

PEMODELAN *INFANT MORTALITY RATE* (IMR) DENGAN PENDEKATAN *ZERO INFLATED POISSON REGRESSION* BERBASIS ALGORITMA EM

Indah Manfaati Nur¹

¹ FMIPA Universitas Muhammadiyah Semarang, Jl. Kedungmundu Raya No.22 Semarang
Email: indahmnur@unimus.ac.id

Abstract One of the most sensitive benchmarks of all government intervention efforts, especially in the field of health is the Infant Mortality Rate (IMR). If the IMR in a region is high, then health status in the region is low. Efforts continue to be made to reduce the Infant Mortality Rate (IMR) in Indonesia. One way that can be done to reduce the Infant Mortality Rate (IMR) is to examine the factors that cause it. The number of death events is a response variable in the form of discrete data then one of the regression model that can be used is Poisson regression model. A characteristic of the Poisson distribution is the presence of equidispersion, ie the condition in which the mean and variance values of the response variable are equal. In practice, however, there is often an overdispersion state in Poisson regression. The Zero-Inflated Poisson regression model is a method capable of dealing with overdispersion occurring. The purpose of this research is to get the best model about Infant Mortality Rate (IMR) by using Zero-Inflated Poisson based EM Algorithm. The results showed that the factors that significantly influence the IMR are the number of health facilities in each district / city, the ratio of availability of village midwives to each district / city, and the percentage of deliveries assisted by health personnel.

Keywords: *Infant Mortality Rate (IMR), Overdispersion, Zero Inflated Poisson (ZIP) Regression*

Abstrak Salah satu tolok ukur yang sensitif dari semua upaya intervensi yang dilakukan pemerintah khususnya di bidang kesehatan adalah *Infant Mortality Rate (IMR)*. Apabila IMR di suatu wilayah tinggi, berarti status kesehatan di wilayah tersebut rendah. Berbagai upaya terus dilakukan untuk menurunkan *Infant Mortality Rate (IMR)* di Indonesia. Salah satu cara yang dapat dilakukan untuk menurunkan *Infant Mortality Rate (IMR)* adalah dengan mengkaji faktor-faktor penyebabnya. Banyaknya kejadian kematian merupakan variabel respon yang berupa data diskrit maka salah satu model regresi yang dapat digunakan adalah model regresi Poisson. Suatu ciri dari distribusi Poisson adalah adanya *equidispersi*, yakni kondisi dimana nilai mean dan varians dari variabel respon bernilai sama. Namun pada praktiknya, sering ditemukan suatu keadaan overdispersi pada regresi Poisson. Model regresi *Zero-Inflated Poisson* merupakan suatu metode yang mampu untuk menangani overdispersi yang terjadi. Tujuan dari penelitian ini adalah untuk mendapatkan model terbaik tentang *Infant Mortality Rate (IMR)* dengan menggunakan *Zero-Inflated Poisson* berbasis Algoritma EM. Hasil penelitian menunjukkan bahwa faktor-faktor yang berpengaruh secara signifikan terhadap IMR adalah jumlah sarana kesehatan pada tiap kabupaten/kota, rasio ketersediaan bidan desa pada tiap kabupaten/kota, dan presentase persalinan ditolong tenaga kesehatan.

Kata kunci: *Infant Mortality Rate (IMR), Overdispersi, Regresi Zero Inflated Poisson (ZIP)*

1. Pendahuluan

Indonesia sebagai salah satu negara berkembang mempunyai suatu permasalahan yang masih sulit untuk ditanggulangi yaitu *Infant Mortality Rate (IMR)* yang tinggi. *Infant Mortality Rate (IMR)* memberikan kontribusi yang besar pada tingkat mortalitas penduduk suatu negara atau suatu daerah. *Infant Mortality Rate (IMR)* atau Angka Kematian Bayi merupakan jumlah kematian bayi (0-11 bulan) per 1000 kelahiran hidup dalam kurun waktu satu tahun. Apabila IMR di suatu wilayah tinggi, berarti status kesehatan di wilayah tersebut rendah. Diperkirakan setiap jam 18 bayi dan 24 balita di Indonesia meninggal dunia. Berdasarkan Survei Demografi dan Kesehatan Indonesia (SDKI, 2012), angka kematian bayi (AKB) di Indonesia yaitu 32 bayi

per 1000 kelahiran. Bila dirincikan 157.000 bayi meninggal per tahun atau 430 bayi per hari. (Profil Kesehatan Indonesia, 2012)

