

Perbandingan Regresi *Zero Inflated Negatif Binomial* dan Regresi *Hurdle Negatif Binomial* pada Data Overdispersi (Studi Kasus: Kejadian Difteri di Indonesia)

Rini Luciani Rahayu¹, Asnaffiya Asrof², Shaila Rustiana³,
Widia Puspitasari⁴, Yusep Suparman⁵

^{1,2,3,4,5} Prodi Magister Statistika Terapan, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Padjadjaran, Jl. Dipati ukur No. 35, Jawa Barat, Indonesia
E-mail: *al.ashry1211@gmail.com*

ABSTRAK

Difteri merupakan penyakit menular yang menyerang saluran pernafasan bagian atas dan terlihat selaput putih kotor yang semakin lama akan membesar yang akan mempersempit saluran pernafasan. Penyakit difteri merupakan kejadian luar biasa (KLB), yaitu penyakit yang sebelumnya memiliki jumlah kasus yang sedikit tetapi mengalami peningkatan pesat. Jumlah kasus difteri merupakan data *count* yang mengikuti distribusi poisson sehingga untuk mengetahui faktor-faktor yang berpengaruh terhadap kasus difteri menggunakan analisis regresi Poisson. Regresi Poisson mensyaratkan kondisi dimana nilai *mean* dan variansi dari variabel respon bernilai sama. Namun, adakalanya terjadi fenomena overdispersi dalam data yang dimodelkan dengan distribusi Poisson. Pada penelitian ini dilakukan analisis dengan memperhatikan *excess zero* dan overdispersi menggunakan regresi *Zero Inflated Negative Binomial* (ZINB) dan model regresi *Hurdle Negative Binomial* (HNB) pada kasus difteri di Indonesia. Hasil penelitian menunjukkan bahwa model ZINB memiliki nilai AIC sebesar 123,40 lebih kecil dibandingkan model HNB. Sehingga model ZINB merupakan model yang tepat untuk mengendalikan overdispersi pada data kasus difteri di Indonesia.

Kata Kunci: AIC, Difteri, HNB, ZINB

ABSTRACT

Diphtheria is an infectious disease that attacks the upper respiratory tract and visible a dirty white membrane that will grow longer which will narrow the respiratory tract. Diphtheria disease is an extraordinary event (KLB), a disease that previously had a small number of cases but experienced a rapid increase. The number of diphtheria cases is a data count following the poisson distribution so as to know the factors that affect the case of diphtheria using Poisson regression analysis. Poisson regression requires conditions where the mean and variance values of the response variable are equal. However, sometimes there is an overdispersion phenomenon in the data modeled with the Poisson distribution. In this study, the analysis was done by considering excess zero and overdispersion using Zero Inflated Negative Binomial (ZINB) regression and Hurdle Negative Binomial (HNB) regression model in diphtheria case in Indonesia. The results showed that ZINB model has AIC value 123.40 is smaller than HNB. So the ZINB model is the right model for controlling overdispersion on Diphtheria case data in Indonesia.

Keyword: AIC, Diphtheria, HNB, ZINB

Pendahuluan

Kesehatan merupakan salah satu unsur kesejahteraan masyarakat, melalui pembangunan kesehatan diharapkan dapat meningkatkan derajat kesehatan masyarakat dalam rangka memperbaiki kualitas hidup manusia. Berdasarkan *Sustainable Development Goals* (SDGs) 2030, *United Nations* menargetkan pada tahun 2030, mengakhiri kematian bayi dan anak di bawah usia 5 tahun yang dapat dicegah, dengan semua negara bertujuan untuk mengurangi angka kematian neonatal sebesar 12 per 1.000 kelahiran hidup dan angka kematian di bawah 5 tahun setidaknya mencapai 25 per 1.000 kelahiran hidup. Hal ini dapat dilakukan dengan mengurangi resiko terjadinya penyakit dengan dilakukannya imunisasi. Salah satu penyakit yang dapat dicegah dengan

imunisasi adalah penyakit difteri dengan pemberian vaksin *Difteri Pertusis Tetanus* (DPT). Difteri merupakan penyakit menular yang menyerang saluran pernafasan bagian atas dan terlihat selaput putih kotor yang semakin lama akan membesar yang akan mempersempit saluran pernafasan.

