

PEMODELAN JUMLAH KEMATIAN AKIBAT DIFTERI DI PROVINSI JAWA TIMUR DENGAN REGRESI BINOMIAL NEGATIF DAN *ZERO-INFLATED* POISSON

Nurul Fitriyah¹, Alfian F. Hadi², Yuliani Setia Dewi³

^{1,2,3} Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Jember

²E-mail: afhadi@unej.ac.id

Abstrak

Penyakit Difteri merupakan salah satu penyakit menular yang berbahaya, karena terdapat 37 kasus kematian dari 955 kasus. Bakteri *Corynebacterium diphtheriae* menyerang saluran pernafasan atas, racun menyebar melalui darah dan dapat menyebabkan kerusakan jaringan di seluruh tubuh terutama jantung dan saraf. Analisis regresi yang digunakan untuk variabel takbebas berupa data *count* adalah analisis regresi Poisson, namun seringkali terjadi overdispers pada regresi Poisson. Hal ini dapat diatasi dengan menggunakan regresi Binomial Negatif, namun seringkali overdispersi pada data cacahan dapat disebabkan oleh *excess zeros* dan untuk mengatasinya digunakan regresi *Zero-Inflated* Poisson (ZIP). Keterkaitan antara prosentase cakupan desa/kelurahan UCI, jumlah kasus gizi buruk, prosentase masyarakat miskin dan hampir miskin, prosentase rumah tangga yang berperilaku hidup bersih dan sehat, serta jumlah puskesmas dengan banyaknya kematian akibat penyakit difteri dapat didekati dengan analisis statistika yang mengkaji tentang hubungan variabel takbebas dan variabel bebas, yaitu analisis regresi. Langkah-langkah dalam penelitian ini adalah, pertama melakukan kajian pustaka tentang difteri. Kedua, melakukan pengujian model regresi Poisson pada data. Ketiga, mengidentifikasi overdispersi serta *excess zeros*. Keempat melakukan pengujian model regresi Binomial Negatif dan ZIP secara *saturated* dan *full* model dengan bantuan program R. Langkah terakhir membandingkan nilai log-likelihood dari model yang didapatkan untuk mendapatkan model terbaik. Hasil penelitian ini menunjukkan bahwa model terbaik diperoleh dari model regresi ZIP dengan nilai log-likelihood sebesar -29,29.

Kata Kunci: Difteri, *Excess zeros*, *Overdispersi*, *Regresi Binomial Negatif*, *Regresi Zero-Inflated Poisson*.

1 Pendahuluan

Kesehatan merupakan salah satu unsur kesejahteraan bagi masyarakat, melalui pembangunan kesehatan diharapkan dapat meningkatkan derajat kesehatan masyarakat dalam rangka memperbaiki kualitas hidup manusia. Pembangunan Kesehatan di Indonesia mempunyai beban ganda, dimana penyakit menular masih menjadi masalah. Salah satu upaya pencegahannya adalah dengan pemberian vaksin. Difteri merupakan salah satu penyakit menular yang telah ditemukan vaksinnya serta telah dijadikan program imunisasi nasional, namun adanya peningkatan jumlah kasus sebanyak dua

kali lipat dalam waktu 1 tahun pada tahun 2012 di Provinsi Jawa Timur menjadikan status difteri sebagai kejadian luar biasa. Provinsi Jawa Timur merupakan penyumbang kasus difteri terbesar di Indonesia yaitu sebesar 74%, meskipun demikian terdapat kabupaten/kota yang jumlah kematian akibat difteri bernilai nol atau tidak ada kematian akibat difteri [1].

Keterkaitan faktor-faktor penyebab banyaknya kematian yang diakibatkan difteri dapat didekati dengan analisis statistika yang mengkaji tentang hubungan variabel takbebas dan variabel bebas, yaitu analisis regresi. Analisis regresi yang digunakan untuk variabel takbebas berupa data *count* adalah analisis regresi Poisson [2]. Pada regresi Poisson terdapat asumsi yang harus dipenuhi yaitu mean dan varian harus bernilai sama, namun seringkali dijumpai data *count* dengan varian lebih besar dari meannya yang disebut overdispersi. Salah satu metode yang dapat digunakan untuk mengatasi overdispersi pada regresi Poisson adalah regresi Binomial Negatif. Pada distribusi Binomial Negatif varian tidak konstan seperti pada model Poisson, sehingga jika data yang dianalisis memiliki varian yang terlalu menyebar lebih baik digunakan model regresi Binomial Negatif [3]. Namun apabila dalam data penelitian terdapat banyak nilai nol maka dapat menyebabkan overdispersi pada regresi Poisson. Model regresi *Zero-Inflated* Poisson merupakan suatu metode untuk memodelkan data yang memiliki *excess zeros*, model ini juga sekaligus menangani overdispersi yang terjadi [4]. Tujuan dari penelitian ini adalah untuk mendapatkan model terbaik terhadap jumlah kematian akibat difteri tahun 2012 di Provinsi Jawa Timur dengan menggunakan regresi Binomial Negatif dan *Zero-Inflated* Poisson.

