results of the theory estimation of the parameters of the censored Hurdle Negative Binomial regression model using the maximum likelihood method are obtained implicit nonlinear equations, so the estimation of the model parameters using Newton Rapshon method. Censored Hurdle Negative Binomial regression model of diphtheria cases in Riau Province with the variables of basic immunization coverage, coverage of toddler health services and coverage of health facilities affect the number of diphtheria cases.

Keywords: Difteri, Censored Hurdle Negative Binomial, Excess Zero, Right Censored.

ABSTRAK Model regresi Hurdle Negative Binomial tersensor merupakan metode yang dapat digunakan untuk variabel respon berupa data cacah dengan nilai nol berlebih. Model ini merupakan alternatif solusi untuk mengatasi overdispersi pada regresi Poisson akibat banyak pengamatan bernilai nol. Model ini menggunakan pendekatan dua bagian yaitu bagian pertama memodelkan variabel respon yang bernilai nol dan bagian kedua memodelkan variabel respon bernilai bilangan bulat non-negatif. Jenis sensor yang digunakan pada penelitian ini sensor kanan. Penelitian ini akan melakukan kajian teori dan kajian terapan pada model regresi Hurdle Negative Binomial tersensor. hasil kajian teori estimasi parameter model regresi Hurdle Negative Binomial tersensor menggunakan metode kemungkinan maksimum diperoleh persamaan nonlinier bersifat implisit, sehingga estimasi parameter model menggunakan metode Newton Rapshon. Pemodelan regresi Hurdle Negative Binomial tersensor terhadap kasus difteri di Provinsi Riau dengan variabel cakupan imunisasi dasar, cakupan pelayanan kesehatan balita dan cakupan fasilitas kesehatan (posyandu) berpengaruh terhadap jumlah kasus difteri.

Kata-kata Kunci: Difteri, Hurdle Negative Binomial, Nol Berlebih, Tersensor Kanan

PENDAHULUAN

Analisis regresi Poisson merupakan analisis statistika yang bertujuan untuk memodelkan hubungan antara variabel respon Y berupa data cacah (*count data*) yang berdistribusi Poisson dengan satu atau lebih variabel prediktor X. Distribusi Poisson sering digunakan untuk memodelkan peristiwa yang memiliki peluang kejadian kecil dengan kejadiannya tergantung pada interval waktu tertentu (Osgood 2000). Casella & Berger (1990) menyatakan data cacah yang diasumsikan menyebar Poisson memiliki fungsi peluang sebagai berikut :

$$P(y; \mu) = \frac{e^{-\mu} \mu^y}{y!} ; \quad y = 0,1,2, \dots \text{ dan } \mu > 0 \quad (1)$$

dengan μ adalah rata-rata banyaknya kejadian dalam suatu interval tertentu. Nilai harapan dan ragam dari sebaran Poisson : $E[y] = \text{Var}[y] = \mu$.

Model regresi Poisson merupakan *Generalized Linier Model* dengan data variabel respon mengikuti distribusi Poisson. GLM terdiri atas tiga komponen, yaitu komponen acak, komponen sistematis, dan fungsi penghubung (*link function*). McCullagh dan Nelder (1989) menyebutkan fungsi penghubung untuk regresi Poisson yang biasa digunakan adalah fungsi logaritma yang dinyatakan sebagai berikut:

$$\eta_i = \log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} = \sum_{j=1}^m x_{ij} \beta_j \quad (2)$$

Cameron dan Trivedi (1998) menyatakan bahwa model regresi Poisson memenuhi asumsi nilai tengah (*mean*) variabel respon yang sama dengan ragamnya (*variance*) dan dikenal dengan istilah *equidispersi*. Pemenuhan asumsi tersebut pada kasus nyata sangat sulit ditemukan. Sellers *et al* (2010) menyatakan bahwa pada beberapa kasus nyata seringkali ditemukan data cacah memiliki ragam lebih besar dari nilai tengah (*overdispersi*) atau ragam lebih kecil dari nilai tengah (*underdispersi*).

