



PEMODELAN PENYAKIT DIFTERI DI SUMATERA BARAT MENGUNAKAN REGRESI ZERO INFLATED DAN REGRESI HURDLE

Fitri Mudia Sari¹, Pardomuan Robinson Sihombing²

¹Universitas Negeri Padang (UNP), Jln. Prof. Dr. Hamka, Air Tawar, Padang

²Badan Pusat Statistik, Jln.dr Sutomo No 6-8 Jakarta

email: robinson@bps.go.id

ABSTRACT

Data that states the number of events in a certain period of time is called count data. Poisson regression is one of the regression models included in the application of GLM that can be used to model the count data. In Poisson regression, there are assumptions that must be met, namely the mean and variance of the response variables must be the same (equidispersion). Several models that are able to overcome overdispersion due to excess zero are the Zero Inflated model and the Hurdle model. This study examines the characteristics of parameter estimation in the modeling of quantified data that is overdispersed due to excess zero using the Zero Inflated Poisson (ZIP), Zero Inflated Negative Binomial (ZINB), Hurdle Poisson (HP) model and the Hurdle Negative Binomial (HNB) model in cases of diphtheria in West Sumatra in 2018. Based on individual parameter testing and AIC values, the HP model provides better performance than the ZIP, ZINB, and HNB models. So the Hurdle Poisson model is better used in this case than other models.

Keywords : excess zero, hurdle, overdispersion, poisson, zero inflated

ABSTRAK

Data yang menyatakan banyaknya kejadian dalam kurun waktu tertentu disebut dengan data cacah. Regresi Poisson adalah salah satu model regresi yang termasuk ke dalam penerapan GLM yang dapat digunakan untuk memodelkan data cacah. Dalam regresi Poisson terdapat asumsi yang harus terpenuhi yaitu rataan dan ragam peubah respon harus sama (equidispersion). Beberapa model yang mampu mengatasi overdispersi akibat nol berlebih adalah model Zero Inflated dan model Hurdle. Penelitian ini mengkaji karakteristik pendugaan parameter pada pemodelan data cacah yang mengalami overdispersi akibat nol berlebih menggunakan model Zero Inflated Poisson (ZIP), Zero Inflated Negative Binomial (ZINB), Hurdle Poisson (HP) dan model Hurdle Negative Binomial (HNB) pada kasus penyakit difteri di Sumatera Barat tahun 2018. Berdasarkan pengujian parameter secara individu maupun nilai AIC, model HP memberikan performa yang lebih baik dibandingkan model ZIP, ZINB, dan HNB. Sehingga model Hurdle Poisson lebih baik digunakan pada kasus ini dibandingkan model lainnya.

Kata kunci: excess zero, hurdle, overdispersi, poisson, zero inflated

1. LATARBELAKANG

Data yang menyatakan banyaknya kejadian dalam kurun waktu tertentu disebut dengan data cacah (count data). Data cacah selalu bernilai positif, karena suatu kejadian tidak mungkin terjadi dalam jumlah negatif. Data cacah tidak dapat

dimodelkan dengan menggunakan regresi OLS (Ordinary Least Square) karena akan terjadi pelanggaran asumsi kenormalan dan asumsi kehomogenan ragam. Pelanggaran asumsi kenormalan terjadi karena sebaran data cacah cenderung menceng ke kanan, hal ini terjadi karena tidak ada kejadian di bawah nol dan jumlah kejadian dengan frekuensi tinggi sangat sedikit. Pelanggaran asumsi kehomogenan ragam terjadi karena pada data cacah sering terjadi peningkatan ragam bersyarat sebagai akibat peningkatan nilai prediktor (Ramadhan, 2019). Data cacah dapat dimodelkan dengan model linier terampat (Generalized linier Model/ GLM). GLM merupakan perluasan dari model regresi linier dengan asumsi peubah penjelas memiliki efek linier akan tetapi tidak mengasumsikan distribusi tertentu dari peubah respon dan digunakan ketika peubah respon merupakan anggota keluarga eksponensial (Nelder dan Weddeburn 1972 dalam Jamilatuzzahro 2018).

Regresi Poisson adalah salah satu model regresi yang termasuk ke dalam penerapan GLM. Regresi Poisson digunakan untuk memodelkan hubungan antara peubah respon Y yang mengikuti sebaran Poisson dengan satu atau lebih peubah bebas X. Sebaran Poisson merupakan sebaran diskrit dengan nilai peubah acak berupa bilangan bulat positif sehingga menjadi pilihan yang baik untuk pemodelan data cacah. Sebaran poisson hanya ditentukan oleh satu parameter yang mendefinisikan baik rata-rata maupun ragam dari sebaran tersebut, sehingga dalam regresi Poisson terdapat asumsi yang harus terpenuhi yaitu rata-rata dan ragam peubah respon harus sama (equidispersion). Namun dalam kenyataannya asumsi tersebut sering dilanggar, dimana ragam lebih kecil dari rata-rata (underdispersion) atau ragam lebih besar dari rata-rata (overdispersion). Pada kebanyakan data count terkadang ditemukan kasus overdispersi (Consul dan Famoye, 1992).