2 OVERDISPERSI: REGRESI BINOMIAL NEGATIF DAN REGRESI ZERO-INFLATED POISSON

Suatu peristiwa *count* mengacu pada berapa kali peristiwa itu terjadi. Metode utama yang dikembangkan untuk memahami distribusi dari data cacah (*count*) adalah regresi Poisson, dan kemudian menjadi metode standar yang digunakan untuk memodelkan data cacah [5]. Peubah acak Y dikatakan mengikuti distribusi Poisson, jika fungsi kepadatan peluangnya berbentuk:

$$p(y) = \frac{e^{-\mu} \mu^y}{y!}$$

Pada regresi Poisson terdapat asumsi yang harus dipenuhi yaitu mean dan varian harus bernilai sama, namun seringkali dijumpai data *count* dengan varian lebih besar dari meannya yang disebut overdispersi [6].

Overdispersi pada regresi Poisson dapat menyebabkan *standard error* dari taksiran parameter regresi yang dihasilkan memiliki kecenderungan untuk menjadi lebih rendah dari seharusnya. Overdispersi pada regresi Poisson dapat dideteksi dengan melihat nilai dari *pearson Chi-square* dan *residual deviance* yang dibagi dengan derajat bebasnya. Apabila kedua nilai ini lebih besar dari satu maka dikatakan overdispersi pada data [7]. Salah satu metode yang dikembangkan untuk mengatasi masalah overdispersi adalah model regresi Binomial Negatif.

Model regresi Binomial Negatif dapat digunakan baik dalam keadaan *equidispersion* ataupun overdispersi. Model regresi Binomial Negatif yang dibangun memiliki sebaran Binomial Negatif dengan parameter μ dan k , dimana $\mu = \alpha\beta$ dan $k = 1/\alpha$, sehingga mean dan varian masing-masing adalah sebagai berikut.

$$E(Y) = \mu \quad \text{dan} \quad \text{Var}(Y) = \mu + k\mu^2$$

Sehingga sebaran Y menjadi:

$$f(y, \mu, k) = \frac{\Gamma(y + k^{-1})}{y! \Gamma(k^{-1})} \left(\frac{k\mu}{1 + k\mu}\right)^y \left(\frac{1}{1 + k\mu}\right)^{1/k}$$

Salah satu permasalahan regresi Poisson yaitu *excess zeros*, dimana jumlah respon “0” yang diamati melebihi jumlah respon “0” yang diperkirakan oleh model. Data yang overdispersi terkadang disebabkan oleh *excess zeros*. Dalam hal ini, munculnya kelebihan nol dapat diatasi dengan menggunakan model yang mengakomodasi overdispersi. Dalam banyak kasus nilai nol memiliki arti penting dalam penelitian yang bersangkutan, apabila nilai nol memiliki arti penting dalam penelitian maka data tersebut harus dimasukkan dalam analisis. *Excess zeros* dapat dilihat pada proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya [8].

Model *Zero-Inflated Poisson* pertama kali diperkenalkan oleh Lambert (1992). Jika Y_i merupakan variabel acak takbebas yang berdistribusi *Zero-Inflated Poisson*, maka penelitian nol dapat dikembangkan dalam dua langkah, yaitu [9]:

$$Y_i \sim \begin{cases} 0, & \text{dengan peluang } \omega \\ \text{Poisson}(\lambda_i), & \text{dengan peluang } (1 - \omega) \end{cases}$$

dengan mean dan variannya adalah sebagai berikut:

$$E(Y_i) = (1 - \omega)\lambda_i = \mu_i \quad \text{dan} \quad \text{var}(Y_i) = \mu_i + \left(\frac{\omega}{1-\omega}\right)\mu_i^2$$

Jika Y_i merupakan variabel acak independen yang berdistribusi 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 dan menghasilkan hanya observasi bernilai nol, sementara keadaan kedua disebut *Poisson state* terjadi dengan probabilitas $(1 - \omega)$. Sehingga model regresi *Zero-Inflated Poisson* didefinisikan sebagai berikut:

$$P(Y_i = y_i) = \begin{cases} \omega + (1 - \omega)e^{-\lambda} & \text{untuk } Y_i = 0 \\ (1 - \omega)\frac{e^{-\lambda}\lambda^{y_i}}{y_i!} & \text{untuk } Y_i = 1, 2, \dots, n \end{cases}$$

yang dinotasikan dengan $Y_i \sim ZIP(\lambda, \omega)$, dan untuk memodelkan ω umumnya digunakan model logit, yaitu:

$$\omega = \frac{\exp(\mathbf{Z}_i^T \boldsymbol{\lambda})}{1 + \exp(\mathbf{Z}_i^T \boldsymbol{\lambda})}$$

Untuk menerapkan model *Zero-Inflated Poisson* dalam model yang lebih praktis, Lambert (1992) menyarankan hubungan model untuk dan adalah sebagai berikut.

$$\log(\lambda) = \mathbf{x}_i^T \boldsymbol{\beta} \quad \text{dan} \quad \text{logit}(\omega) = \log\left(\frac{\omega}{1-\omega}\right) = \mathbf{x}_i^T \boldsymbol{\gamma}$$

dimana \mathbf{x} merupakan matriks kovarian sedangkan $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ adalah matriks berukuran $(p + 1) \times 1$ dan $(q + 1) \times 1$ dari parameter yang tidak diketahui atau yang akan ditaksir.