Menurut Hinde dan Demetrio (2007), terdapat beberapa kemungkinan tidak dipenuhinya asumsi *equidispersi* tersebut yaitu keragaman hasil pengamatan keragaman individu sebagai komponen yang tidak dijelaskan oleh model, korelasi antar respon individu, terjadinya pengelompokan dalam populasi dan peubah teramati yang dihilangkan. Selain itu, untuk kasus yang memiliki peluang kejadian kecil maka akan ditemukan nilai nol yang berlebih pada variabel responnya. Hal ini juga menjadi salah satu indikator penyebab terjadinya *overdispersi*. Akibatnya dapat menyebabkan pendugaan galat baku yang terlalu rendah dan akan menghasilkan pendugaan parameter yang bias ke bawah (*underestimate*). Dengan demikian model regresi Poisson tidak tepat digunakan untuk memodelkan data cacah tersebut.

Mullahy (1986) pertama kali memperkenalkan model data cacah *Hurdle* yang sangat fleksibel dalam mengatasi masalah *underdispersi* dan *overdispersi*. Saffari (2010) membahas tentang model *zero-inflated* dengan data cacah tersensor kanan

sebagai *two part-model* dalam mengatasi overdispersi. Berbagai pendekatan muncul untuk menangani data cacah yang mengandung overdispersi ataupun underdispersi. Salah satunya dengan mengasumsikan model dua tahap untuk variabel respon berupa model campuran Poisson (*mixed Poisson*). Antara lain *Generalized Poisson*, *Gamma-Poisson* (*Binomial Negatif*), *COM Poisson* dan sebagainya. Hilbe (2007) membahas model regresi binomial negatif yang merupakan campuran antara distribusi Poisson dan distribusi Gamma. Selain itu, Afri (2017) membandingkan Regresi Binomial Negatif dan Conway Maxwell Poisson dalam Mengatasi Overdispersi pada Regresi Poisson.

Model *Hurdle Negative Binomial* (HNB) bersifat fleksibel dalam kondisi overdispersi dan underdispersi. Nilai variabel dependen terpisah menjadi dua kondisi yaitu kondisi *zero state* terjadi dengan peluang w_0 , dan kondisi *truncated negative binomial state* terjadi dengan peluang $(1 - w_0)$. Model HNB dengan variabel respon Y_i dapat dinyatakan sebagai berikut (Saffari *et al*, 2012):

$$P(Y_i = y_i) = \begin{cases} w_0 & y_i = 0 \\ (1 - w_0) \frac{\Gamma(y_i + \phi^{-1})}{\Gamma(\phi^{-1})\Gamma(y_i + 1)} \left(\frac{\phi\mu_i}{1 + \phi\mu_i}\right)^{y_i} \left(\frac{1}{1 + \phi\mu_i}\right)^{\phi^{-1}} \frac{1}{1 - (1 + \phi\mu_i)^{-\phi^{-1}}} & y_i > 0 \end{cases} \quad (3)$$

Dengan ϕ adalah parameter dispersi yang tidak bergantung pada variabel independen dengan $\phi > 0$ dan $0 < w_0 < 1$ dan $w_0 = w_0(z_i)$ dimodelkan dengan fungsi penghubung logit untuk model *zero hurdle* dan *truncated negative binomial* masing-masing sebagai berikut :

$$\text{logit}(w_0) = \log\left(\frac{w_0}{1 - w_0}\right) = \sum_{j=1}^m x_{ij}\delta_j$$

$$\log \mu_i = \sum_{j=1}^m x_{ij}\beta_j$$

Dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, p$.

Kejadian variabel respon banyak bernilai nol, maka perlu adanya batasan. Pembatasan atau penyensoran variabel respon untuk asumsi dalam tujuan tertentu. Penyensoran dapat dibedakan atas tersensor kanan, tersensor kiri, tersensor interval dan tersensor acak. Penyensoran atau pembatasan bawah pada variabel respon dikenal dengan tersensor kanan (*right censoring*).

