

UNDERDISPERSION REGRESI POISSON APLIKASI PADA KASUS KEMATIAN BAYI DI PROVINSI JAWA TENGAH

Indah Manfaati Nur, S.Si, M.Si¹⁾, Tiani Wahyu Utami, S.Si, M.Si²⁾

¹Fakultas MIPA, Universitas Muhammadiyah Semarang
email: indahmanfaatinur_statistics@unimus.ac.id

² Fakultas MIPA, Universitas Muhammadiyah Semarang
email: tiani.utami88@gmail.com

Abstract

In a Poisson regression analysis, the response variable (Y) must meet the assumptions equidispersion (variance value is equal to the mean). However, in real data often occurs overdispersion / underdispersion (variance value is not equal to the mean). Overdispersion is the emergence of greater diversity in the set of data compared with the expected variance based models. The implication, for the correct model, the value of Pearson Chi-square statistic divided by the degree of freedom will to be equal to 1. Overdispersion occurs if the value exceeds 1, and underdispersion occurs if the value is less than 1. Overdispersion / underdispersion cause the resulting model be less precise. One alternative to overcome that by replacing assuming a Poisson distribution with a more flexible distribution. The purpose of this article is to show the condition underdispersion Poisson regression modeling on Infant Mortality Case By Regency / City in Central Java province. The predictor variables used are the number of health facilities health centers in each district / city, village midwives availability ratio in each district/city, the percentage of births assisted health personnel, the baby's health care coverage, and number handling obstetric complications.

Keywords: *Poisson regression, Underdispersion, Infant Mortality*

1. PENDAHULUAN

Dalam analisis regresi Poisson, variabel respon (Y) harus memenuhi asumsi *equidispersion* (nilai varians sama dengan nilai mean). Namun, dalam data riil seringkali terjadi *overdispersion/ underdispersion* (nilai varians tidak sama dengan nilai mean).

Overdispersion/ underdispersion menyebabkan model yang dihasilkan menjadi kurang tepat. Salah satu alternatif untuk mengatasinya yaitu dengan mengganti asumsi distribusi Poisson dengan distribusi yang lebih fleksibel. Tujuan dari artikel ini adalah menunjukkan kondisi *underdispersion* dalam pemodelan regresi Poisson pada Kasus Kematian Bayi Menurut Kabupaten/Kota di Provinsi Jawa Tengah. Variabel prediktor yang digunakan adalah jumlah sarana kesehatan puskesmas pada tiap kabupaten/kota, ratio ketersediaan bidan desa pada tiap kabupaten/kota, persentase persalinan ditolong tenaga kesehatan, cakupan

pelayanan kesehatan bayi, Jumlah penanganan komplikasi kebidanan.

Angka kematian dari waktu ke waktu menggambarkan status kesehatan masyarakat secara kasar, kondisi atau tingkat permasalahan kesehatan, kondisi lingkungan fisik dan biologik secara tidak langsung. Angka tersebut dapat digunakan sebagai indikator dalam penilaian keberhasilan pelayanan kesehatan dan program pembangunan kesehatan.

Data jumlah kematian bayi di Provinsi Jawa Tengah merupakan data berdistribusi Poisson dimana kejadian tersebut jarang terjadi.

Dalam jurnal ini akan dibahas suatu model analisis statistik yang sesuai dengan data yang bersifat diskrit dan berdistribusi poisson. Selanjutnya akan ditunjukkan bilamana kondisi overdispersi atau underdispersi dapat terjadi pada suatu data yang diaplikasikan pada suatu contoh kasus

yaitu jumlah kematian bayi di Provinsi Jawa Tengah.

Penelitian sebelumnya terkait regresi poisson telah dilakukan oleh Indraswari (2011) tentang Faktor- faktor Penyebab Kematian Bayi di Kabupaten Karangasem dengan Pendekatan Regresi Poisson. Penulis tertarik ingin menunjukkan kondisi non equidispersi (underdispersi) pada pemodelan regresi Poisson menggunakan data kematian bayi.

2. KAJIAN LITERATUR

Angka Kematian Bayi (AKB) merupakan jumlah kematian bayi (0-11 bulan) per 1000 kelahiran hidup dalam kurun waktu satu tahun. AKB menggambarkan tingkat permasalahan kesehatan masyarakat yang berkaitan dengan faktor penyebab kematian bayi, tingkat pelayanan antenatal, status gizi ibu hamil, tingkat keberhasilan program KIA dan KB, serta kondisi lingkungan dan sosial ekonomi. Apabila AKB di suatu wilayah tinggi, berarti status kesehatan di wilayah tersebut rendah.