Kasus overdispersi bila diabaikan bisa mengakibatkan terjadinya underestimate pada estimasi standar error, sehingga dapat mengakibatkan kesalahan pada pengambilan keputusan beberapa uji hipotesis, misalnya suatu peubah bebas berpengaruh signifikan padahal dalam kenyataannya peubah tersebut tidak berpengaruh signifikan (Hilbe, 2011). Menurut Hinde dan Demetrio (2007) (dalam Afri 2019), terdapat beberapa kemungkinan tidak dipenuhinya asumsi equidispersi, yaitu keragaman hasil pengamatan (keragaman individu) sebagai komponen yang tidak dijelaskan oleh model, korelasi antar respon individu, terjadinya pengelompokan dalam populasi, dan peubah teramat yang dihilangkan. Selain itu, untuk kasus yang memiliki peluang kejadian kecil maka akan ditemukan nilai nol yang berlebih pada peubah responnya (excess zero). 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). Sehingga model regresi Poisson tidak tepat digunakan untuk memodelkan data cacah tersebut. Beberapa model yang mampu mengatasi overdispersi akibat nol berlebih adalah model Zero Inflated dan model Hurdle. Pada model Zero Inflated, peubah respon dimodelkan dengan mengikuti sebaran campuran antara sebaran Bernaulli dan sebaran Poisson (sebaran data cacahan lainnya). Sedangkan pada model Hurdle dilakukan dua jenis pemodelan, pemodelan pertama memodelkan observasi bernilai nol dengan menggunakan model logistik, pemodelan kedua

memodelkan observasi yang bernilai positif dengan menggunakan truncated binomial.

Penelitian ini mengkaji karakteristik pendugaan parameter pada pemodelan data cacah yang mengalami overdispersi akibat nol berlebih. Model yang digunakan yaitu Zero Inflated Poisson, Zero Inflated Negative Binomial, Hurdle Poisson dan model Hurdle Negative Binomial pada data jumlah kasus penyakit difteri di Sumatera Barat tahun 2018..

2. TINJAUAN PUSTAKA

Regresi Poisson

Regresi poisson adalah bagian dari *Generalized Linear Model* (GLM). GLM memiliki tiga komponen yaitu komponen acak, komponen sistematis dan fungsi penghubung (Dobson, 1983). Fungsi peluang untuk distribusi Poisson adalah :

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

$$= \exp(-\mu + y_i \ln \mu - \ln y_i!) \quad (2)$$

Berdasarkan Persamaan (2) maka fungsi penghubung adalah $\ln \mu$, maka hubungan antara rata-rata variabel respon dengan kombinasi linier variabel prediktor adalah

$$\ln \mu_i = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{atau} \quad \mu_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}} \quad (3)$$

Regresi Poisson digunakan untuk menganalisis hubungan antara variabel respon yang berupa data diskrit dengan satu atau lebih variabel prediktor, dimana nilai rata-rata dan variansnya diasumsikan sama (*equidispersion*). Namun, dalam analisis data diskrit dengan menggunakan model regresi Poisson terkadang terjadi pelanggaran asumsi tersebut, dimana data *count* memiliki varians lebih besar dari rata-ratanya (*overdispersion*) (Cox, 1983; Cameron dan Trivedi, 1998). Overdispersi merupakan nilai dispersi *Pearson ChiSquare* atau *deviance* yang dibagi dengan derajat bebasnya. Jika rasio kedua nilai ini lebih besar dari 1 maka dikatakan terjadi overdispersi pada data. Misalkan θ merupakan parameter dispersi, maka jika $\theta > 1$ artinya terjadi overdispersi pada regresi Poisson, jika $\theta < 1$ artinya terjadi underdispersi dan jika $\theta = 1$ berarti tidak terjadi kasus over/under dispersi yang disebut dengan equidispersi (Famoye, Wulu, dan Singh, 2004). Pelanggaran terhadap asumsi tersebut mengakibatkan taksiran parameter yang diperoleh menjadi tidak efisien walaupun tetap konsisten (Hilbe, 2011). Masalah lain pada regresi Poisson adalah apabila terdapat banyak data yang bernilai nol, sehingga lebih banyak data nol-nya dibanding regresi Poisson yang akan diprediksi. Jika hal ini terjadi, akan menyebabkan regresi Poisson menjadi tidak tepat menggambarkan data yang sebenarnya. Model regresi yang disarankan untuk kasus ini adalah model *Zero Inflated* dan model *Hurdle*.