Variabel respon Y_i terkadang tidak ada pembatasan atau penyensoran, tetapi nilai Y_i dapat disensor untuk asumsi dalam tujuan tertentu. Sensor kanan (*right censoring*) merupakan batas bawah dari data variabel yang akan disensor. Jika ada penyensoran untuk pengamatan ke i , maka $Y_i \geq c$ dan jika tidak penyensoran maka $Y_i = y_i$. Variabel indikator sensor d_i didefinisikan sebagai berikut (Saffari *et al*, 2012):

$$d_i = \begin{cases} 1 & \text{jika } y_i \geq c \\ 0 & \text{lainnya} \end{cases} \quad (4)$$

d_i merupakan variabel dummy mengindikasikan apakah pengamatan ke- i (Y_i) tersensor. Fungsi distribusi peluang model tersensor dapat dinyatakan sebagai berikut:

$$P(Y_i = y_i, d_i, x_i) = P(Y_i = y_i)^{1-d_i} P(Y_i \geq y_i)^{d_i} \quad (5)$$

dengan

$$P(Y_i \geq y_i) = \sum_{j=y_i}^{\infty} P(Y_i = j) = 1 - \sum_{j=0}^{y_i-1} P(Y_i = j)$$

$$P(Y_i = y_i) = f(y_i)$$

Fungsi logaritma kemungkinan (*log-likelihood function*) model regresi tersensor dinyatakan sebagai berikut :

$$\log L(\theta_i; y_i) = \sum_{i=1}^n \left\{ (1 - d_i) [I_{y_i=0} \log f(y_i; \theta_i) + I_{y_i>0} \log f(y_i; \theta_i)] + d_i \left(\sum_{j=y_i}^{\infty} P(Y_i = j) \right) \right\} \quad (6)$$

Saffari (2012) memperkenalkan model regresi hurdle negative binomial dengan data cacah tersensor kanan.

METODE PENELITIAN

Data yang digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari data Profil Kesehatan Provinsi Riau Tahun 2016. Variabel dependen yang digunakan data jumlah kasus difteri Variabel independen yang digunakan sebagai berikut:

- X₁ : persentase cakupan imunisasi dasar
- X₂ : persentase cakupan pelayanan anak balita
- X₃ : persentase cakupan posyandu aktif

Langkah-langkah yang dilakukan dalam penelitian sebagai berikut :

1. Mengkaji ulang estimasi dari model regresi *hurdle negative binomial* tersensor
2. Eksplorasi variabel respon Y secara deskriptif dan plot kuantil normal dan histogram untuk mengetahui indikasi kecocokan sekumpulan data terhadap peluang sebaran tertentu (distribusi Poisson).
3. Identifikasi karakteristik data variabel respon serta melakukan uji skor untuk mengidentifikasi terjadinya peluang nol berlebih.
4. Melakukan pengujian terjadi overdispersi pada regresi Poisson dengan menggunakan statistik *deviance* yang dibagi dengan derajat bebasnya.

$$\phi = \frac{D^2}{db} \quad (7)$$

dengan $D^2 = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\mu_i} \right) - (y_i - \mu_i) \right\}$

5. Melakukan pengujian multikolinieritas variabel penjelas dengan menggunakan kriteria *Variance Inflation Factors (VIF)*.
 6. Memeriksa adanya multikolinieritas antar variabel independen
- Menentukan model regresi *hurdle negative binomial* tersensor.

HASIL DAN PEMBAHASAN

Estimasi Parameter Model Hurdle Negative Binomial

Estimasi parameter model *hurdle negative binomial* untuk data tersensor menggunakan metode kemungkinan maksimum (*maximum likelihood*). Metode ini dilakukan dengan memaksimalkan fungsi kemungkinan yaitu dengan cara mencari turunan pertama fungsi kemungkinan terhadap parameter yang disamakan dengan nol.