Kementrian Kesehatan RI mencatat jumlah kasus difteri pada tahun 2015 sebanyak 252 kasus dengan jumlah kasus meninggal sebanyak 5 kejadian yang hanya tersebar di 13 Provinsi di Indonesia. Dari 13 provinsi yang melaporkan adanya kasus difteri, kasus tertinggi terjadi di Sumatera Barat dengan 110 kasus dan Jawa Timur sebanyak 67 kasus. Terjadi peningkatan kasus yang besar di Provinsi Sumatera Barat (kejadian luar biasa (KLB)) dibandingkan tahun 2014 yang hanya sejumlah 9 kasus. Sedangkan jumlah kasus difteri di Jawa Timur telah menunjukkan penurunan dibandingkan tahun 2014 (396 kasus) dan 2013 (610 kasus). Penyakit difteri merupakan kejadian luar biasa (KLB), yaitu penyakit yang sebelumnya memiliki jumlah kasus yang sedikit tetapi mengalami peningkatan pesat.

Jumlah kasus difteri merupakan data *count* yang mengikuti distribusi *Poisson* sehingga untuk mengetahui faktor-faktor yang berpengaruh terhadap kasus difteri menggunakan analisis regresi Poisson. Dalam analisis regresi Poisson asumsi *mean* sama dengan varian biasanya jarang terpenuhi, karena seringkali muncul adanya fenomena *over/under* dispersi dalam pemodelan tersebut. Overdispersi data *count* juga dapat terjadi jika mengalami *excess zero*, yaitu kondisi dimana proporsi nilai nol pada data lebih besar dari nilai lainnya. Jika terjadi *excess zero* dan overdispersi, regresi Poisson tidak sesuai untuk memodelkan data dan model yang akan terbentuk menghasilkan estimasi yang bias. Oleh karena itu pada penelitian ini dilakukan analisis dengan memperhatikan *excess zero* dan overdispersi menggunakan regresi ZINB dan model regresi HNB, dimana kedua metode tersebut akan dibandingkan, untuk mengetahui model yang tepat digunakan pada data.

Berdasarkan permasalahan di atas, maka pada penelitian ini akan dilakukan analisis terhadap variabel respon jumlah kasus difteri di Indonesia dengan variabel bebas yang terdiri dari persentase rumah tangga dengan sanitasi layak, persentase penduduk yang mengakses air bersih, persentase balita yang mendapat imunisasi dasar lengkap, persentase tempat-tempat umum (TTU) yang memenuhi syarat kesehatan, persentase tempat pengelolaan makanan (TPM) yang memenuhi syarat kesehatan dan persentase desa/kelurahan dengan sanitasi total berbasis masyarakat (STBM) menggunakan metode ZINB dan HNB. Kedua model tersebut akan dibandingkan, untuk mengetahui model yang tepat dalam mengatasi *excess zero* dan overdispersi pada data kasus difteri di Indonesia.

Metode dan Bahan Penelitian

1. Regresi Poisson

Hilbe [6] menyatakan bahwa regresi Poisson merupakan suatu bentuk analisis regresi yang digunakan untuk memodelkan data *count*, seperti banyaknya kejadian yang terjadi dalam suatu periode waktu tertentu. Baharuddin [2] mengatakan bahwa metode regresi Poisson biasanya diterapkan pada penelitian kesehatan masyarakat, biologi, dan teknik dimana variabel responnya (y) berupa cacahan objek yang merupakan fungsi dari sejumlah karakteristik tertentu (x). Misalnya $y_i, i = 1, 2, \dots$ merupakan jumlah kejadian yang muncul dalam selang waktu dengan rata-rata μ_i . Jika y adalah variabel acak Poisson dengan parameter $\mu > 0$, maka fungsi massa peluangnya adalah:

$$f(y : \mu) = \frac{e^{-\mu} \mu^y}{y!},$$

dengan $E(y) = \text{Var}(y) = \mu$.