Berbagai upaya harus dilakukan untuk menurunkan *Infant Mortality Rate* (IMR) di Indonesia. Salah satu cara yang dapat dilakukan untuk menurunkan IMR adalah dengan mengetahui faktor-faktor penyebabnya. Beberapa penelitian sebelumnya juga telah menghasilkan banyak faktor terutama sosial ekonomi yang menyebabkan kematian bayi. Pemodelan IMR pernah dilakukan oleh penelitian sebelumnya [8] dengan menggunakan komponen utama. Berawal dari keadaan tersebut, peneliti ingin mengkaji tentang analisis faktor-faktor yang mempengaruhi *Infant Mortality Rate* (IMR). Keterkaitan faktor-faktor tersebut dengan jumlah kematian bayi dapat didekati oleh analisis statistik yaitu dengan menggunakan pemodelan regresi. Analisis regresi digunakan untuk menganalisis data variabel respon yang berupa data diskrit maupun kontinu. Karena banyaknya angka kematian bayi merupakan variabel respon yang berupa data diskrit maka salah satu model regresi yang dapat digunakan adalah model regresi Poisson. Suatu ciri dari distribusi Poisson adalah adanya *equidispersi*, yakni keadaan dimana nilai mean dan varians dari variabel respon bernilai sama. Namun pada praktiknya, kadang-kadang ditemukan suatu keadaan dimana nilai varians lebih besar dari nilai meannya sehingga menyebabkan overdispersi pada regresi Poisson. Hal ini dapat terjadi karena sumber keragaman yang tak teramati, pada variabel respon ada data yang hilang, adanya penciliran data, adanya interaksi antar pengamatan dan kesalahan spesifikasi.

Data tentang *Infant Mortality Rate* (IMR) pada wilayah Provinsi Jawa Tengah menunjukkan ciri-ciri terjadinya overdispersi Model regresi *Zero-Inflated Poisson* merupakan suatu metode yang mampu untuk menangani overdispersi yang terjadi.

2. Metode

2.1 Metode Analisis

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. Distribusi zip memiliki fungsi probabilitas sebagai berikut : [5]

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

Model gabungan untuk μ dan ω sebagai berikut[6] :

$$\log(\mu) = \mathbf{X}\boldsymbol{\beta} \quad \text{dan} \quad \text{logit}(\omega) = \log\left(\frac{\omega}{1-\omega}\right) = \mathbf{X}\boldsymbol{\gamma}$$

Dimana \mathbf{X} adalah matriks variabel prediktor, $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ adalah parameter yang akan ditaksir, dan ω adalah probabilitas observasi bernilai nol.

Estimasi parameter regresi ZIP dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Fungsi *Log-likelihood* gabungan untuk model regresi ZIP diberikan oleh :

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\gamma} | y_i, x_i) = \sum_{\substack{i=1 \\ y_i=0}}^n \ln(\exp(x_i' \boldsymbol{\gamma}) + \exp(-\exp(x_i' \boldsymbol{\beta}))) - \sum_{i=1}^n \ln(1 + \exp(x_i' \boldsymbol{\gamma})) + \sum_{\substack{i=1 \\ y_i>0}}^n ((y_i x_i' \boldsymbol{\beta}) - \exp(x_i' \boldsymbol{\beta})) - \sum_{\substack{i=1 \\ y_i>0}}^n \ln(y_i!)$$

Estimasi maximum likelihood untuk $\boldsymbol{\beta}$ dan $\boldsymbol{\gamma}$ dapat diperoleh dengan menggunakan pendekatan standar untuk model campuran, yaitu algoritma EM. Algoritma EM memerlukan prosedur sederhana yang dapat diimplementasikan dalam software standar atau metode estimasi langsung seperti metode *Newton Raphson*.

Pengujian kesesuaian model dan pengujian parameter regresi ZIP adalah dengan menggunakan LR (*Likelihood Ratio*) test.

Uji kesesuaian model

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_r = \gamma_1 = \gamma_2 = \dots = \gamma_r = 0$$

$$H_1 : \text{Paling sedikit ada satu } \beta_i \neq 0 \text{ atau } \gamma_i \neq 0$$

dengan β_j adalah parameter model $\ln(\mu) = X\beta$ ke-j, γ_j adalah parameter model $\text{logit}(\omega) = X\gamma$

Statistik Uji :

$$G = 2 \ln \left[\frac{L(y; \hat{\omega})}{L(y; \hat{\eta})} \right] \sim \chi^2_{(2p)} \tag{1}$$

Kriteria uji : Tolak H_0 pada taraf signifikansi α jika $G_{hitung} > \chi^2_{\alpha, 2p}$