Fungsi peluang *Hurdle Negative Binomial* Tersensor (*Censored Hurdle Negative Binomial*) yang disingkat dengan CHNB. Fungsi peluang ini diperoleh dengan mensubstitusikan fungsi peluang *Hurdle Negative Binomial* (HNB) dari persamaan (3) pada fungsi peluang data tersensor kanan dari persamaan (5) dinyatakan sebagai

$$\begin{aligned}
 P(Y_i = y_i, d_i, x_i) &= P(Y_i = y_i)^{1-d_i} P(Y_i \geq y_i)^{d_i} \\
 &= [P(Y_i = y_i, \boldsymbol{\beta})]^{1-d_i} P(Y_i \geq y_i)^{d_i} \\
 &= [I_{y_i=0} f(y_i; \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta}) I_{y_i>0} f(y_i; \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta})]^{1-d_i} \left[\sum_{j=y_i}^{\infty} P(Y_i = j) \right]^{d_i} \quad (8)
 \end{aligned}$$

Langkah awal pendugaan parameter adalah membentuk fungsi kemungkinan (*likelihood function*) dari model CHNB dinyatakan sebagai

$$L_{CHNB} = \prod_{i=1}^n [I_{y_i=0} f(y_i, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta}) I_{y_i>0} f(y_i, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta})]^{1-d_i} \left[\sum_{j=y_i}^{\infty} P(Y_i = j) \right]^{d_i}$$

Selanjutnya dilakukan operasi logaritma natural sehingga diperoleh logaritma natural fungsi kemungkinan sebagai berikut

$$LL_{CHNB} = \ln \left(\prod_{i=1}^n [I_{y_i=0} f(y_i, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta}) I_{y_i>0} f(y_i, \boldsymbol{\phi}, \boldsymbol{\delta}, \boldsymbol{\beta})]^{1-d_i} \left[\sum_{j=y_i}^{\infty} P(Y_i = j) \right]^{d_i} \right)$$

Logaritma natural fungsi kemungkinan model regresi CHNB untuk data tersensor kanan dinyatakan sebagai

$$LL_{CHNB} = \sum_{i=1}^n \left\{ (1-d_i) \left[I_{y_i=0} \ln w_0 + I_{y_i>0} \left\{ \ln(1-w_0) + \ln(g) - \ln \left(1 - (1 + \phi e^{x_i^T \beta})^{\phi^{-1}} \right) \right\} \right] + d_i \ln(F) \right\} \quad (9)$$

dengan

$$F = \sum_{j=y_i}^{\infty} P(Y_i = j) = \sum_{j=y_i}^{\infty} \frac{(1-w_0)(g)}{1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}}}$$

$$w_0 = \frac{e^{x_i^T \delta}}{1 + e^{x_i^T \delta}}$$

$$g(y_i, \boldsymbol{\phi}, \boldsymbol{\beta}) = \frac{\Gamma(y_i + \phi^{-1})}{\Gamma(\phi^{-1}) \Gamma(y_i + 1)} \left(\frac{\phi e^{x_i^T \beta}}{1 + \phi e^{x_i^T \beta}} \right)^{y_i} \left(\frac{1}{1 + \phi e^{x_i^T \beta}} \right)^{\phi^{-1}}$$

Estimasi parameter ϕ, δ, β diperoleh dengan memaksimum logaritma natural fungsi kemungkinan pada persamaan (11). Syarat perlu untuk memaksimumkan logaritma natural fungsi kemungkinan adalah mencari turunan pertamanya terhadap masing-masing parameter yaitu ϕ, δ, β dan disamakan dengan nol. Secara matematis dinyatakan sebagai berikut

$$\begin{aligned}\frac{\partial LL_{CHNB}}{\partial \phi} &= 0 \\ \frac{\partial LL_{CHNB}}{\partial \delta} &= 0 \\ \frac{\partial LL_{CHNB}}{\partial \beta} &= 0\end{aligned}\quad (10)$$