3 Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder. Data tersebut diperoleh dari Dinas Kesehatan Provinsi Jawa Timur. Data tersebut merupakan data-data setiap kabupaten/kota di Provinsi Jawa Timur tahun 2012. Data yang digunakan adalah data jumlah kematian yang diakibatkan oleh difteri sebagai variabel respon (Y), dengan variabel-variabel prediktor (x) meliputi:

1. Cakupan desa/kelurahan UCI (x_1)
2. Jumlah kasus gizi buruk (x_2)
3. Prosentase masyarakat miskin dan hampir miskin (x_3)
4. Rumah tangga berperilaku hidup bersih dan sehat (x_4)
5. Jumlah Puskesmas (x_5)

Pengolahan data dalam penelitian ini menggunakan *software* R, dan paket yang digunakan adalah paket MASS dan *pscl*. Langkah-langkah yang dilakukan dalam menyelesaikan penelitian ini adalah sebagai berikut:

1. Melakukan kajian pustaka tentang KLB difteri di Provinsi Jawa Timur tahun 2012 serta menentukan faktor apa saja yang diduga mempengaruhi jumlah kematian akibat difteri
2. Memodelkan jumlah kematian akibat difteri menggunakan model regresi Poisson dengan *software* program R
3. Mengidentifikasi overdipersi dan *excess zeros*
4. Pengujian model regresi Binomial Negatif dengan menggunakan *software* program R secara *full* dan *saturated* model
5. Pengujian model regresi *Zero-Inflated* Poisson dengan menggunakan *software* program R secara *full* dan *saturated* model
6. Membandingkan model-model yang telah didapatkan pada pengujian model regresi Binomial Negatif dan *Zero-Inflated* Poisson dengan melihat nilai *Log-likelihood* sehingga didapatkan model terbaik.

4 Hasil dan Pembahasan

Model regresi Poisson, Binomial Negatif dan *Zero-Inflated* Poisson dapat dibentuk dari variabel takbebas dan beberapa variabel bebas yang diduga mempengaruhi. Pemodelan jumlah kematian akibat difteri mempunyai beberapa kombinasi model yang diperoleh dari *full* dan *saturated* model.

Pengujian kesesuaian model regresi pada penelitian ini menggunakan uji rasio *likelihood* yang didefinisikan sebagai berikut [10]:

$$G = -2 \log \left[\frac{L(y; \hat{\omega})}{L(y; \hat{\Omega})} \right]$$

Perumusan hipotesis pengujian kesesuaian model regresi Binomial Negatif adalah sebagai berikut [11]:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_k = 0$$

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

statistik ujinya adalah:

$$G = 2 \sum_{i=1}^n \left\{ y_i \log \left(\frac{y_i}{\mu_i} \right) - (y_i + \alpha^{-1}) \log \left(\frac{1 + \alpha y_i}{1 + \alpha \mu_i} \right) \right\}$$

$$G \sim \chi^2$$

Sedangkan perumusan hipotesis untuk model regresi ZIP adalah sebagai berikut [12]:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = \gamma_1 = \gamma_2 = \dots = \gamma_p = 0$$

$$H_1: \text{paling sedikit ada satu } \beta_j \text{ atau } \gamma_j \neq 0, j = 1, 2, \dots, p$$

Dengan statistik uji sebagai berikut.

$$G = \left(2 \sum_{i=1}^n z_i \mathbf{x}_i^T \hat{\boldsymbol{\gamma}} - \log(1 + \exp(\mathbf{x}_i^T \hat{\boldsymbol{\gamma}})) + 2 \sum_{i=1}^n (1 - z_i)(y_i \mathbf{x}_i^T \hat{\boldsymbol{\beta}}) \right) - \left(2 \sum_{i=1}^n z_i \hat{y}_0 - \log(1 + \mathbf{x}_i^T \hat{\boldsymbol{\gamma}}) + 2 \sum_{i=1}^n (1 - z_i)(y_i \hat{\beta}_0 - \exp(\hat{\beta}_0)) \right)$$

Pengujian parameter pada model regresi ZIP terbagi menjadi dua yaitu model log dan model logit. Perumusan hipotesis untuk model log adalah sebagai berikut:

$$H_0: \beta_r = 0, r = 1, 2, \dots, k$$

$$H_1: \beta_r \neq 0$$

Sedangkan perumusan untuk model logit adalah sebagai berikut:

$$H_0: \gamma_r = 0, r = 1, 2, \dots, k$$

$$H_1: \gamma_r \neq 0$$

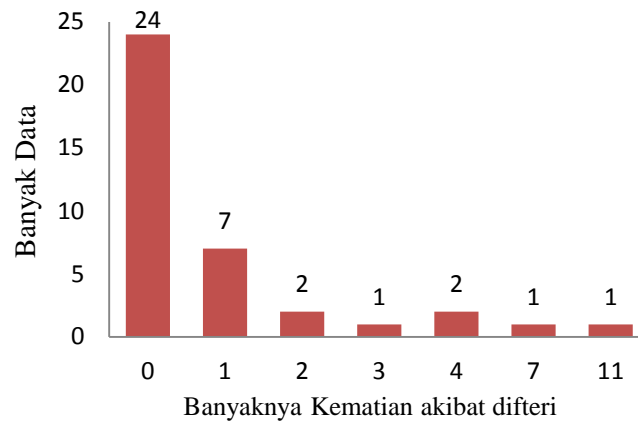
Model regresi Poisson untuk *full* model adalah sebagai berikut:

$$\mu = \exp(-1,2006086 - 0,0177546x_{1i} + 0,0394781x_{2i} + 0,0394781x_{3i} - 0,0030006x_{4i} + 0,0365877x_{5i})$$

dengan nilai *log-likelihood* dari *full* model Poisson adalah sebesar -60,9902.