Prosedur signifikansi parameter regresi ZIP secara individu

Prosedur untuk menguji signifikansi parameter secara individu sebagai berikut

- a. Uji signifikansi parameter model $\ln(\mu) = X\beta$

Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_0 : \beta_j \neq 0 \text{ untuk suatu } j = 1, 2, \dots, p$$

$$\text{Statistika Uji : } W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2 \sim \chi^2_{(1)} \tag{2}$$

Kriteria uji : Tolak H_0 pada taraf signifikansi α jika $W_j > \chi^2_{\alpha, 1}$

- b. Uji signifikansi parameter model $\text{logit}(\omega) = X\gamma$

Hipotesis :

$$H_0 : \gamma_j = 0$$

$$H_0 : \gamma_j \neq 0 \text{ untuk suatu } j = 1, 2, \dots, p$$

$$\text{Statistika Uji : } W_j = \left(\frac{\hat{y}_j}{SE(\hat{y}_j)} \right)^2 \sim \chi^2_{(1)} \quad (3)$$

Kriteria uji : Tolak H_0 pada taraf signifikansi α jika $W_j > \chi^2_{\alpha,1}$

Pemilihan model terbaik untuk regresi *zero inflated poisson* salah satunya adalah dengan metode AIC (*Akaike's Information Criterion*). Nilai AIC adalah sebagai berikut : $G+(k+1)$

Dimana G adalah statistik uji kesesuaian model dan $k+1$ adalah jumlah parameter. Model terbaik regresi ZIP adalah model dengan AIC terkecil. Menurut Bozdogan (2000), AIC didefinisikan sebagai berikut :

$$AIC = -2\ell + 2p$$

Dengan ℓ adalah nilai ln likelihood dari model dan p adalah banyaknya parameter. Semakin kecil nilai AIC maka model semakin baik.

2.2 Sumber Data dan Variabel Penelitian

Sumber data utama yang digunakan dalam penelitian adalah data sekunder yang diambil dari data Profil Kesehatan Jawa Tengah dan Jawa Tengah dalam Angka yang dirilis oleh BPS (Badan Pusat Statistik).

Adapun variabel respon yang menjadi perhatian dalam penelitian ini adalah angka kematian bayi (AKB) di Provinsi Jawa Tengah. Sedangkan variabel prediktor yang di duga mempengaruhi dan menggambarkan angka kematian bayi antara lain :

- Jumlah sarana kesehatan puskesmas pada tiap kabupaten/kota (X_1)
- Ratio ketersediaan bidan desa pada tiap kabupaten/kota (X_2)
- Persentase persalinan ditolong tenaga kesehatan (X_3)
- Cakupan pelayanan kesehatan bayi (X_4)
- Jumlah penanganan komplikasi kebidanan (X_5)
- Jumlah penanganan komplikasi neonatal (X_6)
- Jumlah bayi yang diberikan ASI eksklusif (X_7)
- Persentase rumah tangga ber-PHBS pada tiap kab/kota (X_8).

2.3 Langkah-langkah Penelitian

Berikut ini adalah langkah-langkah analisis data yang digunakan dalam penelitian ini

1. Menentukan model regresi Poisson
2. Menaksir parameter model regresi Poisson
Estimasi parameter model regresi Poisson dengan menggunakan pendekatan *Iteratively Reweighted Least Square* (IRWLS). Prosedur ini dilakukan dengan metode *Newton Raphson*. Metode ini digunakan untuk menyelesaikan persamaan
3. Menentukan devians (simpangan) model
4. Menentukan model terbaik regresi Poisson
Model regresi Poisson yang layak digunakan dipilih berdasarkan nilai devians yang kecil
5. Melakukan uji overdispersi

Jika terbukti adanya overdispersi maka metode analisis yang digunakan selanjutnya adalah model regresi *Zero Inflated Poisson*. Namun jika tidak terbukti adanya overdispersi maka model terbaik dari model regresi Poisson yang digunakan

6. Menaksir parameter model regresi ZIP
Estimasi parameter model regresi ZIP dilakukan dengan menggunakan algoritma EM
7. Mengkaji kesesuaian model regresi ZIP
Statistik uji yang digunakan adalah G_{hitung}
8. Menguji hipotesis kesesuaian model regresi ZIP
Pengujian yang dimaksud adalah pengujian parameter secara individu
9. Menentukan model terbaik regresi ZIP
Model terbaik yang dipilih adalah model dengan nilai AIC terkecil

3. Hasil dan Pembahasan

3.1 Regresi Poisson

Berikut ini adalah nilai Estimasi Parameter dan Devian pada Tiap Kemungkinan Model Regresi Poisson, yang disajikan dalam Tabel 1.