Zero Inflated Poisson (ZIP)

Model regresi ZIP merupakan model campuran untuk data cacah dengan banyak nilai nol pada peubah respon. Model ini merupakan kombinasi dari sebaran poisson dengan sebaran kejadian yang bernilai nol (Cameron dan Trivedi 1998). Persamaan ZIP yaitu:

$$P(Y = y_i) = \begin{cases} \phi_i + (1 - \phi_i)e^{-\lambda}, & y_i = 0 \\ (1 - \phi_i) \frac{e^{-\lambda} \lambda^{y_i}}{y_i!}, & y_i = 1, 2, \dots \end{cases}$$

1.1 Zero-Inflated Negative Binomial (ZINB)

Regresi ZINB merupakan model yang dibentuk dari distribusi campuran poisson gamma (Hilbe, 2007). Pada ZINB data cacahan (*counting*) diasumsikan diperoleh dari dua proses, yaitu proses pertama merupakan munculnya data *count* yang hanya berisi nilai nol ditentukan dengan probabilitas p . Sedangkan pada proses kedua yaitu nilai nol dan nilai positif pada data cacahan, kedua-duanya dihasilkan oleh suatu proses yang mengikuti distribusi *Negative Binomial* (NB). Model ZINB dapat ditulis sebagai berikut:

$$P(Y_i = y_i | \mathbf{x}_i, \mathbf{z}_i) = \begin{cases} \frac{\exp(\mathbf{z}_i \gamma)}{1 + \exp(\mathbf{z}_i \gamma)} + \frac{1}{1 + \exp(\mathbf{z}_i \gamma)} (1 + \exp(\mathbf{x}_i \beta))^{-\alpha^{-1}}; & y_i = 0 \\ \frac{1}{1 + \exp(\mathbf{z}_i \gamma)} \frac{\Gamma(y_i + \alpha^{-1})}{y_i! \Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \exp(\mathbf{x}_i \beta)} \right)^{\alpha^{-1}}; & y_i > 0 \end{cases}$$

Hurdle Poisson

Menurut Mullahy (1986) kedua bagian model *Hurdle Poisson* terbentuk berdasarkan fungsi masa peluang pada bilangan bulat non-negatif dengan dua bagian. Pemodelan yang pertama adalah tahapan memodelkan nilai observasi yang bernilai nol dengan menggunakan pendekatan model logistik, dengan model logistik sebagai berikut:

$$\text{logit} \left(\frac{\pi_i}{1 - \pi_i} \right) = \mathbf{z}_i^T \boldsymbol{\alpha}$$

dengan: \mathbf{z}_i^T : vektor kovariat untuk variabel prediktornya

$\boldsymbol{\alpha}$: parameter model logit

Pemodelan yang kedua adalah tahapan memodelkan nilai observasi yang bernilai positif dengan menggunakan *truncated Poisson*. Dengan fungsi sebagai berikut:

$$\log(\mu_i) = \mathbf{X}_i^T \boldsymbol{\beta}$$

dengan: \mathbf{X}_i^T : vektor kovariat untuk variabel prediktornya

$\boldsymbol{\beta}$: parameter model *truncated*

Sehingga model umum untuk fungsi peluang model *Hurdle Poisson* adalah sebagai berikut dimana fungsi peluang untuk model *Hurdle* adalah gabungan antara model logit dan model *truncated Poisson* (Cantoni dan Zedini, 2010):

$$P(Y_i = y_i) = \begin{cases} \frac{1}{1 + \exp(\mathbf{z}_i^T \boldsymbol{\alpha})} \\ [C] \left[\frac{(\exp(\mathbf{X}_i^T \boldsymbol{\beta}))^{y_i}}{[(\exp(\mathbf{X}_i^T \boldsymbol{\beta})) - 1] + y_i!} \right] \end{cases}$$

Hurdle Negative Binomial (HNB)