Model regresi Poisson dapat ditulis sebagai [1]:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k,$$

dimana μ_i adalah rata-rata peubah respon, β adalah koefisien regresi, x adalah peubah prediktor dan k adalah banyaknya peubah prediktor.

2. Pengujian Overdispersi

Winkelmann [9] menyatakan bahwa regresi Poisson mengasumsikan *equidispersi* yaitu kondisi saat nilai rata-rata dan ragam peubah respon bernilai sama. Namun, adakalanya terjadi fenomena overdispersi dalam data yang dimodelkan dengan distribusi Poisson. Namun, seringkali regresi Poisson mengalami overdispersi yaitu nilai ragam atau varians melebihi nilai rata-rata. Data overdispersi jika taksiran dispersi lebih besar dari 1 dan underdispersi jika taksiran dispersi kurang dari 1.

Pengujian overdispersi menggunakan statistik uji *Pearson's Chi Square* [1].

$$\sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{y}_i)}$$

Overdispersi terjadi ketika nilai statistik uji *Pearson's Chi Square* lebih dari χ_{n-k-12} [6], dimana n adalah ukuran sampel, y_i adalah nilai peubah respon pengamatan ke- i , $\hat{\mu}_i$ adalah penduga rata-rata bagi peubah respon pengamatan ke- i dan $V(\hat{y}_i)$ adalah penduga ragam bagi peubah respon pengamatan ke- i .

3. Regresi Zero Inflated Poisson

Salah satu penyebab terjadinya overdispersi adalah lebih banyak observasi bernilai nol daripada yang ditaksir untuk model regresi Poisson. Salah satu metode analisis yang diusulkan untuk lebih banyak observasi bernilai nol daripada yang ditaksir adalah model regresi *Zero Inflated Poisson* (ZIP) [7].

Jika y_i adalah variabel random independen yang mempunyai distribusi ZIP, nilai nol pada observasi diduga muncul dalam dua cara yang sesuai untuk keadaan (*state*) yang terpisah. Keadaan pertama disebut *zero state* terjadi dengan probabilitas ω_i dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Poisson state* terjadi dengan probabilitas $(1-\omega_i)$ dan berdistribusi Poisson dengan *mean* μ_i .

Proses dua keadaan ini memberikan distribusi campuran dua komponen dengan fungsi probabilitas sebagai berikut [7]:

$$\Pr(Y = y) \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i} & , y_i = 0 \\ (1 - \omega_i) \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} & , y_i = 1, 2, \dots, 0 \leq \omega_i \leq 1 \end{cases}$$

Lambert dalam Jansakul dan Hinde [7] menunjukkan model gabungan untuk μ dan ω sebagai berikut:

$$\log(\mu) = X\beta \text{ dan } \text{Logit}(\omega) = \text{Log}\left(\frac{\omega}{1-\omega}\right) = X\gamma,$$

dengan X adalah matriks variabel prediktor, β dan γ adalah vektor parameter yang akan ditaksir, dan ω adalah probabilitas observasi bernilai nol.

4. Regresi Zero Inflated Negative Binomial

Model regresi ZINB adalah sebagai berikut ditunjukkan pada persamaan di bawah ini.

$$f(Y_i = y_i) = q_i + (1 - q_i) \left[\frac{1}{1 + \alpha \mu_i} \right]^{\frac{1}{\alpha}} : y_i = 0$$

$$f(Y_i = y_i) = (1 - q_i) \frac{\Gamma \left[y_i + \frac{1}{\alpha} \right] \left[\frac{1}{1 + \alpha \mu_i} \right]^{\frac{1}{\alpha}} \left[\frac{\alpha \mu_i}{1 + \alpha \mu_i} \right]^{y_i}}{\Gamma \left[\frac{1}{\alpha} \right] y_i!} : y_i = 1, \dots, n$$

dengan

$$\log \left[\frac{q_i}{1 - q_i} \right] = \tau \sum_1^p x_{ij} \beta_j$$

dimana

$f(Y_i = y_i)$ = Fungsi probabilitas terjadinya y kejadian pada lokasi-i
 μ_i = Angka ekspektasi
 y_i = Jumlah kejadian yang terjadi pada lokasi-i
 τ = Parameter skalar

