



PEMODELAN *NEGATIVE BINOMIAL REGRESSION* PADA DATA JUMLAH KEMATIAN BAYI DI KABUPATEN JOMBANG

Dian Kusuma Wardani¹, Anggun Wulandari²

¹ Prodi Pendidikan Agama Islam, Universitas KH A Wahab Hasbullah

² Prodi Pendidikan Biologi, Universitas KH A Wahab Hasbullah

email korespondensi : dianwardani@unwaha.ac.id

Diterima : (07-10-2020), **Revisi:** (03-11-2020), **Diterbitkan :** (24-12-2020)

ABSTRAK

Regresi *Poisson* merupakan salah satu metode regresi untuk menganalisis apabila variabel terikat berupa data diskrit. Asumsi bahwa nilai rata-rata dan nilai varian dari variabel terikat harus sama terpenuhi akan menghasilkan kesimpulan yang valid. Kasus overdispersi terjadi jika nilai varian lebih besar dari nilai rata-rata sedangkan nilai varian kurang dari nilai rata-rata disebut underdispersi. Pada penelitian ini data yang digunakan adalah data jumlah kematian bayi di Kabupaten Jombang Tahun 2018. Tujuan dari penelitian ini untuk mengetahui faktor-faktor yang berpengaruh terhadap jumlah kematian bayi. Model yang digunakan untuk mengetahui faktor-faktor yang mempengaruhi jumlah kematian bayi adalah model *Negative Binomial Regression*. Berdasarkan hasil analisis diketahui bahwa model *Negative Binomial Regression* yang terbentuk bahwa variabel bebas yang memberikan pengaruh signifikan yaitu variabel persentase ibu hamil mendapatkan tablet Fe3, variabel persentase komplikasi kebidanan yang ditangani, variabel persentase ibu hamil melaksanakan program K4 dan persentase Berat Badan Lahir Rendah (BBLR).

Kata kunci : *poisson, overdispersi, underdispersi, binomial negatif*

ABSTRACT

Poisson regression is a regression method to analyze the dependent variable is discrete data (counting). The assumption that mean value and variance value of the dependent variable must be the same is fulfilled will produce valid conclusions. Overdispersion cases occur when the variance value is greater than mean value while the variance value is less than mean value is called underdispersion. In this study data used is number of infant deaths in Kabupaten Jombang 2018. The purpose of this study is to determine the factors that affect number of infant deaths. The model used to determine the factors that affect number of infant deaths is Negative Binomial Regression model. Based on the results of analysis, it is known that Negative Binomial Regression model which is formed by independent variables which have a significant effect is variable percentage of pregnant women getting Fe3 tablets, variable percentage of obstetric complications handled, variable percentage of pregnant women implementing K4 program and percentage of BBLR

Key words : *poisson, overdispersion, underdispersion, negative binomial*

Pendahuluan

Asumsi yang harus dipenuhi dalam analisis regresi *Poisson* yaitu nilai rata-rata dan nilai varian dari variabel terikat harus sama/*equidispersion* (Myers, dkk., 2010). Kesimpulan akan tidak valid jika asumsi *equidispersi* tidak terpenuhi. Pelanggaran asumsi jika nilai varian lebih besar dari nilai rata-rata disebut *overdispersi* sedangkan nilai ragam kurang dari nilai rata-rata disebut *underdispersi*. Melliana, dkk. (2013) membandingkan metode regresi binomial negatif dan metode regresi *poisson* tergeneralisasi pada kasus *overdispersi* yaitu data jumlah penderita kanker serviks. Hasil penelitian menunjukkan bahwa metode regresi binomial negatif memberikan nilai Akaike Information Criteria (AIC) lebih kecil daripada metode regresi *poisson* tergeneralisasi. Menurut Astuti, dkk. (2013), model binomial negatif lebih sesuai jika digunakan pada data yang mengalami *overdispersi* dan *underdispersi*.