Model regresi *Hurdle Negative Binomial* merupakan model *Hurdle* yang pada model keduanya (model log) menggunakan sebaran *Negative Binomial*. *Negative Binomial* pada model *Hurdle* menjadikan model ini mampu mengatasi data yang overdispersi dan memiliki amatan bernilai nol yang banyak (*excess zero*). Misalkan y_i ($i = 1, 2, \dots, n$) merupakan peubah respon berupa data cacah ($y_i = 0, 1, 2, \dots$), maka fungsi peluang dari model regresi HNB sebagai berikut (Saffari et al. 2012; Desjardin 2013):

$$P(Y_i = y_i) = \begin{cases} \pi_i & ; y_i = 0 \\ \frac{1}{1 + \left(\frac{k}{\mu_i + k}\right)^k} \frac{\Gamma(y + k)}{\Gamma(y + 1)\Gamma(k)} \left(\frac{k}{\mu_i + k}\right)^{\alpha-1} \left(1 - \frac{k}{\mu_i + k}\right)^y & ; y_i > 0 \end{cases}$$

dimana π_i merupakan peluang munculnya nilai nol dan $0 < \pi_i < 1$. Nilai π_i dan μ_i bergantung pada peubah penjelas yang dapat didefinisikan sebagai berikut:

$$\pi_i = \frac{\exp(X_i^T \delta)}{1 + \exp(X_i^T \delta)} \quad \text{dan} \quad \mu_i = \exp(X_i^T \beta)$$

Sehingga model umum untuk fungsi peluang model *Hurdle Negative Binomial* adalah sebagai berikut:

$$P(Y_i = y_i) = \begin{cases} \left(\frac{\exp(X_i^T \delta)}{1 + \exp(X_i^T \delta)}\right) & ; y_i = 0 \\ \frac{\left(\frac{1}{1 + \exp(X_i^T \delta)}\right)}{1 + \left(\frac{k}{\exp(X_i^T \beta) + k}\right)^k} \frac{\Gamma(y + k)}{\Gamma(y + 1)\Gamma(k)} \left(\frac{k}{\exp(X_i^T \beta) + k}\right)^{\alpha-1} \left(1 - \frac{k}{\exp(X_i^T \beta) + k}\right)^y & ; y_i > 0 \end{cases}$$

Pengujian Estimasi Parameter

Pengujian serentak (*overall*) untuk parameter model regresi *Zero-Inflated* dan *Hurdle* dapat menggunakan *Likelihood Ratio Test* (LRT). Hipotesis yang diuji sebagai berikut:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0 ; \gamma_1 = \gamma_2 = \dots = \gamma_k = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_i \neq 0 ; \gamma_i \neq 0, \quad i = 1, 2, \dots, k$$

statistik uji LRT adalah

$$G^2 = -2 \ln \left(\frac{L(\Omega_0)}{L(\Omega)} \right)$$

Kriteria pengujian yang digunakan adalah tolak H_0 jika $G^2 \geq \chi^2_{\alpha; db}$ dengan db adalah derajat bebas yang nilainya sama dengan perbedaan dimensi parameter antara Ω_0 , dan Ω , dan α adalah tingkat signifikansi dalam pengujian.

Pengujian parsial koefisien regresi dilakukan dengan menguji masing-masing parameter pada regresi *Zero-Inflated* dan *Hurdle*. Hipotesis yang diuji sebagai berikut:

$$H_0: \beta_j = 0 ; \gamma_j = 0$$

$$H_1: \beta_j \neq 0; \gamma_j \neq 0 \quad j = 1, 2, \dots, k$$

Statistik uji yang digunakan adalah statistik Wald

$$W_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} ; W_j = \frac{\hat{\gamma}_j}{SE(\hat{\gamma}_j)}$$

Kriteria pengujiannya yaitu tolak H_0 jika $|W_j| > z_{\alpha/2}$ atau p-value $< \alpha$, di mana α adalah tingkat signifikansi dalam pengujian dan SE adalah standard error.

Pemilihan Model Terbaik

Model terbaik dapat ditentukan dengan beberapa ukuran/statistik, di antaranya adalah perbandingan nilai *Akaike's Information Criterion* (AIC). Akaike (1973) dalam Cameron dan Trivedi (1998) mengusulkan kriteria pemilihan berdasarkan fungsi \ln likelihood untuk membandingkan model berdasarkan maximum likelihood. AIC dituliskan sebagai berikut:

$$AIC = -2l(\Omega) + 2p$$

Dengan l adalah nilai \ln likelihood dari model, p adalah banyaknya parameter dalam model dan $\Omega = (\beta, \gamma, \alpha)$. Pemilihan model terbaik dilihat dari nilai terkecil dari AIC.