Tabel 1. Nilai Estimasi Parameter dan Devian pada Tiap Kemungkinan Model Regresi Poisson

No	Model	Nilai Dugaan Parameter				Devians
		β_0	β_1	β_2	β_3	
1.	$\mu = \exp(\beta_0 + x_1)$	0.1955	0.0199			65.2413
2.	$\mu = \exp(\beta_0 + x_2)$	1.8471		-0.0223		59.6090
3.	$\mu = \exp(\beta_0 + x_3)$	-0.6557			0.0194	66.0848
4.	$\mu = \exp(\beta_0 + x_1 + x_2)$	1.4866	0.0105	-0.0205		59.1173
5.	$\mu = \exp(\beta_0 + x_1 + x_3)$	-0.6184	0.0167		0.0128	64.7486
6.	$\mu = \exp(\beta_0 + x_2 + x_3)$	0.4404		-0.0218	0.0197	58.3619
7.	$\mu = \exp(\beta_0 + x_1 + x_2 + x_3)$	0.3821	0.0063	-0.0209	0.0176	58.1946

Berdasarkan model yang dibentuk, ada 4 model yang layak digunakan mengingat nilai devians untuk 4 model tersebut merupakan nilai devians yang kecil untuk tiap-tiap kelompok model berdasarkan jumlah variabel independent. Terdapat tiga kelompok model berdasarkan jumlah variabel independent yang dimasukkan ke dalam model, yaitu kelompok model dengan satu variabel, dua variabel dan tiga variabel. Model-model yang layak digunakan untuk pemodelan Regresi Poisson selanjutnya adalah sebagai berikut :

1. $\mu = \exp(\beta_0 + x_2)$
2. $\mu = \exp(\beta_0 + x_1 + x_2)$
3. $\mu = \exp(\beta_0 + x_2 + x_3)$
4. $\mu = \exp(\beta_0 + x_1 + x_2 + x_3)$

Statistik uji yang digunakan untuk uji kesesuaian model seperti pada persamaan berikut :

$$G = 2 \sum_{i=1}^n \left[y_i \ln \left(\frac{y_i}{\hat{y}_i} \right) - (y_i - \hat{y}_i) \right]$$

Pada tabel 2 berikut, dapat dilihat nilai devians dan $\chi^2_{(n-k-1, \alpha)}$ untuk 4 model regresi Poisson.

No.	Model	Devians	db	$\chi^2_{(db, \alpha)}$	Devians/db
1.	$\mu = \exp(\beta_0 + x_1)$	59.6090	33	47.39988	1.806333
2.	$\mu = \exp(\beta_0 + x_1 + x_2)$	59.1173	32	46.19426	1.847416
3.	$\mu = \exp(\beta_0 + x_1 + x_2)$	58.3619	32	46.19426	1.823809
4.	$\mu = \exp(\beta_0 + x_1 + x_2 + x_3)$	58.1946	31	44.98534	1.877245

Nilai Devians dari ke-4 model Regresi Poisson tersebut memiliki seluruhnya lebih besar dari nilai *Chi-square* tabel, yang artinya seluruh model signifikan pada tingkat signifikansi 0.05, sehingga dapat disimpulkan bahwa ke-4 model layak digunakan. Model-model tersebut mengalami kondisi *overdispersi* karena berdasarkan nilai Devians yang dibagi dengan derajat bebasnya lebih besar dari 1, sehingga penggunaan Regresi Poisson tidak layak digunakan karena tidak memenuhi asumsi *equidispersi*.

3.2 Zero Inflated Poisson (ZIP) Regression

Nilai Estimasi Parameter Regresi ZIP dengan Tiga Variabel Independent dan Analisis Kesesuaian (x_1, x_2, x_3) disajikan pada Tabel 3 sebagai berikut :

Tabel 3. Nilai Estimasi Parameter Regresi ZIP

Parameter	Estimasi	df	t	P-Value	G_{hitung}
γ_0	-7.4864	35	-0.50	0.6176	125.7
γ_1	0.0538	35	0.33	0.7444	
γ_2	-0.0338	35	-0.54	0.5939	
γ_3	0.0772	35	0.31	0.7590	
β_0	-0.03015	35	-0.02	0.9839	
β_1	0.0116	35	0.67	0.5087	
β_2	-0.0227	35	-2.57	0.0147	
β_3	0.0245	35	1.09	0.2833	

Pengujian secara serentak menghasilkan nilai G_{hitung} lebih besar dari nilai $\chi^2_{(1;0.1)} = 2.706$, sehingga model regresi ZIP dengan variabel independen x_1 , x_2 dan x_3 layak digunakan. Parameter yang memiliki nilai P-value yang kurang dari 0.05 ada satu parameter, yaitu β_2 sehingga parameter yang signifikan adalah β_2 .