Ismail dan Jemain (2007) melakukan penelitian tentang perbandingan metode regresi binomial negatif dan metode regresi *poisson* tergeneralisasi. Hasil penelitian menunjukkan bahwa kedua metode digunakan untuk menangani masalah pelanggaran asumsi *equidispersion* khususnya *overdispersi* yaitu kondisi dimana nilai varian lebih besar daripada nilai rata-rata. Model regresi binomial negatif merupakan salah satu model yang dapat digunakan saat terjadi *overdispersi* pada data count. Pada penelitian ini membahas mengenai metode maksimum likelihood

dimana solusi dari fungsi *likelihoodnya* diselesaikan dengan metode Newton Raphson. Uji kesesuaian model yang digunakan mencakup statistik *pseudo-R²*, uji rasio *likelihood* dan uji Wald (Wahyuni, 2011). Pada penelitian ini menggunakan pemodelan regresi binomial negatif, data diduga terjadi underdispersi atau overdispersi.

Angka kematian bayi (AKB) atau *Infant Mortality Rate* adalah jumlah penduduk yang meninggal sebelum mencapai usia 1 tahun per 1000 kelahiran hidup (KH) pada tahun yang sama. AKB dapat menggambarkan kondisi sosial ekonomi masyarakat setempat, karena bayi adalah kelompok usia yang paling rentan terkena dampak dari perubahan lingkungan maupun sosial ekonomi. AKB Kabupaten Jombang menunjukkan tren yang menurun dalam beberapa tahun terakhir dari tahun 2010-2017, namun AKB meningkat lagi di tahun 2018. Dari 8,50 per 1000 KH pada tahun 2017 menjadi 10,28 per 1000 KH pada tahun 2018. Penyebab peningkatan AKB adalah banyaknya bayi lahir dengan BBLR dan Asfiksia dan kelainan bawaan. Beberapa upaya yang telah dilakukan untuk menurunkan AKB antara lain : (1) adanya kegiatan orientasi terpadu bagi ibu hamil dalam penurunan AKI dan AKB, (2) deteksi dini resiko tinggi bayi bagi kader, (3) pengkajian kasus ibu dan bayi, (4) review program KIA bagi bidan desa, (5) penguatan kelas ibu hamil dan ibu balita (Dinas Kesehatan Kabupaten Jombang, 2018).

Regresi Binomial Negatif

Regresi Binomial Negatif adalah salah satu model regresi yang merupakan terapan dari GLM. Menurut Greene (2008), sebagai penerapan dari GLM maka sebaran *Negative Binomial* memiliki ketiga komponen random, komponen sistematis dan fungsi penghubung. Pada model regresi *Negative Binomial* variabel terikat diasumsikan memiliki sebaran *Negative Binomial* yang dihasilkan dari sebaran campuran *Poisson Gamma*. Fungsi peluang model regresi *Negative Binomial* dapat dinyatakan sebagaimana persamaan (1):

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \frac{1}{\kappa})}{\Gamma(\frac{1}{\kappa})y_i!} \left(\frac{1}{1 + \kappa\mu_i}\right)^{\frac{1}{\kappa}} \left(\frac{\kappa\mu_i}{1 + \kappa\mu_i}\right)^{y_i} \text{ dengan } i = 0, 1, 2, \dots, n \quad (1)$$

Pada saat $\kappa \rightarrow 0$ maka sebaran *Negative Binomial* memiliki ragam $V[Y] \rightarrow \mu$. Sebaran *Negative Binomial* akan mendekati suatu sebaran Poisson yang mengasumsikan rata-rata dan ragam yang sama yaitu $E[Y] = V[Y] = \mu$. Dalam model regresi *Negative Binomial*, y_i adalah variabel yang berupa *count data*. Menurut Hilbe (2011), model regresi *Negative Binomial* pada umumnya menggunakan fungsi

penghubung logaritma atau *log link* yaitu $\ln \mu_i = X_i^T \beta$. Model regresi *Negative Binomial* dapat menggunakan *log link* karena $\ln \mu_i$ dan $X_i^T \beta$ akan terdefinisi di dalam interval $(0, \infty)$ dan interpretasi parameter regresi akan menjadi lebih mudah. Setelah diperoleh fungsi penghubung yang tepat, selanjutnya dapat dinyatakan model regresi *Negative Binomial* untuk memodelkan *count data* yaitu : $\ln[E(Y_i, X_i)] = \ln(\mu_i) = X_i^T \beta$ untuk $i = 1, 2, \dots, n$. Sehingga dapat diperoleh $\mu_i = \exp(X_i^T \beta)$.