Asumsi yang harus dipenuhi dari regresi Poisson adalah kesetaraan mean dan varian. Apabila variabel respon mengalami overdispersi maka model regresi Poisson tidak sesuai. Taksiran dispersi dapat diukur dengan nilai *Residual deviance* yang dibagi dengan derajat bebasnya, apabila nilai taksiran dispersi lebih dari 1 maka ada indikasi overdispersi. *Full* model dari regresi Poisson di atas mengalami overdispersi, karena nilai taksiran dispersi dari model tersebut adalah 2,662. Seluruh kombinasi model regresi Poisson yang diperoleh dari *saturated* model diketahui mengalami overdispersi, karena nilai-nilai dari taksiran dispersi model-model tersebut lebih besar dari satu.

Salah satu penyebab terjadinya overdispersi adalah adanya *excess zeros*. *Excess zeros* dapat dilihat dari proporsi variabel respon yang bernilai nol lebih besar dari data diskrit lainnya [9]. Berikut disajikan grafik proporsi jumlah kematian akibat difteri tahun 2012 di Provinsi Jawa Timur:



Gambar 1 Grafik Proporsi Data

Dari Gambar 1 diketahui bahwa proporsi variabel respon yang bernilai nol adalah sebanyak 24 data atau 75% dari data kematian tersebut bernilai nol, sehingga disimpulkan bahwa data tersebut mengalami *excess zeros*. Selain *excess zeros*, data juga diperiksa terdapat *outlier* atau tidak. Karena jika dalam data terdapat *outlier* akan menyebabkan model tidak sesuai untuk menggambarkan data. Untuk memeriksa terdapat *outlier* digunakan bantuan program R dengan fungsi sebagai berikut:

```
> outlier.test(glm(Y~x1+x2+x3+x4+x5, family=poisson(log),
data=DIFTERI))
```

```
max|rstudent| = 6.588096,
unadjusted p = 4.455036e-11, Bonferroni p = 1.692914e-09
```

```
Observation: 17
```

Dari keluaran program tersebut diketahui bahwa pengamatan ke-17 dideteksi sebagai outlier, sehingga harus dikeluarkan.

Untuk mengatasi masalah overdispersi dan *excess zeros* model regresi Poisson tersebut, maka perlu dilakukan analisis dengan model regresi Binomial Negatif dan ZIP. Berikut ini adalah model regresi Binomial Negatif untuk *full* model setelah pengamatan ke-17 dikeluarkan:

$$\mu = \exp(-4,0564935 - 0,0085024x_{1i} + 0,0011643x_{2i} + 0,0790228x_{3i} + 0,0001641x_{4i} + 0,0300732x_{5i})$$

dengan nilai *log-likelihood* dari model di atas adalah -34,83748. Untuk mendapatkan terbaik dari model Binomial Negatif maka akan dibandingkan nilai-nilai *log-likelihood* dari model yang didapatkan.

Berikut disajikan tabel hasil penelitian tentang pengujian hipotesis untuk kesesuaian model regresi Binomial Negatif.

Tabel 1. Rangkuman Model Regresi Binomial Negatif dengan Dua Variabel Bebas

No.	Variabel	Log- Likelihood	G_{Hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	P-value	Sig
1	x_1	-40,91248	12,15	4	9,488	Tolak H_0	0,308	ns
	x_2						0,025	*
2	x_1	-37,96837	6,26178			Terima H_0	0,606094	ns
	x_3						0,000377	***
3	x_1	-43,63126	17,5876			Tolak H_0	0,547	ns
	x_4						0,753	ns
4	x_1	-43,47221	17,2695			Tolak H_0	0,887	ns
	x_5						0,472	ns
5	x_2	-36,1389	2,60284			Terima H_0	0,038311	*
	x_3						0,000725	***
6	x_2	-41,20467	12,7344	Tolak H_0	0,0286	*		
	x_4				0,6826	ns		
7	x_2	-41,06224	12,4495	Tolak H_0	0,0519	ns		
	x_5				0,4792	ns		
8	x_3	-37,30031	4,92566	Terima H_0	0,000152	***		
	x_4				0,192280	ns		
9	x_3	-36,65963	3,6443	Terima H_0	0,000150	***		
	x_5				0,131667	ns		
10	x_4	-43,22947	16,7839	Tolak H_0	0,512	ns		
	x_5				0,262	ns		

Berdasarkan Tabel 1, dari sepuluh model regresi Binomial Negatif dengan dua variabel bebas yang dibentuk menghasilkan empat model yang sama dengan *full* model, karena $G_{Hitung} < \chi^2_{v,\alpha}$ sehingga keputusan yang diperoleh dari model-model tersebut adalah Terima H_0 . Selanjutnya diseleksi untuk memperoleh model terbaik dari model-model tersebut, model dengan nilai *log-likelihood* terbesar dan tingkat signifikansi yang tinggi dapat dipilih sebagai model terbaik. Berdasarkan Tabel 1, dapat diketahui bahwa model terbaik regresi Binomial Negatif dengan dua variabel adalah model dengan variabel bebas x_2 dan x_3 dengan nilai *log-likelihood* -36,1389.