Dengan angka ekspektasi kejadian dan simpangan bakunya sesuai yang diberikan pada persamaan di bawah ini:

$$E(Y_i) = \mu_i = (1 - q_i) \lambda_i$$

$$Var(Y_i) = \mu_i + \left[\frac{q_i + \alpha}{1 - q_i} \right] \mu_i^2$$

dimana

$$\lambda_i = \exp[X_i \beta] = \exp \left[\sum_1^p x_{ij} \beta_j \right] \text{ dengani } = 1, \dots, n$$

$E(Y_i)$ = Angka ekspektasi kejadian
 $Var(Y_i)$ = Simpangan baku dari suatu kejadian
 X_i = Variabel bebas dari model
 β = Koefisien yang diestimasi
 x_{ij} = Nilai parameter j untuk lokasi i

5. Regresi Hurdle

Salah satu pendekatan yang dapat digunakan untuk mengatasi overdispersi adalah model Hurdle. Pada model Hurdle dilakukan dua jenis pemodelan. Pemodelan pertama memodelkan observasi bernilai nol dengan menggunakan model logistik. *Link function* model logistik sebagai berikut:

$$\text{logit} \left(\frac{\pi_i}{1 - \pi_i} \right) = Z_i^T \alpha.$$

Pemodelan kedua memodelkan observasi yang bernilai positif dengan menggunakan *truncated binomial*. Fungsi hubung yang digunakan adalah log yang ditunjukkan sebagai berikut

$$\log(\mu_i) = X_i^T \beta.$$

6. Sumber Data dan Langkah Penelitian

Semua data yang digunakan dalam penelitian adalah data sekunder yang terdapat pada profil Kesehatan Indonesia Tahun 2015 yang dipublikasikan oleh Kementerian Kesehatan RI.

Variabel penelitian yang digunakan terdiri dari variabel respon (Y) yaitu jumlah kasus difteri di Provinsi dan beberapa variabel prediktor Persentase rumah tangga dengan sanitasi layak (X_1), Persentase penduduk dengan air bersih (X_2), persentase balita yang mendapat imunisasi dasar lengkap (X_3), persentase tempat-tempat umum (TTU) yang memenuhi syarat kesehatan (X_4), persentase tempat Pengelolaan Makanan (TPM) yang memenuhi syarat kesehatan (X_5), dan persentase desa/kelurahan yang memiliki sanitasi total berbasis masyarakat (X_6).

Adapun langkah-langkah penelitian sebagai berikut:

- a. Analisa deskriptif variabel respon dan variabel prediktor
- b. Mendeteksi adanya multikolinearitas dari variabel prediktor dengan menggunakan kriteria uji VIF
- c. Menentukan model regresi Poisson
- d. Melakukan pengujian *overdispersion* pada model regresi Poisson
- e. Menangani kelebihan data nol dengan menggunakan *zero inflated Poisson*
- f. Melakukan pengujian *oversdispersion* pada model *zero inflated Poisson*
- g. Mengatasi overdispersi dan kelebihan data nol pada model *zero inflated Poisson* dengan model *zero inflated negatif binomial (ZINB)* dan *hurdle negatif binomial (HNB)*
- h. Melakukan pengujian parameter regresi secara serentak dan parsial pada model ZINB dan HNB
- i. Melakukan interpretasi model ZINB dan HNB banyak kejadian Difteri di Indonesia yang telah didapatkan
- j. Pemilihan model yang tepat antara ZINB dan HNB
- k. Membuat kesimpulan dari hasil analisis

Hasil dan Pembahasan

Pada bagian ini akan dilakukan pemodelan jumlah kasus difteri yang mengalami over dispersi dengan model ZINB dan HNB. Pada perhitungannya menggunakan program R versi 3.3.2. Untuk mengukur ketepatan model digunakan *Akaike's Information Criterion (AIC)*.