Ringkasan hasil turunan pertama logaritma natural fungsi kemungkinan terhadap masing-masing parameter pada persamaan (12).

$$\frac{\partial LL_{CHNB}}{\partial \phi} = \sum_{i=1}^n \left\{ (1-d_i) I_{y_i > 0} \left[\frac{g'}{g} \phi - \{ \ln(1-w_0) + \ln(g) - \ln(1 - (1 + \phi \mu_i)^{\phi^{-1}}) \} \right] + d_i \ln(F) \right\} \quad (11)$$

$$\frac{\partial LL_{CHNB}}{\partial \delta^T} = \sum_{i=1}^n \left\{ (1-d_i) \left[I_{y_i=0} \frac{w_0'}{w_0} + I_{y_i > 0} \frac{-w_0'}{1-w_0} \right] + d_i \frac{1}{F} \frac{\partial F}{\partial \delta^T} \right\} \quad (12)$$

$$\frac{\partial LL_{CHNB}}{\partial \beta^T} = \sum_{i=1}^n \left\{ (1-d_i) I_{y_i > 0} \left[\frac{1}{g} \frac{\partial g}{\partial \beta^T} + \frac{(1 + \phi e^{x_i^T \beta})^{\phi^{-1}} \phi x_i e^{x_i^T \beta}}{\phi (1 + \phi e^{x_i^T \beta}) (1 - (1 + \phi e^{x_i^T \beta})^{\phi^{-1}})} \right] + \frac{d_i}{F} \frac{\partial F}{\partial \beta^T} \right\} \quad (13)$$

Dengan

$$g'_\phi = g \left[-\frac{\Psi(y_i + \phi^{-1})}{\phi^2} + \frac{\Psi(\phi^{-1})}{\phi^2} + \frac{y_i}{\phi} + \frac{\ln(1 + \phi e^{x_i^T \beta})}{\phi^2} - \frac{e^{x_i^T \beta}}{\phi (1 + \phi e^{x_i^T \beta})} \right]$$

$$F'_\phi = \sum_{j=y_i}^{\infty} (1-w_0) \left[\frac{g'_\phi \left(1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}} \right)}{\left(1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}} \right)^2} + g (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}} \left(\frac{e^{x_i^T \beta} / (1 + \phi e^{x_i^T \beta}) - \ln(1 + \phi e^{x_i^T \beta}) / \phi^2}{\left(1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}} \right)^2} \right) \right]$$

$$w_{0' \delta^T} = \frac{e^{x_i^T \delta}}{1 + e^{x_i^T \delta}} = \frac{x_i e^{x_i^T \delta}}{1 + e^{x_i^T \delta}} - \frac{(e^{x_i^T \delta})^2 x_i}{(1 + e^{x_i^T \delta})^2} = \frac{x_i e^{x_i^T \delta}}{(1 + e^{x_i^T \delta})^2}$$

$$F'_\delta = \sum_{j=y_i}^{\infty} g \left[\frac{-w_0'}{1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}}} \right]$$

$$g'_{\beta^T} = g \left[(-y_i - \phi^{-1}) \left(\frac{\phi x_i e^{x_i^T \beta}}{1 + \phi e^{x_i^T \beta}} \right) + y_i x_i \right]$$

$$F'_{\beta} = \sum_{j=y_i}^{\infty} (1 - w_0) \frac{g'_{\beta^T} \left(1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}} \right) - g \left(x_i e^{x_i^T \beta} (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}-1} \right)}{\left(1 - (1 + \phi e^{x_i^T \beta})^{-\phi^{-1}} \right)^2}$$