Tabel 2. Rangkuman Model Regresi Binomial Negatif dengan Tiga Variabel Bebas

No.	Variabel	Log- Likelihood	G_{Hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	P-value	Sig
1	x_1	-35,74402	1,81308	5	11,070	Terima H_0	0,344923	ns
	x_2						0,023435	*
	x_3						0,00733	***
2	x_1	-40,55702	11,43908			Tolak H_0	0,183	ns
	x_2						0,015	*
	x_4						0,401	ns
3	x_1	-40,83042	11,98588	Tolak H_0	0,4208	ns		
	x_2				0,0371	*		
	x_5				0,6707	ns		
4	x_1	-37,29976	4,92456	Terima H_0	0,971500	ns		
	x_3				0,000163	***		
	x_4				0,232466	ns		
5	x_1	-36,65935	3,64374	Terima H_0	0,979301	ns		
	x_3				0,000158	***		
	x_5				0,134635	ns		
6	x_1	-43,18819	16,70142	Tolak H_0	0,732	ns		
	x_4				0,480	ns		
	x_5				0,308	ns		

No.	Variabel	Log- Likelihood	G_{Hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	P-value	Sig
7	x_2	-35,66989	1,66482			Terima H_0	0,054086	ns
	x_3						0,000453	***
	x_4						0,303815	ns
8	x_2	-35,02608	0,3772			Terima H_0	0,065482	ns
	x_3						0,000398	***
	x_5						0,163858	ns
9	x_2	-40,62233	11,5697			Tolak H_0	0,0382	*
	x_4						0,3609	ns
	x_5						0,2834	ns
10	x_3	-36,52124	3,36752			Terima H_0	0,000168	***
	x_4						0,681314	ns
	x_5						0,263999	ns

Berdasarkan Tabel 2, dari kesepuluh model regresi Binomial Negatif dengan tiga variabel bebas keputusan yang diperoleh adalah Terima H_0 untuk enam model, maka model-model tersebut dianggap sama dengan *full* model. Selanjutnya diseleksi untuk memperoleh model terbaik dari model-model tersebut, model dengan nilai log-likelihood terbesar dan tingkat signifikansi yang tinggi dapat dipilih sebagai model terbaik. Berdasarkan Tabel 2, dapat diketahui bahwa model terbaik regresi Binomial Negatif dengan tiga variabel adalah model dengan variabel bebas bebas x_2, x_3 dan x_5 . Meskipun hanya variabel x_3 saja yang signifikan namun nilai log-likelihood untuk model tersebut adalah yang terbesar yaitu -35,02608.

Tabel 3. Rangkuman Model Regresi Binomial Negatif dengan Empat Variabel Bebas

No.	Variabel	Log- Likelihood	G_{Hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	P-value	Sig
1	x_1	-35,50826	1,34156			Terima H_0	0,535565	ns
	x_2						0,035409	*
	x_3						0,000579	***
	x_4						0,466131	ns
2	x_1	-34,83908	$3,2 \times 10^{-3}$			Terima H_0	0,501168	ns
	x_2						0,039665	*
	x_3						0,000474	***
	x_5						0,198577	ns
3	x_1	-40,18871	10,70246	6	12,592	Terima H_0	0,2675	ns
	x_2						0,0208	*
	x_4						0,2606	ns
	x_5						0,3925	ns
4	x_1	-36,51057	3,34618			Terima H_0	0,87228	ns
	x_3						0,00016	***
	x_4						0,60798	ns
	x_5						0,25515	ns
5	x_2	-34,98806	0,30116			Terima H_0	0,070468	ns
	x_3						0,000543	***
	x_4						0,785169	ns
	x_5						0,273857	ns

Berdasarkan Tabel 3, dari kelima model regresi Binomial Negatif dengan empat variabel bebas keputusan yang diperoleh adalah terima H_0 untuk semua model, maka model-model tersebut dianggap sama dengan *full* model. Selanjutnya diseleksi untuk memperoleh model terbaik dari model-model tersebut, model dengan nilai log-

likelihood terbesar dan tingkat signifikansi yang tinggi dapat dipilih sebagai model terbaik. Berdasarkan Tabel 3, dapat diketahui bahwa model terbaik regresi Binomial Negatif dengan empat variabel adalah model dengan variabel bebas x_1, x_2, x_3 dan x_5 , nilai log-likelihood untuk model tersebut adalah -34,83908.

Dari uraian di atas dapat diketahui model Binomial Negatif terbaiknya adalah model dengan nilai log-likelihood terbesar, model tersebut adalah full model Binomial Negatif dengan nilai log-likelihood -34,83748. Meskipun demikian, model regresi Binomial Negatif dengan empat variabel bebas dan full model memiliki selisih nilai log-likelihood yang sangat kecil, dan keputusan dari uji kesesuaian model adalah terima, sehingga model dengan empat variabel bebas dan dapat dikatakan sebagai model terbaik karena pada full model variabel tidak signifikan.

Metode lain yang digunakan untuk mengatasi overdispersi adalah regresi Zero-Inflated Poisson. Berikut adalah full model dari regresi Zero-Inflated Poisson setelah pengamatan ke-17 dikeluarkan:

$$\log(\mu) = 4,0939156 - 0,0496945x_{1i} + 0,00115216x_{2i} + 0,0068677x_{3i} - 0,0030702x_{4i} - 0,0267398x_{5i}$$

dan

$$\text{logit}(\mu) = 313,0070 - 2,0146x_{1i} + 0,0423x_{2i} - 3,4232x_{3i} - 0,1391x_{4i} - 2,1828x_{5i}$$

dengan nilai log-likelihood dari model ZIP di atas adalah -23,84. Untuk mendapatkan model terbaik dari model ZIP, maka dilakukan langkah yang sama untuk mencari model terbaik dari model Binomial Negatif yaitu dengan membandingkan nilai-nilai log-likelihood dari model ZIP yang didapatkan.