Deskripsi Variabel Penelitian

Jumlah kasus difteri di Indonesia sebanyak 250 kasus yang tersebar pada 34 provinsi. Terdapat 22 provinsi yang tidak ditemukan kasus difteri. Sementara itu jumlah kasus yang paling banyak di Provinsi Sumatera Barat sebanyak 110 kasus. Pada tahun 2015 di Indonesia, rata-rata persentase rumah tangga dengan akses sanitasi layak (X_1) sebesar 59,07 persen. Rata-rata persentase penduduk yang memiliki akses air bersih (X_2) sebesar 68,62 persen. Rata-rata persentase penduduk yang mendapat imunisasi DPT (X_3) sebesar 82,41 persen. Rata-rata persentase TTU yang memenuhi syarat kesehatan (X_4) sebesar 58,84 persen. Rata-rata persentase TPM yang memenuhi syarat kesehatan (X_5) sebesar 10,23 dan rata-rata persentase sanitasi berbasis masyarakat (X_6) sebesar 32,63 persen.

Tabel 1. Statistik Deskriptif Variabel Prediktor

Keterangan	X_1	X_2	X_3	X_4	X_5	X_6
Rata-Rata	59,07	68,62	82,41	58,84	10,23	32,63
Minimum	23,90	41,08	47,30	29,23	0,00	1,87
Maksimum	89,28	93,40	99,80	81,21	28,32	93,84

Identifikasi Multikolinearitas

Pada variabel prediktor yang digunakan dilakukan penghitungan nilai Variance Inflation Factors (VIF). Nilai VIF ini digunakan untuk mengecek ada tidaknya multikolinieritas. Nilai $VIF \geq 10$ menyatakan bahwa terjadi multikolinieritas. Pada Tabel 2 menunjukkan nilai VIF dari masing-masing variabel prediktor kurang dari 10, yang berarti tidak terjadi multikolinieritas

Tabel 2. Nilai VIF Variabel Prediktor

Statistik Uji	X_1	X_2	X_3	X_4	X_5	X_6
VIF	3,1214	2,1842	2,382	1,5716	1,1923	1,3099

Regresi Poisson

Berikut merupakan penaksiran parameter model regresi Poisson:

Tabel 3. Estimasi Parameter Model Regresi Poisson

Variabel	Nilai Estimasi Parameter	Standar Error	Nilai Z	p-value
Intercept	-14,4187	1,4749	-9,776	< 2e-16
X_1	-0,1255	0,0118	-10,627	< 2e-16
X_2	0,1539	0,0132	11,6690	< 2e-16
X_3	0,1273	0,0145	8,7920	< 2e-16
X_4	-0,0184	0,0070	-2,6300	0.009
X_5	0,1935	0,0126	15,4160	< 2e-16
X_6	0,0163	0,0031	5,2300	0.0000
$AIC = 551,93$		$Devians = 491,16$		$df = 27$

Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5% didapatkan devian = 491,16 lebih besar dari $\chi_{0,05,27}^2 = 40.113$, sehingga keputusan dari pengujian parameter secara simultan menyatakan tolak H_0 , yang artinya bahwa ada salah satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Berdasarkan pengujian secara individu dengan taraf signifikansi 5% didapatkan semua parameter memiliki nilai p-value < 5%, berarti semua variabel prediktor dalam model secara individu memberikan pengaruh yang signifikan terhadap jumlah difteri di Indonesia. Sehingga model yang diperoleh sebagai berikut:

$$\ln(\mu_i) = -14,42 - 0,13X_{1i} + 0,15X_{2i} + 0,13X_{3i} - 0,02X_{4i} + 0,19X_{5i} - 0,02X_{6i}$$

Selanjutnya akan dilakukan pengujian overdispersi. Overdispersi dapat diindikasikan dengan nilai *Chi Kuadrat Pearson* (Devian) yang dibagi dengan derajat bebas. Jika nilai tersebut lebih dari 1, maka dikatakan terjadi overdispersi. Nilai $\frac{Devian}{db} = \frac{491.16}{27} = 18,19 > 1$, sehingga dapat dikatakan terjadi overdispersi. Untuk menangani masalah overdispersi karena *excess zero* pada regresi Poisson digunakan model *zero inflated poisson*.