Berdasarkan persamaan (10), (11), (12) dan (13) diperoleh persamaan nonlinier yang berbentuk implisit sehingga untuk mencari solusi dari persamaan tersebut diperlukan suatu metode iterasi numerik yaitu metode iterasi Newton Raphson. Secara umum dapat dinyatakan sebagai

$$\theta_{(m+1)} = \theta_{(m)} - H_{(m)}^{-1}(\theta_{(m)})g_{(m)}(\theta_{(m)}) \quad 14$$

Dengan $\theta_{(m)} = [\phi \quad \delta \quad \beta]^T$

Langkah awal prosedur iterasi dimulai dengan menentukan nilai taksiran awal parameter yaitu pada saat $m = 0$. Selanjutnya membentuk vektor kemiringan (*slope*) g yang merupakan turunan pertama logaritma natural fungsi kemungkinan terhadap parameter θ . Matriks Hessian H merupakan turunan kedua logaritma natural fungsi kemungkinan terhadap parameter θ . Dengan mensubstitusikan nilai taksiran awal ke elemen vektor g dan matriks H sehingga diperoleh vektor $g_{(0)}$ dan matriks $H_{(0)}$. Iterasi terus dilakukan hingga mencapai penduga parameter yang konvergen dengan pendugaan parameter yang konvergen diperoleh jika $\|\theta_{(m+1)} - \theta_{(m)}\| < \varepsilon$ (ε merupakan bilangan yang sangat kecil).

Deskripsi Data

Variabel respon yang digunakan dalam penelitian ini adalah jumlah kasus difteri (Y) di Provinsi Riau Tahun 2016. Pada tahun 2016 jumlah kasus difteri di Provinsi Riau meningkatkan tajam dari tahun sebelumnya dengan kasus terbanyak di kabupaten Bengkalis tercatat 7 kasus, selanjutnya disusul oleh kabupaten Rokan hulu dan kota Pekanbaru sebanyak 1 kasus. Sedangkan di kabupaten/kota lain bebas dari kasus ini. Terdapat tiga variabel penjelas yang di duga berpengaruh terhadap variabel respon jumlah kasus difteri. Ketiga variabel tersebut adalah persentase cakupan imunisasi dasar (X_1), persentase cakupan pelayanan anak balita (X_2), dan persentase cakupan posyandu aktif (X_3). Tabel 1 menjelaskan secara deskriptif dari variabel yang digunakan pada penelitian.

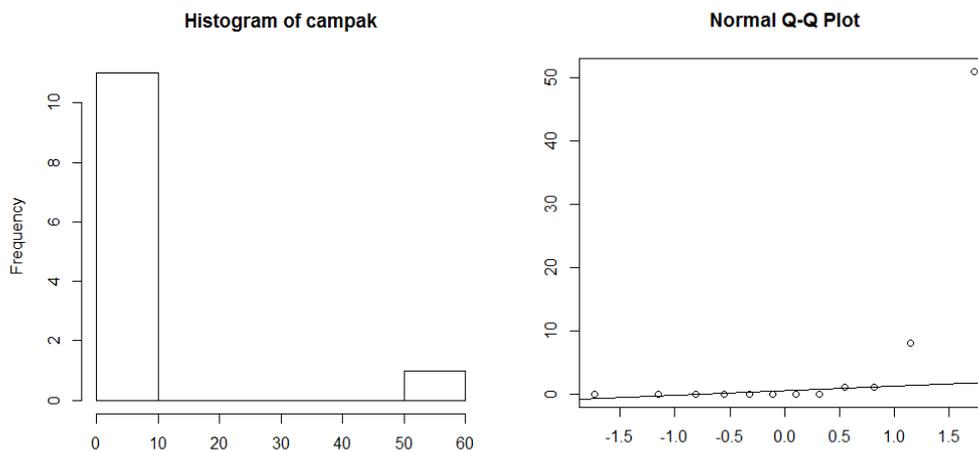
Tabel 1. Statistika Deskriptif Variabel

Sumber	Varaibel			
	Y	X_1	X_2	X_3
Rata-rata	0.75	73.12	67.55	53.21
Minimum	0	10.05	17.30	20.49
Maksimum	7	100	89.40	100

Berdasarkan tabel 1, diketahui bahwa cakupan imunisasi dasar, pelayanan kesehatan balita dan cakupan posyandu aktif masih ada yang rendah di beberapa kabupaten. Bahkan ada juga yang sudah mencapai target. Namun pembangunan kesehatan perlu diupayakan oleh semua lapisan masyarakat agar terwujud derajat

kesehatan yang setinggi-tingginya. Keberhasilan pembangunan kesehatan sangat ditentukan oleh kesinambungan antara upaya program dan sektor.