Tabel 4. Rangkuman Model Regresi Zero-Inflated Poisson dengan Dua Variabel Bebas

No.	Variabel	Log-Likelihood	G_{hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	Sig	
							Log	Logit
1	x_1	-41	34,32			Tolak H_0	ns	ns
	x_2						**	ns
2	x_1	-39,2	30,72			Tolak H_0	ns	ns
	x_3						ns	*
3	x_1	-43,59	39,5			Tolak H_0	ns	ns
	x_4						ns	ns
4	x_1	-41,04	34,4			Tolak H_0	ns	ns
	x_5						*	ns
5	x_2	-35,66	23,64	6	12,592	Tolak H_0	*	ns
	x_3						*	ns
6	x_2	-40,58	33,48			Tolak H_0	**	ns
	x_4						ns	ns
7	x_2	-39,02	30,36			Tolak H_0	**	ns
	x_5						ns	ns
8	x_3	-37,09	26,5			Tolak H_0	ns	**
	x_4						ns	ns
9	x_3	-36,38	25,08			Tolak H_0	ns	*
	x_5						ns	*
10	x_4	-42,75	37,82			Tolak H_0	ns	ns
	x_5						ns	ns

Berdasarkan Tabel 4, dari kesepuluh model ZIP di atas keputusan yang diperoleh adalah tolak H_0 untuk semua model dengan dua variabel bebas. Berdasarkan keputusan tersebut, model-model dengan dua variabel bebas dianggap berbeda dengan *full* model ZIP. Selanjutnya diseleksi untuk memperoleh model terbaik dari model-model tersebut, model dengan nilai *log-likelihood* terbesar dan tingkat signifikansi yang tinggi dapat dipilih sebagai model terbaik. Berdasarkan Tabel 4, dapat diketahui bahwa model terbaik regresi ZIP dengan dua variabel adalah model dengan variabel bebas x_2 dan x_3 dengan nilai *log-likelihood* -35,66.

Tabel 5. Rangkuman Model Regresi *Zero-Inflated* Poisson dengan Tiga Variabel Bebas

No.	Variabel	Log- <i>Likelihood</i>	G_{Hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	Sig	
							Log	Logit
1	x_1	-32,26	16,84			Tolak H_0	ns	ns
	x_2						**	ns
	x_3						**	ns
2	x_1	-32,65	17,62			Tolak H_0	**	ns
	x_2						***	ns
	x_4						**	ns
3	x_1	-34,94	22,2			Tolak H_0	**	ns
	x_2						***	ns
	x_5						**	ns
4	x_1	-35,37	23,06			Tolak H_0	ns	ns
	x_3						ns	ns
	x_4						ns	ns
5	x_1	-33,8	19,92	8	15,507	Tolak H_0	ns	ns
	x_3						ns	ns
	x_5						ns	ns
6	x_1	-40,71	33,74			Tolak H_0	**	ns
	x_4						ns	*
	x_5						ns	ns
7	x_2	-34,2	20,72			Tolak H_0	**	ns
	x_3						ns	*
	x_4						ns	ns
8	x_2	-33,37	19,06			Tolak H_0	***	ns
	x_3						*	ns
	x_5						ns	ns
9	x_2	-38,95	30,22			Tolak H_0	**	ns
	x_4						ns	ns
	x_5						ns	ns
10	x_3	-35,79	23,9			Tolak H_0	ns	*
	x_4						ns	ns
	x_5						ns	ns

Berdasarkan Tabel 5, dari kesepuluh model ZIP di atas keputusan yang diperoleh adalah tolak H_0 untuk semua model dengan tiga variabel bebas. Berdasarkan keputusan tersebut, model-model dengan dua variabel bebas dianggap berbeda dengan *full* model ZIP. Selanjutnya diseleksi untuk memperoleh model terbaik dari model-model tersebut, model dengan nilai *log-likelihood* terbesar dan tingkat signifikansi yang tinggi dapat dipilih sebagai model terbaik. Berdasarkan Tabel 5, dapat diketahui bahwa model terbaik regresi ZIP dengan tiga variabel adalah model dengan variabel bebas x_1 , x_2 dan x_3 dengan nilai *log-likelihood* -32,26.

Tabel 6 Rangkuman Model Regresi ZIP dengan Empat Variabel Bebas

No.	Variabel	Log-Likelihood	G_{Hitung}	df	$\chi^2_{v,\alpha}$	Keputusan	Sig	
							Log	Logit
1	x_1	-26,76	5,84			Terima H_0	**	ns
	x_2						***	ns
	x_3						ns	ns
	x_4						ns	ns
2	x_1	-25,33	2,98			Terima H_0	**	ns
	x_2						***	ns
	x_3						ns	ns
	x_5						ns	ns
3	x_1	-33,21	18,74	10	18,307	Tolak H_0	**	ns
	x_2						***	ns
	x_4						ns	ns
	x_5						ns	ns
4	x_1	-28,49	9,3			Terima H_0	*	ns
	x_3						ns	ns
	x_4						ns	ns
	x_5						ns	ns
5	x_2	-33,18	18,68			Tolak H_0	**	ns
	x_3						ns	*
	x_4						ns	ns
	x_5						ns	ns

Berdasarkan Tabel 6, dari kelima model ZIP diketahui terdapat dua model dengan keputusan tolak H_0 , model tersebut dianggap berbeda dengan *full* model. Sedangkan untuk tiga model lainnya keputusan yang diperoleh adalah terima H_0 , berarti model tersebut dianggap sama dengan *full* model. Selanjutnya diseleksi untuk memperoleh model terbaik dari model-model tersebut. Model dengan nilai *log-likelihood* terbesar dan tingkat signifikansi yang tinggi dapat dipilih sebagai model terbaik. Berdasarkan Tabel 6, dapat diketahui bahwa model terbaik regresi ZIP dengan empat variabel adalah model dengan variabel bebas x_1, x_2, x_3 dan x_5 dengan nilai *log-likelihood* -25,33.