Zero Inflated Poisson

Model regresi ZIP terbagi menjadi dua jenis. Model pertama disebut *count* model atau model log yang digunakan untuk menentukan peluang dari variabel respon suatu amatan bernilai selain nol sedangkan model kedua yaitu zero-inflation model atau model logit yang digunakan untuk menentukan peluang dari variabel respon suatu amatan bernilai nol (Long 1997 dalam Ramadhani 2014).

Tabel 4. Estimasi Parameter Model Zero Inflated Poisson

Variabel	Nilai Estimasi Parameter	Standar Error	Nilai Z	p-value
Model Count				
Intercept	2,805	1,267	2,213	0,027
X ₁	-0,066	0,015	-4,495	0,000
X ₂	0,078	0,019	4,026	0,000
X ₃	-0,004	0,016	-0,267	0,790
X ₄	-0,096	0,014	-6,667	0,000
X ₅	0,094	0,015	6,421	0,000
X ₆	0,107	0,010	10,974	< 2e-16
Model Zero				
Intercept	12,299	5,858	2,099	0,036
X ₁	0,069	0,065	1,062	0,288
X ₂	-0,099	0,081	-1,231	0,218
X ₃	-0,091	0,056	-1,623	0,105
X ₄	-0,072	0,044	-1,645	0,100
X ₅	0,054	0,072	0,747	0,455
X ₆	0,070	0,036	1,924	0,054
<i>AIC</i> = 121,4027		<i>LR</i> = 424,32		<i>df</i> = 12

Berdasarkan hasil pengujian serentak dengan taraf signifikansi 5% didapatkan $LR = 424,32$ lebih besar dari $\chi^2_{0,05,12} = 21,026$, sehingga keputusan dari pengujian parameter secara simultan menyatakan tolak H_0 , yang artinya bahwa ada salah satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Berdasarkan pengujian secara individu dengan taraf signifikan 5% didapatkan hampir semua parameter pada model count kecuali β_3 memiliki nilai p-value < 5%, berarti semua variabel prediktor kecuali X_3 dalam model count secara individu memberikan pengaruh yang signifikan terhadap jumlah difteri di Indonesia. Sedangkan pada model *zero* hanya γ_0 yang signifikan.

Selanjutnya akan dilakukan pengujian overdispersi. Pengujian overdispersi menggunakan statistik uji *Pearson's Chi Square*, nilai statistik uji *Pearson's Chi Square* = 293,56 lebih besar dari $\chi^2_{0,05;20} = 31,41$, sehingga dapat dikatakan pada model ZIP terjadi overdispersi. Untuk mengatasi overdispersi pada model *zero inflated Poisson* digunakan model ZINB dan Regresi HNB.

Model Zero Inflated Negative Binomial (ZINB)

Model ZINB digunakan untuk menanggapi overdispersi pada model *zero inflated Poisson*. Model ZINB dibagi menjadi dua komponen, yaitu model count untuk μ_i dan model zero inflation untuk p_i .

Tabel 5. Estimasi Model Regresi *Zero Inflated Negative Binomial*

Variabel	Nilai Estimasi Parameter	Standar Error	Nilai Z	p-value
Model <i>Count</i>				
Intercept	2,805	1,267	2,213	0,027
X ₁	-0,066	0,015	-4,495	0,000
X ₂	0,078	0,019	4,026	0,000
X ₃	-0,004	0,016	-0,267	0,790
X ₄	-0,096	0,014	-6,667	0,000
X ₅	0,094	0,015	6,420	0,000
X ₆	0,107	0,010	10,974	< 2e-16
Log (<i>theta</i>)	16,008	144,675	0,111	0,912
Model <i>Zero</i>				
Intercept	12,298	5,858	2,099	0,036
X ₁	0,069	0,065	1,062	0,288
X ₂	-0,099	0,081	-1,231	0,218
X ₃	-0,091	0,056	-1,623	0,105
X ₄	-0,072	0,044	-1,645	0,100
X ₅	0,054	0,072	0,747	0,455
X ₆	0,070	0,036	1,924	0,054