Multikolinieritas

Menurut Li (2000), pendeteksian multikolinieritas dapat dilakukan menggunakan nilai *Variance Inflation Factor* (VIF). Untuk regresi dengan lebih dari dua variabel definisi VIF adalah

$$VIF_j = \frac{1}{1-R_j^2} \quad (2)$$

di mana :

R_j^2 : koefisien determinasi dari *auxiliary regression*

Auxiliary regression adalah regresi dengan X_j sebagai variabel respon, dan X selainnya sebagai variabel prediktor. Nilai R_j^2 berkisar antara 0 sampai dengan 1 sehingga nilai VIF akan naik seiring dengan kenaikan koefisien determinasi dari *auxiliary regression*. Nilai VIF yang lebih dari 10 merupakan bukti cukup untuk mendeteksi adanya multikolinieritas (Li, 2000).

Metode Penelitian

Data dalam penelitian ini adalah data sekunder didapatkan dari Profil Kesehatan Kabupaten Jombang yang dikeluarkan Dinas Kesehatan Kabupaten Jombang. Unit pengamatan sebanyak 21 unit pengamatan yang terdiri dari 21 kecamatan yang ada di Kabupaten Jombang. Variabel penelitian yang digunakan dalam penelitian ini terdiri dari dua yaitu variabel terikat (Y) dan lima variabel bebas (X). Tabel 1 menyajikan uraian dari setiap variabel :

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y ₁	Jumlah kematian bayi
X ₁	Persentase persalinan oleh tenaga kesehatan
X ₂	Persentase ibu hamil mendapatkan tablet Fe3

X_3	Persentase komplikasi kebidanan yang ditangani
X_4	Persentase ibu hamil melaksanakan program K4
X_5	Persentase BBLR

Langkah-langkah untuk mengidentifikasi faktor-faktor yang berpengaruh terhadap jumlah kematian bayi dan kematian ibu dengan pendekatan model *Negative Binomial Regression* adalah :

1. Melakukan uji distribusi poisson pada variabel terikat
2. Melakukan uji asumsi equidispersi
3. Mendeteksi multikolinieritas dari variabel bebas dengan menggunakan kriteria uji VIF
4. Mendapatkan penduga parameter model *Negative Binomial Regression* dengan menggunakan *Maximum Likelihood Estimation* (MLE)
5. Melakukan pengujian hipotesis untuk *Negative Binomial Regression*
6. Melakukan interpretasi model
7. Membuat kesimpulan dari hasil analisis

Hasil dan Pembahasan

Uji Asumsi Distribusi Poisson

Sebelum melakukan pemodelan *Negative Binomial Regression* maka data variabel terikat harus diuji apakah mengikuti distribusi poisson. Pengujian distribusi ini menggunakan *Kolmogorov Smirnov Test* yang ada pada SPSS. Berikut hipotesis untuk pengujian ini :

H_0 : Sampel berasal dari populasi yang berdistribusi poisson

H_1 : Sampel berasal dari populasi yang tidak berdistribusi poisson

Taraf signifikansi yang digunakan adalah 0,05. Kriteria uji H_0 ditolak jika nilai hasil output memiliki nilai kurang dari 0,05.

Tabel 2. Kolmogorov Smirnov Test

		Y
N		21
Poisson Parameter a,b	Rata-rata	9.48
Most Extreme Difference	Absolute	0.262
	Positive	0.262
	Negative	-0.179
Kolmogorov-Smirnov Z		1.200
Asymp. Sig. (2-tailed)		0.112

Berdasarkan Tabel 2 diperoleh nilai sig 0.112 > 0.05 maka keputusan Terima Ho berarti sampel berasal dari populasi yang berdistribusi poisson.

Parameter Dispersi

Parameter dispersi (ϕ) diperoleh dari rumus $\phi = \frac{\text{nilai deviance}}{df}$ dengan **df = degree of freedom**. Apabila nilai $\phi > 1$ maka terjadi overdispersi dan apabila $\phi < 1$ maka terjadi underdispersi. Tabel 3 berikut akan menyajikan nilai deviance.