3. METODE PENELITIAN

Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari buku Sumatera Barat dalam angka tahun 2019 yang meliputi 19 kabupaten/kota. Peubah penelitian yang digunakan adalah sebagai berikut:

- Y : Jumlah kasus difteri di Sumatera Barat tahun 2018
- X₁ : Persentase rumah tangga dengan sanitasi layak
- X₂ : Persentase balita yang mendapat imunisasi DPT
- X₃ : Kepadatan penduduk per km²
- X₄ : Persentase penduduk dengan air bersih

Langkah-Langkah Analisis

Langkah-langkah yang dilakukan dalam penelitian ini adalah:

- Melakukan analisis deskriptif.
- Melakukan pembakuan data pada peubah bebas dengan menggunakan persamaan normal baku. Hal ini dilakukan karena peubah bebas X₃ memiliki satuan yang berbeda dengan peubah bebas lainnya serta memiliki rentang data yang sangat besar, sehingga tidak ada hasil yang tersedia pada model *Zero-Inflated* dan model Hurdle.
- Menentukan model awal regresi Poisson.
- Mengestimasi parameter model regresi Poisson.
- Melakukan uji overdispersi pada model regresi Poisson.
- Mengatasi overdispersi dan kelebihan data nol dengan model *Zero-Inflated* dan model Hurdle.
- Melakukan pemilihan model yang tepat antara model *Zero-Inflated* dan model Hurdle.

Membuat kesimpulan dari hasil analisis..

4. HASIL DAN PEMBAHASAN

Tabel 1 menyajikan deskriptif statistik dari peubah-peubah yang digunakan dalam penelitian ini. Jumlah kasus difteri di kabupaten/kota yang terdapat di Sumatera Barat paling banyak yaitu 14 kasus yang ditemukan di Kota Padang dan paling sedikit adalah tidak ditemukannya kasus difteri di tujuh kabupaten/kota, yaitu Kabupaten Kepulauan Mentawai, Kabupaten Sijunjung, Kabupaten Padang Pariaman, Kabupaten Solok Selatan, Kota Solok, Kota Sawahlunto, dan Kota Padang Panjang. Rata-rata persentase rumah tangga dengan sanitasi layak adalah 60,27%. Rata-rata persentase balita yang mendapatkan imunisasi DPT adalah 77,97%. Rata-rata kepadatan penduduk adalah 765,16 per km². Dan rata-rata persentase penduduk dengan air bersih adalah 72,18%.

Tabel 1. Deskripsi Statistik dari Peubah-Peubah yang Digunakan dalam Penelitian

Peubah	N	Minimum	Maksimum	Rataan
Jumlah kasus difteri (Y)	19	0	14	1,63
Persentase RT dengan sanitasi layak (X ₁)	19	28,49	91,21	60,27
Persentase balita mendapat imunisasi DPT (X ₂)	19	48,88	93,85	77,97
Kepadatan penduduk (X ₃)	19	15,03	5102,54	765,16
Persentase penduduk dengan air bersih (X ₄)	19	48,86	95,46	72,18

Hasil pemodelan regresi Poisson menunjukkan bahwa peubah bebas yang berpengaruh secara signifikan pada alpha 10% adalah persentase balita mendapat imunisasi DPT (X₂). Tabel 2 menunjukkan hasil taksiran parameter dari model awal regresi Poisson.

Tabel 2. Taksiran Parameter Model Regresi Poisson

Variabel	Taksiran parameter	Standar error	Nilai-p	Keterangan
Intersep	0,2684	0,2221	0,2267	Tidak signifikan
Z ₁	0,1501	0,3521	0,6699	Tidak signifikan
Z ₂	0,5515	0,2701	0,0412	Signifikan
Z ₃	-0,1875	0,2484	0,4505	Tidak signifikan
Z ₄	0,2669	0,3795	0,4818	Tidak signifikan
AIC = 91,26		Deviance = 51,902		df = 14

Berdasarkan hasil pengujian overdispersi diperoleh nilai *deviance* sebesar 51,902 dengan derajat bebas 14, sehingga rasio antara *residual deviance* dengan derajat bebasnya adalah 3,7073. Nilai tersebut lebih besar dari satu, maka dapat disimpulkan bahwa terjadi overdispersi pada data. Karena terjadi masalah overdispersi, maka dapat disimpulkan bahwa model regresi Poisson tidak cocok digunakan pada data kasus difteri di Provinsi Sumatera Barat tahun 2018. Model alternatif yang dapat digunakan untuk memodelkan data tersebut adalah model *Zero-Inflated* dan model *Hurdle* karena ditemukan nilai nol yang berlebih pada peubah responnya.

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 Rahayu *et al.* 2018).