Dari uraian di atas dapat diketahui model ZIP terbaiknya adalah model dengan nilai *log-likelihood* terbesar, model tersebut adalah *full* model ZIP dengan nilai *log-likelihood* -23,84. Model-model terbaik dari masing-masing metode dapat dirangkum pada Tabel 7.

Tabel 7. Model Terbaik untuk Regresi Poisson, Binomial Negatif dan ZIP

Model Regresi	Log-likelihood
Poisson	-60,99021
Binomial Negatif	-34,83748
ZIP	-23,84

Berdasarkan Tabel 7 dapat diketahui bahwa model terbaik untuk memodelkan jumlah kematian akibat penyakit difteri di Provinsi Jawa Timur pada tahun 2012 adalah dengan regresi ZIP. Nilai *log-likelihood* dari model ZIP lebih besar dari model Poisson maupun Binomial Negatif. Dari uraian tersebut didapatkan model terbaiknya adalah sebagai berikut:

$$\log(\mu) = 4,0939156 - 0,0496945x_{1i} + 0,00115216x_{2i} + 0,0068677x_{3i} - 0,0030702x_{4i} - 0,0267398x_{5i}$$

dan

$$\text{logit}(\mu) = 313,0070 - 2,0146x_{1i} + 0,0423x_{2i} - 3,4232x_{3i} - 0,1391x_{4i} - 2,1828x_{5i}$$

dimana x_1 adalah prosentase cakupan Desa/Kelurahan UCI (*Universal Child Immunization*), x_2 adalah banyaknya kasus gizi buruk, x_3 adalah prosentase masyarakat miskin dan hampir miskin, x_4 adalah prosentase rumah tangga yang berperilaku hidup bersih dan sehat, dan x_5 merupakan banyaknya Puskesmas yang tersedia di setiap Kabupaten/Kota di Provinsi Jawa Timur.

Tabel 8 Interpretasi Model Regresi ZIP

β atau γ	$\exp(\beta)$ atau $\exp(\gamma)$	Pengaruh
Model Log		
$\beta_1 = -0,0496945$	0,95152	Menurun sebesar 4,85%
$\beta_2 = 0,00115216$	1,00153	Meningkatkan sebesar 0,15%
$\beta_3 = 0,0068677$	1,00689	Meningkatkan sebesar 0,69%
$\beta_4 = -0,0030702$	0,99693	Menurun sebesar 0,31%
$\beta_5 = -0,0267398$	0,97362	Menurun sebesar 2,64%
Model Logit		
$\gamma_1 = -2,0146$	0,13337	Menurun sebesar 86,66%
$\gamma_2 = 0,0423$	1,04321	Meningkatkan sebesar 4,32%
$\gamma_3 = -3,4232$	0,03261	Menurun sebesar 96,74%
$\gamma_4 = -0,1391$	0,87014	Menurun sebesar 12,99%
$\gamma_5 = -2,1828$	0,11273	Menurun sebesar 88,73%

Dari Tabel 8 di atas dapat diketahui bahwa terdapat interpretasi yang tidak sesuai dengan kajian pustaka difteri, yaitu pada variabel bebas prosentase masyarakat miskin dan hampir miskin (x_3) pada model logit. Variabel tersebut seharusnya meningkatkan, namun hasil estimasi variabel tersebut menurunkan resiko terjadi kematian akibat difteri. Oleh karena itu, selanjutnya dicari model yang dianggap sama dengan *full* model dan sesuai dengan kajian pustaka tentang difteri. Pada model regresi *Zero-Inflated Poisson*, variabel bebas yang terdapat pada model log dan logit tidak harus sama sehingga dimungkinkan terdapat model yang sesuai dan dapat dianggap sama dengan *full* model untuk menggambarkan data. Dan berikut adalah model regresi ZIP yang sesuai dan dapat dianggap sama dengan *full* model:

$$\text{log}(\mu) = 0,8343938 - 0,04492720x_{1i} + 0,0017490x_{2i} + 0,0475583x_{3i} - 0,002109x_{4i}$$

dan

$$\text{logit}(\mu) = 165,580921 - 2,840897x_{1i} + 0,004202x_{2i} - 0,121245x_{4i}$$

dengan nilai *log-likelihood* -29,29. Interpretasi dari model tersebut adalah sebagai berikut.