Tabel 3. Nilai Deviance

	Value	df	Value/df
Deviance	2.903	15	0.194
Pearson Chi-Square	2.903	15	0.237

Berdasarkan Tabel 3 diperoleh nilai $\phi = 0,194$ nilai ini menandakan bahwa terjadi kasus underdispersi karena $\phi < 1$.

Asumsi Non-Multikolinieritas

Tabel 4. Nilai Tolerance dan VIF

Model	Collinearity Statistics	
	Tolerance	VIF
X ₁	0.222	4.495
X ₂	0.640	1.562
X ₃	0.804	1.244
X ₄	0.226	4.421
X ₅	0.686	1.458

Berdasarkan Tabel 4 diperoleh nilai dari masing-masing variabel memiliki nilai VIF kurang dari 10 sehingga dapat dikatakan bahwa tidak terdapat masalah multikolinieritas antar variabel bebas.

Model Negative Binomial Regression

Model *Negative Binomial Regression* merupakan regresi yang diaplikasikan pada data yang mengalami kasus overdispersi atau underdispersi. Pemodelan *Negative Binomial Regression* dapat dilakukan setelah pengujian asumsi non-multikolinieritas. Data yang digunakan yaitu data kematian bayi di Kabupaten Jombang Tahun 2018 menggunakan *Negative Binomial Regression* dengan lima variabel bebas yaitu persentase persalinan oleh tenaga kesehatan (X_1), persentase ibu hamil mendapatkan tablet Fe3 (X_2), persentase komplikasi kebidanan yang ditangani (X_3), persentase ibu hamil melaksanakan program K4 (X_4) dan persentase BBLR (X_5). Pendugaan parameter menggunakan *Maximum Likelihood Estimation* (MLE) di mana membutuhkan iterasi *Newton Raphson*. Berikut disajikan parameter dari *Negative Binomial Regression*.

Tabel 5. Hasil Pendugaan Parameter *Negative Binomial Regression*

Parameter	B	Std. Error	Hypotesis Test		
			Wald Chi Square	df	Sig
(Intercept)	-5.632	2.7760	4.116	1	0.042
X ₁	0.130	0.0689	4.042	1	0.060
X ₂	-0.801	0.0277	8.327	1	0.004
X ₃	-0.022	0.0146	6.022	1	0.028
X ₄	-0.259	0.0885	9.901	1	0.002
X ₅	-0.017	0.1138	10.014	1	0.009
(Scale)	1 ^a				

Pengujian parameter secara serentak dengan menggunakan hipotesis sebagai berikut:

Parameter β

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_{j4} = 0$$

H_1 : paling sedikit ada satu $\beta_l \neq 0$, dengan $l = 1,2,3,4$

Hasil pengujian hipotesis secara serentak dengan menggunakan statistik uji G diperoleh sebesar 13,532. Nilai statistik uji G lebih besar dibandingkan dengan

$\chi_4^2 = 11.070$. Hal ini menunjukkan bahwa secara serentak variabel bebas memberikan pengaruh signifikan pada model yang terbentuk. Pengujian parameter secara parsial. adapun hipotesis dalam pengujian parameter secara parsial sebagai berikut :

Parameter β_l

$$H_0 : \beta_l = 0$$

$$H_1 : \beta_l \neq 0, l = 1,2,3,4$$

Berdasarkan hasil pada Tabel 5 parameter yang signifikan untuk jumlah kematian bayi adalah $\beta_2, \beta_3, \beta_4, \beta_5$. Hal ini terlihat dari nilai *p-value* dari parameter kurang dari alpha (0,05). Berdasarkan hasil pengujian parameter pada Tabel 5 maka variabel bebas yang signifikan untuk jumlah kematian bayi adalah variabel persentase ibu hamil mendapatkan tablet Fe3, persentase komplikasi kebidanan yang ditangani, variabel persentase ibu hamil melaksanakan program K4 dan persentase BBLR. Model yang terbentuk dari *Negative Binomial Regression* sebagai berikut :

$$\ln \hat{\mu} = -5.632 + 0.130X_1 - 0.801X_2 - 0.022X_3 - 0.259X_4 - 0.017X_5$$

Interpretasi model *Negative Binomial Regression* sebagai berikut :