Tabel 3. Taksiran Parameter Model *Zero-Inflated Poisson*

Variabel	Taksiran parameter	Standar error	Nilai-p	Keterangan
Model Count				
Intersep	0,4695	0,2435	0,0538	Signifikan
Z_1	-0,0921	0,3603	0,7983	Tidak signifikan
Z_2	1,3911	0,3781	0,0002	Signifikan
Z_3	-0,2369	0,2774	0,3931	Tidak signifikan
Z_4	0,2699	0,3615	0,4554	Tidak signifikan
Model Zero				
Intersep	-129,05	289,87	0,646	Tidak signifikan
Z_1	-40,80	120,82	0,736	Tidak signifikan
Z_2	141,09	317,24	0,657	Tidak signifikan
Z_3	27,86	75,52	0,712	Tidak signifikan
Z_4	10,13	97,70	0,917	Tidak signifikan
<i>AIC</i> = 68,60		Log L = -24,3		<i>df</i> = 10

Berdasarkan hasil pengujian serentak dengan taraf signifikansi 10% didapatkan $|Log L| = 24,3$ lebih besar dari $\chi_{0,1;10}^2 = 15,98$, sehingga keputusan dari pengujian parameter secara simultan menyatakan bahwa belum cukup bukti untuk menerima H_0 , yang artinya bahwa paling tidak ada satu peubah prediktor yang berpengaruh signifikan terhadap peubah respon. Berdasarkan pengujian secara individu dengan taraf signifikan 10% didapatkan hanya intersep dan parameter peubah persentase balita yang mendapat imunisasi DPT (X_2) yang berpengaruh signifikan terhadap jumlah kasus difteri di Sumatera Barat pada model *count*. Sedangkan pada model *zero* tidak ada satupun parameter yang signifikan.

Tabel 4. Taksiran Parameter Model *Zero-Inflated Negative Binomial*

Variabel	Taksiran parameter	Standar error	Nilai-p	Keterangan
Model Count				
Intersep	0,5031	0,2705	0,0629	Signifikan
Z_1	-0,0521	0,4216	0,9017	Tidak signifikan
Z_2	1,2739	0,4420	0,0039	Signifikan
Z_3	-0,2479	0,3554	0,4855	Tidak signifikan
Z_4	0,2104	0,4719	0,6557	Tidak signifikan
Log (theta)	1,5247	0,9761	0,1183	Tidak signifikan
Model Zero				
Intersep	140,27	391,92	0,720	Tidak signifikan
Z_1	-44,18	162,47	0,786	Tidak signifikan
Z_2	153,35	428,63	0,721	Tidak signifikan
Z_3	30,37	101,82	0,765	Tidak signifikan
Z_4	10,83	131,54	0,082	Tidak signifikan
<i>AIC</i> = 67,66		Log L = -22,83		<i>df</i> = 11

Model ZINB dibagi menjadi dua komponen, yaitu model *count* untuk μ_i dan model *zero inflation* untuk π_i . Tabel 4 menunjukkan taksiran parameter untuk model *Zero Inflated Negative Binomial*. Pengujian kesesuaian model ZINB dilakukan dengan melihat nilai *Likelihood Ratio Test* (LRT). Dari tabel 4, dapat dilihat nilai LRT sebesar 22,83 dan nilai $\chi^2_{0,1;11} = 17,27$. Karena nilai LR lebih besar dari nilai $\chi^2_{0,1;11}$, maka belum cukup bukti untuk menolak H_0 , artinya paling tidak ada satu peubah prediktor yang berpengaruh signifikan terhadap peubah respon. Berdasarkan pengujian secara individu dengan taraf signifikan 10% didapatkan hanya intersep dan parameter peubah persentase balita yang mendapat imunisasi DPT (X_2) yang berpengaruh signifikan terhadap jumlah kasus difteri di Sumatera Barat pada model *count*. Sedangkan pada model *zero* tidak ada satupun parameter yang signifikan.

Model *Hurdle Poisson* terdiri dari model logit dan model *truncated Poisson*. Uji kesesuaian model ini dilakukan dengan melihat nilai *Likelihood Ratio Test* (LRT). Dari tabel 5, dapat dilihat nilai LRT sebesar 20,77 dan nilai $\chi^2_{0,1;10} = 15,98$. Karena nilai LR lebih besar dari nilai $\chi^2_{0,1;10}$, maka belum cukup bukti untuk menolak H_0 , artinya paling tidak ada satu peubah prediktor yang berpengaruh signifikan terhadap peubah respon. Berdasarkan pengujian secara individu dengan taraf signifikan 10% didapatkan hampir semua parameter peubah prediktor pada model *Count* berpengaruh signifikan terhadap jumlah kasus difteri di Sumatera Barat, kecuali peubah kepadatan penduduk (X_3). Sedangkan pada model *Zero Hurdle* tidak ada satupun parameter yang signifikan.