3.3 Pemilihan Model Regresi ZIP Terbaik

Metode yang dilakukan untuk pemilihan model Regresi ZIP terbaik adalah *Akaike's Informaton Criterion* (AIC). Model terbaik adalah model dengan nilai AIC terkecil. Berikut ini adalah tabel yang menunjukkan nilai AIC untuk ke-enam model Regresi ZIP.

Tabel 4. Pemilihan Model Terbaik

Variabel Model regresi ZIP	AIC
x_2	129.2
x_1, x_2	126.7
x_2, x_3	126.0
x_1, x_2, x_3	125.6*

(keterangan * Model dengan nilai AIC terkecil)

Model dengan nilai AIC terkecil adalah model Regresi ZIP dengan tiga variabel independen, yaitu x_1, x_2 dan x_3 dengan parameter yang signifikan adalah β_2 .

4. Kesimpulan

Variabel yang berpengaruh secara signifikan pada kasus *Infant Mortality Rate* (IMR) di Jawa Tengah antara lain jumlah sarana kesehatan pada tiap kabupaten/kota, rasio ketersediaan bidan desa pada tiap kabupaten/kota, dan presentase persalinan ditolong tenaga kesehatan.

Berikut adalah model Regresi ZIP terbaik yang terpilih :

$$\text{logit}(\omega_i) = -7.4864 + 0.0538 x_{1i} - 0.0338x_{2i} + 0.0772x_{3i}$$

$$\log(\mu_i) = -0.03015 + 0.0116x_{1i} - 0.0227x_{2i} + 0.0245x_{3i}$$

Model logit menjelaskan bahwa peluang variabel respon bernilai nol dipengaruhi oleh rasio ketersediaan bidan desa tiap kabupaten. Sedangkan model log menjelaskan bahwa jumlah kematian bayi akan bertambah sebesar 1.0116 jika jumlah sarana kesehatan berkurang sebesar satu satuan, jumlah kematian bayi akan berkurang sebesar 1.0227 jika rasio ketersediaan bidan desa bertambah sebesar satu satuan, dan jumlah kematian bayi akan bertambah sebesar 1.0245 jika presentase persalinan bertambah sebesar satu satuan, dengan syarat bahwa variabel independen yang lain adalah konstan.

Ucapan Terima Kasih

Penelitian ini didanai oleh Direktorat Riset dan Pengabdian Masyarakat Direktorat Jendral Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi dan Perguruan Tinggi dalam skema Penelitian Dosen Pemula.

Daftar Pustaka

- [1] Badan Pusat Statistik.(2001). *Estimasi Fertilitas, Mortalitas dan Migrasi Hasil Sensus Penduduk Tahun 2000*. Jakarta :Badan Pusat Statistik.
- [2] Baharuddin. (2005). “Ukuran R^2 dalam Model Regresi Poisson”. *Integral* .10,(3),114-121
- [3] Bozdogan, H. (2000). *Akaike's Information Criterion and Recent Developments in Information Complexity, Mathematical Psychology*, 44, 62-91.
- [4] Cameron, A.C, dan Trivedi, P.K. 1998. *Regression Analysis of Count Data*. Cambridge:Cambridge University Press.
- [5] Jansakul, N. & Hinde, J. P. “Score Tests for Zero-Inflated Poisson Models,” *Computational Statistics & Data Analysis*, vol. 40, pp. 75-96, 2002.
- [6] Lambert, D. 1992. *Zero Inflated Poisson Regression, With An Application To Detect In Manufacturing, Journal Technometrics, Feb 1992 Vol 32 no 1*.
- [7] Myers, R.H. 1990. *Classical and Modern Regression with Applications, second. edition* Boston: PWS-KENT Publishing Company.
- [8] Pramasita (2005). Faktor-faktor yang Mempengaruhi Angka. Kematian Bayi di Wilayah Jawa Timur Tahun 2002.
- [9] Rusli, S. 2012. Pengantar Ilmu Kependudukan. Jakarta : LP3ES
- [10] [www.dinkesjatengprov.go.id/buku profil kesehatan provinsi jawa tengah 2014](http://www.dinkesjatengprov.go.id/buku%20profil%20kesehatan%20provinsi%20jawa%20tengah%202014)