Gambar 1 merupakan diagram kuantil normal dan histogram dari jumlah kasus difteri di Provinsi Riau. Diagram ini dapat digunakan untuk mengetahui pola distribusi data. Diagram kuantil normal terlihat distribusi data tidak mengikuti garis lurus dan histogram variabel Y terlihat tidak simetris. Hal ini menunjukkan bahwa distribusi data jumlah kasus difteri menyimpang dari distribusi normal. Dengan kata lain asumsi awal bahwa data kasus difteri menyebar mengikuti distribusi Poisson.



Gambar 1. Histogram dan Plot Kuantil Normal

Identifikasi terhadap variabel respon Y disajikan pada tabel 2 yang menunjukkan bahwa variabel respon Y mempunyai nilai nol berlebih yaitu dengan frekuensi sebesar 75%. Dengan menggunakan uji skor diperoleh nilai statistik $S_{\omega} = 3.84$ yang lebih besar $\chi^2_{0.05;1} = 3.84$ pada taraf signifikansi 0.05. Hal ini menunjukkan bahwa terjadi peluang nol berlebih sebagai penyebab terjadinya overdispersi pada variabel respon Y.

Tabel 2. Eksplorasi Distribusi Data Variabel Dependen Y

Jumlah Kasus Difteri	Jumlah	Persentase
0	9	75
1	2	16.67
7	1	8.33

Pemeriksaan terjadinya overdispersi dilakukan dengan nilai *deviance* yang dibagi dengan derajat bebasnya. Hasil pemeriksaan overdispersi pada regresi Poisson diperoleh nilai ϕ sebesar 2.95. nilai tersebut lebih besar dari 1, maka hal mengindikasikan kondisi overdispersi pada kasus difteri di Provinsi Riau.

Model Regresi Hurdle Negative Binomial Tersensor

Sebelum pembentukan model dilakukan pengecekan multikolinieritas variabel penjelas. Hasil pemeriksaan multikolinieritas pada tabel 3 terlihat bahwa nilai VIF untuk semua variabel penjelas kurang dari 10. Hal ini menunjukkan bahwa tidak terdapat multikolinieritas.

Tabel 3. Nilai Variance Inflation Factor (VIF)

Variabel	Varaibel		
	X ₁	X ₂	X ₃
VIF	1.615	1.104	1.734

Pemodelan data kasus difteri di Provinsi Riau dilakukan menggunakan model regresi *hurdle negative binomial* tersensor dengan menggunakan tiga variabel penjelas. Pada taraf signifikansi $\alpha = 5\%$ diperoleh estimasi parameter model seperti yang disajikan pada tabel 4. Dengan nilai AIC sebesar 20.42.

Tabel 4. Estimasi Parameter Model Hurdle Negative Binomial Tersensor

Parameter	Estimasi	P-value	Parameter	Estimasi	P-value
δ_0	-0.831	0.832	β_0	0.181	0.850
δ_1	-0.023	0.047*	β_1	0.045	0.050*
δ_2	0.167	-0.022*	β_2	-0.032	0.014*
δ_3	0.211	0.003*	β_3	-0.057	0.032*
ϕ	0.506*				