Tabel 5. Taksiran Parameter Model *Hurdle Poisson*

Variabel	Taksiran parameter	Standar eror	Nilai-p	Keterangan
Model Truncated				
Intersep	-8,2563	4,7590	0,0828	Signifikan
Z_1	-7,6601	4,4755	0,0870	Signifikan
Z_2	14,6744	7,3341	0,0454	Signifikan
Z_3	2,9276	2,0581	0,1549	Tidak signifikan
Z_4	1,6619	0,9608	0,0837	Signifikan
Model Zero Hurdle				
Intersep	0,5820	0,4985	0,243	Tidak signifikan
Z_1	-0,0603	0,7658	0,937	Tidak signifikan
Z_2	-0,1531	0,5947	0,797	Tidak signifikan
Z_3	0,7366	0,9630	0,444	Tidak signifikan
Z_4	-0,4476	0,8843	0,613	Tidak signifikan
AIC = 61,5471		Log L = -20,77		df = 10

Model *Hurdle Negative Binomial* terdiri dari model logit dan model *truncated Negative Binomial*. Uji kesesuaian model ini dilakukan dengan melihat nilai *Likelihood Ratio Test* (LRT). Dari tabel 6, dapat dilihat nilai LRT sebesar 21,62 dan nilai $\chi^2_{0,1;11} = 17,27$. Karena nilai LR lebih besar dari nilai $\chi^2_{0,1;11}$, maka belum cukup bukti untuk menolak H_0 , artinya paling tidak ada satu peubah prediktor yang berpengaruh signifikan terhadap peubah respon. Namun berdasarkan pengujian secara individu dengan taraf signifikan 10% tidak ada

satupun parameter peubah prediktor, baik pada model *count* maupun model *zero Hurdle*, yang berpengaruh signifikan terhadap jumlah kasus difteri di Sumatera Barat. Hal ini disebabkan karena nilai standar eror dari Log (theta) yang tidak tersedia.

Tabel 6. Taksiran Parameter Model *Hurdle Negative Binomial*

Variabel	Taksiran parameter	Standar eror	Nilai-p	Keterangan
Model Truncated				
Intersep	-7,958	6,172	0,197	Tidak signifikan
Z_1	-7,243	5,817	0,213	Tidak signifikan
Z_2	14,193	9,588	0,139	Tidak signifikan
Z_3	2,693	2,741	0,326	Tidak signifikan
Z_4	1,479	1,183	0,211	Tidak signifikan
Log (theta)	1,871	NA	NA	Hasil tidak tersedia
Model Zero Hurdle				
Intersep	0,5820	0,4985	0,243	Tidak signifikan
Z_1	-0,0603	0,7659	0,937	Tidak signifikan
Z_2	-0,1531	0,5947	0,797	Tidak signifikan
Z_3	0,7366	0,9630	0,444	Tidak signifikan
Z_4	-0,4476	0,8843	0,613	Tidak signifikan
$AIC = 65,23$		$\text{Log L} = -21,62$		$df = 11$

Pemilihan model terbaik dilakukan dengan melihat nilai AIC dari masing-masing model yang disajikan pada tabel 7. Berdasarkan nilai AIC dari masing-masing model, dapat disimpulkan bahwa model terbaik adalah model *Hurdle Poisson*, karena model ini memiliki nilai AIC yang lebih kecil dibandingkan model-model lainnya. Hal ini menunjukkan bahwa model *Hurdle Poisson* lebih baik digunakan pada kasus penyakit difteri di Provinsi Sumatera Barat tahun 2018.

Tabel 7. Nilai AIC

Model	AIC
ZIP	68,60
ZINB	67,66
<i>Hurdle Poisson</i>	61,54
<i>Hurdle Negative Binomial</i>	65,23

Model *Hurdle Poisson* untuk kasus penyakit difteri di Provinsi Sumatera Barat tahun 2018 adalah sebagai berikut:

$$\text{logit} \left(\frac{\pi_i}{1 - \pi_i} \right) = 0,58 - 0,06Z_1 - 0,15Z_2 + 0,74Z_3 - 0,45Z_4$$

$$\log(\mu_i) = -8,26 - 7,66Z_1 + 14,67Z_2 + 2,93Z_3 + 1,66Z_4$$

Berdasarkan hasil pendugaan parameter pada model logit dapat diketahui bahwa setiap penambahan rumah tangga yang memiliki sanitasi layak akan menurunkan peluang terjadinya kasus difteri sebesar $e^{-0,06}$ kali tanpa melibatkan peubah lain. Setiap penambahan bayi yang mendapatkan imunisasi DPT akan menurunkan peluang terjadinya kasus difteri sebesar $e^{0,15}$ kali tanpa melibatkan peubah lain. Setiap penambahan kepadatan penduduk per km² akan menaikkan