Tabel 9 Interpretasi Model Regresi ZIP Terbaik

β atau γ	$\exp(\beta)$ atau $\exp(\gamma)$	Pengaruh
Model Log		
$\beta_1 = -0,0449272$	0,95607	Menurun sebesar 4,39%
$\beta_2 = 0,0017490$	1,00175	Meningkatkan sebesar 0,175%
$\beta_3 = 0,0475583$	1,04871	Meningkatkan sebesar 4,87%
$\beta_4 = -0,002109$	0,99779	Menurun sebesar 0,22%

β atau γ	$\exp(\beta)$ atau $\exp(\gamma)$	Pengaruh
	Model Logit	
$\gamma_1 = -2,840897$	0,05837	Menurun sebesar 94,16%
$\gamma_2 = 0,004202$	1,00421	Meningkatkan sebesar 0,42%
$\gamma_4 = -0,121245$	0,88582	Menurunkan sebesar 11,42%

Dari Tabel 9 di atas dapat diketahui bahwa model log dari regresi ZIP menunjukkan bahwa setiap kenaikan 1% dari prosentase cakupan Desa/Kelurahan UCI (x_1) dapat menurunkan resiko kematian akibat difteri sebesar 4,39%, untuk setiap kenaikan 1 kasus gizi buruk (x_2) berdampak pada peningkatan resiko terjadinya kematian akibat difteri sebesar 0,175%. Untuk setiap kenaikan 1% dari prosentase masyarakat miskin dan hampir miskin (x_3) akan berdampak pada peningkatan terjadinya kematian akibat difteri sebesar 4,87%, dan untuk setiap kenaikan 1% rumah tangga yang berperilaku hidup bersih dan sehat (x_4) dapat menurunkan resiko terjadinya kematian akibat difteri sebesar 0,22%. Dan dari model logit dari regresi ZIP menunjukkan bahwa setiap kenaikan 1% dari prosentase cakupan Desa/Kelurahan UCI (x_1) dapat menurunkan resiko kematian akibat difteri sebesar 94,16%, untuk setiap kenaikan 1 kasus gizi buruk (x_2) berdampak pada peningkatan resiko terjadinya kematian akibat difteri sebesar 0,42%, sedangkan untuk setiap kenaikan 1% rumah tangga yang berperilaku hidup bersih dan sehat (x_4) dapat menurunkan resiko terjadinya kematian akibat difteri sebesar 11,42%.

5 Kesimpulan dan Saran

Berdasarkan hasil analisis dan pembahasan dapat diambil kesimpulan bahwa model regresi ZIP menghasilkan nilai lebih besar dibandingkan dengan model regresi Poisson dan Binomial Negatif. Dari pemodelan regresi Poisson, Binomial Negatif, dan ZIP didapatkan model terbaik untuk memodelkan jumlah kematian akibat difteri di Provinsi Jawa Timur pada tahun 2012 adalah model regresi ZIP dengan variabel bebas model log adalah prosentase cakupan Desa/Kelurahan UCI, jumlah kasus gizi buruk, prosentase masyarakat miskin dan hampir miskin, dan prosentase rumah tangga yang berperilaku bersih dan sehat. Sedangkan pada model logit adalah prosentase cakupan Desa/Kelurahan UCI, jumlah kasus gizi buruk, dan prosentase rumah tangga yang berperilaku bersih dan sehat. Dengan demikian yang perlu ditingkatkan oleh pemerintah setempat untuk dapat menekan terjadinya kematian akibat penyakit difteri adalah cakupan Desa/Kelurahan UCI dan rumah tangga yang berperilaku hidup bersih dan sehat. Selain itu, pemerintah juga perlu menindak lanjuti masalah gizi buruk serta mencari solusi untuk mengurangi prosentase masyarakat miskin dan hampir miskin.

Daftar Pustaka

- [1] Dinas Kesehatan Provinsi Jawa Timur. *Profil Kesehatan Provinsi Jawa Timur Tahun 2012*. Surabaya: Dinas Kesehatan Provinsi Jawa Timur, 2013.
- [2] Berk, R. & MacDonald, J. M. "Overdispersion and Poisson Regression," *Journal Quant Criminol*, vol. 24, pp.269-28, April 2008.
- [3] Ismail, N. & Jemain, A. A. "Generalized Poisson Regression: An Alternative For Risk Classification," *Jurnal Teknologi*, vol. 43, pp. 39-54, 2005.

- [4] Jansakul, N. & Hinde, J. P. "Score Tests for Zero-Inflated Poisson Models," *Computational Statistics & Data Analysis*, vol. 40, pp. 75-96, 2002.
- [5] Hilbe, J. M. *Negative Binomial Regression Second Edition*. New York: Cambridge University Press. 2011.
- [6] Agresti, A. *Categorical Data Analysis Second Edition*. New York: John Wiley and Sons, Inc. 2002.
- [7] Hardin, J. W. dan Hilbe, J. M. *Generalized Linier Models and Extensions*. Texas: Stata Press. 2007.
- [8] Little, T. D. 2013. *The Oxford Handbook of Quantitative Methods, Volume 2 Statistical Analysis*. [serial on line]. http://books.google.co.id/books?id=ulgdl4BPH0C&printsec=frontcover&hl=id&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. [21 Maret 2014].
- [9] Lambert, D. "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing," *Technometrics*, vol. 34, pp. 1-14, February 1992.
- [10] Hosmer, D. W. & Lemeshow, S. *Applied Logistic Regression Second Edition*. John Wiley & Sons, Inc, 2000.
- [11] Wahyuni, W. 2011. "Penaksiran Parameter Model Regresi Binomial Negatif Pada Kasus Overdispersi". Tidak Diterbitkan. Skripsi. Depok: FMIPA Universitas Indonesia.
- [12] Hall, D. B. & Shen, J. "Robust Estimation for Zero-Inflated Poisson Regression," *Scandinavian Journal of Statistics*, **10**.1111/j/1467-9469.2009.00657.x., 2009.