Keterangan: *) signifikan pada $\alpha = 5\%$

Faktor yang memberikan pengaruh yang ditemukan pada kasus difteri yaitu cakupan imunisasi dasar, cakupan pelayanan anak balita dan cakupan fasilitas kesehatan (posyandu). Model hurdle negative binomial tersensor pada kasus difteri di provinsi riau sebagai berikut :

$$\mu = \exp (0.181 - 0.045X_1 + 0.032X_2 - 0.057X_3)$$

Pada model diperlihatkan bahwa setiap penambahan satu persen cakupan imunisasi dasar maka akan menambah jumlah kasus difteri sebanyak 1.045 kali dari jumlah kasus difteri semula, jika variabel lain bernilai konstan. Sebaliknya penambahan satu persen cakupan pelayanan balita maka akan menurunkan peluang kasus difteri sebesar 1.033 kali dari jumlah kasus difteri semula, jika variabel lain bernilai konstan. Selain itu, penambahan satu persen cakupan posyandu maka akan menurunkan peluang kasus difteri sebesar 1.059 kali dari jumlah kasus difteri semula, jika variabel lain bernilai konstan.

Pengujian signifikansi estimasi parameter model regresi hurdle negative binomial tersensor secara simultan dengan taraf signifikansi 5% diperoleh nilai statistik G sebesar 25.4 lebih besar dari $\chi^2_{0.05,6} = 12.59$. Hal ini menunjukkan bahwa minimal ada satu variabel penjelas yang berpengaruh signifikan terhadap variabel respon Y.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Estimasi parameter model regresi *hurdle negative binomial* tersensor dengan variabel respon tersensor kanan menggunakan metode kemungkinan maksimum menghasilkan persamaan nonlinier yang bersifat implisit sehingga untuk menyelesaikan estimasi parameter digunakan metode iterasi Newton Rapshon.
2. Berdasarkan analisis regresi *hurdle negative binomial* tersensor menggunakan kasus differi diperoleh bahwa setiap variabel penjelas memberikan pengaruh yang berbeda.

Saran yang dapat disampaikan berdasarkan hasil penelitian sebagai berikut:

1. Estimasi parameter model regresi *hurdle negative binomial* tersensor menggunakan metode pendugaan parameter metode Bayes.
2. Rendahnya cakupan pelayanan kesehatan anak balita tetap harus dipantau pertumbuhan, perkembangan dan vitamin A.

Cakupan imunisasi selain imunisasi dasar juga perlu ditingkatkan kepada masyarakat demi terciptanya pembangunan kesehatan.

DAFTAR PUSTAKA

- Afri, Lusi Eka. 2017. Perbandingan Regresi Binomial Negatif dan Conway Maxwell Poisson dalam Mengatasi Overdispersi pada Regresi Poisson. *Jurnal Gantang*. II (1): 79 - 88
- Cameron A.C dan Trivedi P.K. 1998. *Regression analysis of count data*. Cambridge: Cambridge University Press.
- Casella G, Berger RL. 1990. *Statistical Inference*. California: Brooks/Cole.
- Dinkes Riau. 2016. *Profil Kesehatan Provinsi Riau 2016*. Pekanbaru: Dinas Kesehatan Provinsi Riau.
- Hardin JW, Hillbe JM. 2007. *Generalized Linier Models and Extensions*. Texas: Stata Press.
- Hinde J, Dem'etrio CGB. 2007. *Overdispersion: Models and Estimation. A Short Course for SINAPE 1998*
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models Second Edition*. London: Chapman and Hall.
- Mullahy, J. 1986. Specification and Testing of Some Modified Count Data Modified Count Data Models. *Journal of Econometrics*.33: 341-365.
- Osgood D Wayne. 2000. Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology* ,16: 21-43.
- Saffari, S. E., dan Robiah, A. (2010). Zero Inflated Poisson Regression Model with Right Censored Count Data. *Journal of Mathematics*,27(1), 21-29.

Saffari, S. E., Robiah, A. dan Greene, W. 2012. Hurdle Negative Binomial Regression Model with Right Censored Count Data. *Journal of Statistics and Operations Research Transactions*. 36(2), 181-194.