peluang terjadinya kasus difteri sebesar $e^{0,74}$ kali tanpa melibatkan peubah lain. Setiap penambahan penduduk yang memiliki akses air bersih akan menurunkan peluang terjadinya kasus difteri sebesar $e^{-0,45}$ kali tanpa melibatkan peubah lain. Berdasarkan hasil pendugaan parameter pada model *truncated* Poisson dapat diketahui bahwa setiap penambahan rumah tangga yang memiliki sanitasi layak akan menurunkan peluang terjadinya kasus difteri sebesar $e^{-7,66}$ kali tanpa melibatkan peubah lain. Setiap penambahan bayi yang mendapatkan imunisasi DPT akan menaikkan peluang terjadinya kasus difteri sebesar $e^{14,67}$ kali tanpa melibatkan peubah lain. Setiap penambahan kepadatan penduduk per km² akan menaikkan peluang terjadinya kasus difteri sebesar $e^{2,93}$ kali tanpa melibatkan peubah lain. Setiap penambahan penduduk yang memiliki akses air bersih akan menaikkan peluang terjadinya kasus difteri sebesar $e^{1,66}$ kali tanpa melibatkan peubah lain.

5. KESIMPULAN

Berdasarkan pengujian parameter secara individu, model *Hurdle Poisson* memberikan perfoma yang lebih baik dibandingkan model *Zero-Inflated Poisson*, *Zero-Inflated Negative Binomial*, dan *Hurdle Negative Binomial*. Pada taraf signifikan 10% didapatkan hampir semua parameter peubah prediktor pada model *truncated* Poisson berpengaruh signifikan terhadap jumlah kasus difteri di Sumatera Barat, kecuali peubah kepadatan penduduk (X_3). Namun tidak satupun peubah prediktor memberikan pengaruh yang signifikan terhadap peubah respon pada model logit (pemodelan yang memodelkan nilai observasi yang bernilai nol) dalam model *Hurdle Poisson*. Begitu juga dengan nilai AIC keempat model, model *Hurdle Poisson* memberikan nilai AIC yang lebih kecil dibandingkan model lainnya. Sehingga model *Hurdle Poisson* lebih baik digunakan pada kasus penyakit difteri di Provinsi Sumatera Barat tahun 2018 dibandingkan model lainnya yang dibahas pada penelitian ini.

REFERENSI

- [1]. Afri, L. A., Arcat. (2019). Pemodelan Regresi Hurdle Negative Binomial Tersensor Pada Kasus Penyakit Difteri. *Jurnal Absis vol 2 no 1*, 148-158.
- [2]. Badan Pusat Statistik (2019). *Provinsi Sumatera Barat dalam Angka 2019*. Sumatera Barat: BPS Sumatera Barat.
- [3]. Cameron, A.C., & Trivedi, P.K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- [4]. Cantoni, E., & Zedini, A. (2010). A Robust Version of the Hurdle Model. *Journal of Statistical Planning and Inference*. 141(3), 1214-1223.
- [5]. Consul, P.C. & Famoye, F. (1992). Generalized Poisson Regression Model. *Communications in Statistics - Theory and Methods*, 21:1, 89-109
- [6]. Cox, D.R. (1983). Some Remarks on Over-Dispersion. *Biometrika*, 70(1), 269 - 274.

- [7]. Desjardins CD. (2013). *Evaluating the Performance of Two Competing Models Of School Suspension Under Simulation-The Zero-Inflated Negative Binomial and the Negative Binomial Hurdle* [disertasi]. Minnesota (US): Minnesota University.
- [8]. Dobson, J.A. (1983). *An Introduction To Statistical Modelling*, USA: Chapman & Hall/CRC.
- [9]. Famoye, F., Wulu, J.T. and Singh, K.P. (2004). On The Generalized Poisson Regression Model with an Application to Accident Data. *Journal of Data Science*, 2 (2004) 287-295
- [10]. Hilbe, J.M., (2011). *Negative Binomial Regression. Second Edition*, New York: Cambridge University Press.
- [11]. Jamilatuzzahro, Caraka, R. E., dan Herliansyah, R. (2018). *Aplikasi Generalized Linear Model pada R*, Yogyakarta: Innosain.
- [12]. Mullahy, J. (1986). Specification and Testing of Some Modified Count Data Modified Count Data Models. *Journal of Econometrics*.33: 341-365.
- [13]. Rahayu, R.L., Asrof, A., Rustiana, S., Puspitasari, W., Suparman, Y., (2018). 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 vol 4 no 1*. 16-25.
- [14]. Ramadhan, M.N. (2019). *Pemodelan Jumlah Penyakit Kusta Di Provinsi Sulawesi Tenggara Menggunakan Metode Regresi Poisson Inverse Gaussian*. Undergraduate thesis, Muhammadiyah University, Semarang.
- [15]. Saffari, S.E., Adnan R. (2011). Zero-Inflated Negative Binomial Regression Model with Right Censoring Count Data. *Journal of Materials Science and Engineering B*. 1:551-554.