Pengujian kesesuaian pada model regresi ZINB dilakukan dengan menggunakan *likelihood ratio (LR)*. Nilai *likelihood ratio* yang didapat adalah 40,53 dan nilai $\chi^2_{0,05,12} = 21,026$, karena nilai $LR = 40,53$ lebih besar dari $\chi^2_{0,05,12} = 21,026$, maka hasil uji menyatakan tolak H_0 . Keputusan menolak H_0 berarti terdapat minimal satu variabel prediktor yang berpengaruh terhadap Y .

Pengujian parameter ZINB secara individu dapat dilihat pada Tabel 5. Berdasarkan Tabel 5 menunjukkan bahwa lima variabel prediktor kecuali X_3 (persentase imunisasi dasar lengkap) berpengaruh secara nyata terhadap kejadian difteri di Indonesia pada model *count*. Namun pada model *zero inflation* semua variabel tidak signifikan.

Intrepretasi model yang terbentuk dari ZINB didasarkan pada nilai *odd rasio* yang terlihat dari nilai $\exp(\beta)$. Untuk variabel X_1 yaitu, persentase rumah tangga dengan sanitasi layak, hasil penelitian menunjukkan bahwa koefisien regresinya bernilai negatif dengan perhitungan sebagai berikut $\exp(-0.066) = 0,93$. Untuk setiap penurunan satu persen rumah tangga dengan sanitasi layak, maka cenderung akan ditemukan kasus difteri sebesar 0,93 kali.

Model *Hurdle Negative Binomial (HNB)*

Model *hurdle* terdiri dari model logit dan model *truncated* Negatif Binomial. Pengujian kesesuaian pada model regresi HNB dilakukan dengan menggunakan *likelihood ratio (LR)*. Nilai *likelihood ratio* yang didapat adalah 38.883 dan nilai $\chi^2_{0,05,12} = 21,026$, karena nilai $LR = 38,883 > \chi^2_{0,05,12} = 21,026$, maka hasil uji menyatakan tolak H_0 . Keputusan menolak H_0 berarti terdapat minimal satu variabel prediktor yang berpengaruh terhadap Y .

Tabel 6. Estimasi Model *Hurdle* Negative Binomial

Variabel	Nilai Estimasi Parameter	Standar Error	Nilai Z	p-value
<i>Model Truncated</i>				
Intercept	2,536	1,270	1,998	0,046
X_1	-0,069	0,015	-4,652	0,000
X_2	0,085	0,020	4,276	0,000
X_3	-0,004	0,017	-0,222	0,825
X_4	-0,097	0,017	-5,859	0,000
X_5	0,095	0,016	-5,859	0,000
X_6	0,106	0,011	9,612	< 2e-16
Log (<i>theta</i>)	14,245	139,162	0,102	0,919
<i>Model Zero Hurdle</i>				
Intercept	-8,428	4,380	-1,924	0,054
X_1	-0,032	0,049	-0,658	0,510
X_2	0,026	0,054	0,489	0,625
X_3	0,091	0,050	1,801	0,072
X_4	0,032	0,033	0,976	0,329
X_5	-0,052	0,068	-0,758	0,448
X_6	-0,034	0,022	-1,528	0,127

Pengujian parameter model *Hurdle* binomial negatif secara individu dapat dilihat pada Tabel 6. Berdasarkan Tabel 6 menunjukkan bahwa lima variabel prediktor kecuali X_3 (persentase imunisasi dasar lengkap) berpengaruh secara nyata terhadap kejadian Difteri di Indonesia pada model count. Namun pada model *zero Hurdle* semua variabel tidak signifikan pada $\alpha = 5\%$.