1. Setiap kenaikan 1 persen persentase ibu hamil mendapatkan tablet Fe3 maka akan menurunkan ln rata-rata jumlah kasus kematian bayi sebesar 0.801 dengan asumsi bahwa variabel yang lain tidak berubah.
2. Setiap kenaikan 1 persen persentase komplikasi kebidanan yang ditangani maka akan menurunkan ln rata-rata jumlah kasus kematian bayi sebesar 0.022 dengan asumsi bahwa variabel yang lain tidak berubah.
3. Setiap kenaikan 1 persen persentase ibu hamil melaksanakan program K4 maka akan menurunkan ln rata-rata jumlah kasus kematian bayi sebesar 0.259 dengan asumsi bahwa variabel yang lain tidak berubah.
4. Setiap kenaikan 1 persen persentase BBLR maka akan menurunkan ln rata-rata jumlah kasus kematian bayi sebesar 0.017 dengan asumsi bahwa variabel yang lain tidak berubah.

Kesimpulan dan Saran

Berdasarkan model *Negative Binomial Regression* yang terbentuk variabel bebas yang memberikan pengaruh signifikan terhadap jumlah kasus kematian bayi di

Kabupaten Jombang Tahun 2018 adalah variabel persentase ibu hamil mendapatkan tablet Fe₃, variabel persentase komplikasi kebidanan yang ditangani, variabel persentase ibu hamil melaksanakan program K4 dan persentase BBLR. Berdasarkan hasil penelitian dapat diberikan saran kepada pihak terkait untuk meningkatkan penyuluhan mengenai kesehatan pada ibu hamil. Sedangkan sebagai bahan evaluasi pada penelitian selanjutnya untuk lebih banyak menggunakan variabel bebas dalam model agar didapatkan model dengan faktor signifikan lebih tepat.

Ucapan Terimakasih

Penulis mengucapkan terima kasih kepada Kemenristek/BRIN yang telah memberikan dukungan dana terhadap penelitian ini dan ibu yang selalu memberikan dukungan dan semangat selama proses penelitian ini berlangsung. Kepada seluruh jajaran pimpinan dan rekan di Universitas KH.A Wahab Hasbullah (UNWAHA) Tambakberas Jombang yang telah memberikan dukungan selama penulisan artikel ini.

Daftar Pustaka

- Astuti C C, Sumarminingsih E dan Soehono, L A. (2013). *Perbandingan Generalized Poisson Regression dan Negative Binomial Regression untuk Data Overdispersi dan Underdispersi pada Regresi Poisson. Skripsi : Program Studi Statistika FMIPA*. Universitas Brawijaya. Malang.
- Dinas Kesehatan Kabupaten Jombang. (2018). *Profil Kesehatan Indonesia*. Dinas Kesehatan Jombang.
- Greene, W. (2008). *Functional Forms For The Negative Binomial Model For Count Data Working Paper Department of Economics-Stren School of Business*, 585-590.
- Hilbe, J. M. (2011). *Negative Binomial Regression*. New York: Cambridge University Press.
- Ismail, N dan Jemain, A A. (2007). *Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models*. Casualty Actuarial Society Forum. <http://semanticscholar.org/paper/>. Tanggal akses : 9 Juli 2020.

- Li, F. (2000). *Multicollinearity*. Department of Statistics, Stockholm University, Hal. 1-10. <http://people.su.se/>. Tanggal Akses: 29 Juni 2019.
- Melliana A, Yeni Setyorini, Haris Eko, Sisty Rosi dan Purhadi. (2013). *The Comparison Of Generalized Poisson Regression And Negative Binomial Regression Methods In Overcoming Overdispersion*. International Journal Of Scientific & Technology Research Volume 2. <http://www.ijstr.org/final-print/aug2013/>. Tanggal akses : 10 Juli 2020.
- Myers, R.H., Montgomery, D.C., Vining, G.G., dan Robinson, T.J. (2010). *Generalized Linier Models with Application in Engineering and Sciences*. John Wiley and Sons, Inc., Publication. Canada.
- Wahyuni W. (2011). *Penaksir Parameter Model Regresi Binomial Negatif pada Kasus Overdispersi*. Skripsi: Program Studi Matematika FMIPA. Universitas Indonesia. Depok.