Intrepretasi model yang terbentuk dari HNB didasarkan pada nilai *odd rasio* yang terlihat dari nilai $\exp(\beta)$. Untuk variabel X_1 yaitu persentase rumah tangga dengan sanitasi layak, hasil penelitian menunjukkan bahwa koefisien regresinya bernilai negatif dengan perhitungan sebagai berikut $\exp(-0.069) = 0,93$. Untuk setiap penurunan satu persen rumah tangga dengan sanitasi layak maka cenderung akan ditemukan kasus difteri sebesar 0,93 kali.

Penentuan Model yang Tepat

Pemilihan model terbaik yang digunakan dengan melihat nilai AIC dari masing-masing model yang disajikan pada Tabel 7. Dilihat dari nilai AIC, dapat disimpulkan bahwa nilai AIC dari model ZINB lebih kecil dibandingkan dengan model *Hurdle* Negatif Binomial. Hal tersebut menunjukkan bahwa model ZINB lebih baik digunakan pada data kejadian Difteri di Indonesia.

Tabel 7. Nilai AIC

Model	AIC
ZINB	123,4027
<i>Hurdle</i> Negatif Binomial	125,0494

Model *Hurdle* negatif binomial sudah mampu mengendalikan nilai nol dan overdispersi, demikian juga dengan model ZINB. Namun pada data kejadian difteri di Indonesia model ZINB lebih tepat digunakan untuk mengendalikan nilai dan overdispersi pada data tersebut.

Kesimpulan

Model regresi Poisson tidak memenuhi asumsi overdispersion, maka digunakan model lain, model yang diusulkan adalah model ZIP, namun pada model ZIP juga masih terjadi overdispersi. Untuk mengatasinya digunakan model ZINB dan model HNB. Nilai AIC dari model ZINB lebih kecil dibandingkan dengan model HNB. Hal tersebut menunjukkan bahwa model ZINB lebih baik digunakan pada data kejadian difteri di Indonesia. Model HNB sudah mampu mengendalikan nilai nol dan overdispersi, demikian juga dengan model ZINB. Namun pada data kejadian difteri di Indonesia model ZINB lebih tepat digunakan untuk mengendalikan nilai dan overdispersi pada data tersebut.

Daftar Pustaka

- [1] Agresti, A., *An Introduction to Categorical Data Analysis*. Second Edition. New Jersey, John Wiley and Sons, Inc., 2007.
- [2] Baharruddin, Ukuran R^2 dalam Regresi Poisson, Universitas Haluoleo, http://staff.undip.ac.id/matematika/dwi_ispriyanti/files/2010/07/regresi_poisson2.pdf, 2005.
- [3] Dewanti, N. P. P., M. Susilawati, dan I.G.A.M. Srinadi, Perbandingan Regresi Zero Inflated Poisson (Zip) Dan Regresi Zero Inflated Negative Binomial (Zinb) pada Data Overdispersion (Studi Kasus: Angka Kematian Ibu di Provinsi Bali), *E-Jurnal Matematika*, Vol. 5 (4), pp. 133-138 ISSN: 2303-1751, 2016.
- [4] Gandhahusada, Srisasi, Illhude, H. Henry D., Pribadi, dan Wita, Parasitologi Kedokteran, Jakarta: Fakultas Kedokteran Universitas Indonesia, 1998.
- [5] Hall, D.B., and Shen, J., Robust Estimation for Zero-Inflated Poisson Regression, *Scandinavian Journal of Statistics*, Blackwell Publishing Ltd., 1-16, 2009.
- [6] Hilbe, J. M., *Negative Binomial Regression*, Second Edition, New York: Cambridge University Press, 2011.
- [7] Jansakul, N. and Hinde, J.P., Score Tests for Zero-Inflated Models, *Computational Statistics and Data Analysis*, Vol.40 : 75-96, 2002.
- [8] Kemenkes. (2016). *Profil Kesehatan Indonesia 2015*. Jakarta: Kemenkes RI.
- [9] Winkelmann, R., *Econometric Analysis of Count Data*, Springer-Verlag Berlin Heidelberg, 2008.