Salah satu metode statistika yang dapat digunakan untuk mengetahui pola hubungan antara jumlah kematian bayi dengan faktor-faktor yang mempengaruhinya adalah analisis regresi.

Analisis regresi merupakan suatu metode yang digunakan untuk menganalisis pola hubungan antara variabel respon dengan beberapa variabel prediktor. Analisis regresi digunakan untuk menganalisis data variabel respon yang berupa data diskrit dan data kontinu. Metode analisis regresi yang umumnya sering digunakan untuk menganalisis data dengan variabel respon Y yang berupa data diskrit dan variabel prediktor X berupa data diskrit, kontinu, kategorik atau campuran berdistribusi Poisson adalah model regresi Poisson [8].

Jika Y merupakan data diskrit yang berdistribusi Poisson dengan parameter $\mu > 0$ maka fungsi massa peluangnya adalah

$$f_{y_i, \mu} = \frac{e^{-\mu} \mu^y}{y!}; y = 0, 1, 2, 3, \dots$$

Persamaan 1 dapat dirubah sebagai berikut:

$$f_{y_i, \mu} = \exp\{y \ln \mu - \mu - \ln(y_i!)\}$$

$$f_{y_i, \mu} = \exp\{y_i \theta_i - b(\theta_i) - \ln(y_i!)\} \quad (2)$$

persamaan fungsi distribusi keluarga eksponensial. Dengan menggunakan fungsi link diperoleh model regresi Poisson dalam bentuk:

$$\ln \mu = \mathbf{X}\boldsymbol{\beta}$$

$$E y_i = \mu_i$$

$$\mu_i = \exp\{\beta_0 + \sum_{j=1}^m \beta_j x_j\}; 1, 2, \dots, n \quad (3)$$

dengan i berupa unit eksperimen, unit eksperimen sebanyak n dan m menyatakan banyaknya variabel prediktor. Dalam bentuk matrik model regresi Poisson ditulis adalah sebagai berikut :

$$y_i = \exp \mathbf{X}\boldsymbol{\beta} \text{ dengan } i = 1, 2, 3, \dots, n \quad (4)$$

dimana y berupa variabel respon berdistribusi Poisson yang berbentuk vektor dengan ukuran $(n \times 1)$, \mathbf{X} menyatakan variabel prediktor berbentuk matrik dengan ukuran $(n \times (p+1))$ dan $\boldsymbol{\beta}$ adalah parameter dengan ukuran $((p+1) \times 1)$ [8].

Dalam regresi Poisson, ada asumsi yang harus terpenuhi yaitu asumsi *equidispersion* [10]. Akan tetapi dalam kasus data riil, asumsi *equidispersion* jarang terpenuhi karena data yang bertipe diskrit seringkali mengalami kasus *overdispersion* (penyimpangan tinggi) maupun *underdispersion* yaitu nilai *mean* dan *variance* tidak sama, dengan kata lain nilai *variance* lebih besar daripada nilai *mean* ($\text{var}(Y) > \mu$) atau nilai *variance* lebih kecil daripada nilai *mean* ($\text{var}(Y) < \mu$) [9].

Overdispersion/Underdispersion adalah munculnya keragaman yang lebih besar pada sekumpulan data dibandingkan dengan ragam yang diharapkan berdasarkan model. Overdispersion sering terjadi ketika melakukan model fitting berdasarkan sebaran Binomial atau Poisson. Implikasinya, untuk model yang benar, nilai statistic Chi-square

Pearson dibagi dengan derajat kebebasnya akan bernilai sama dengan 1. Overdispersion terjadi jika nilai tersebut melebihi dari 1, dan underdispersion terjadi jika nilai tersebut kurang dari 1.

Overdispersion menyebabkan nilai devians model menjadi sangat besar dan menyebabkan model yang dihasilkan menjadi kurang tepat. Salah satu cara untuk mengatasi adanya kasus *overdispersion* dalam regresi Poisson adalah dengan mengganti asumsi distribusi Poisson dengan distribusi lain yang lebih fleksibel.

3. METODE PENELITIAN

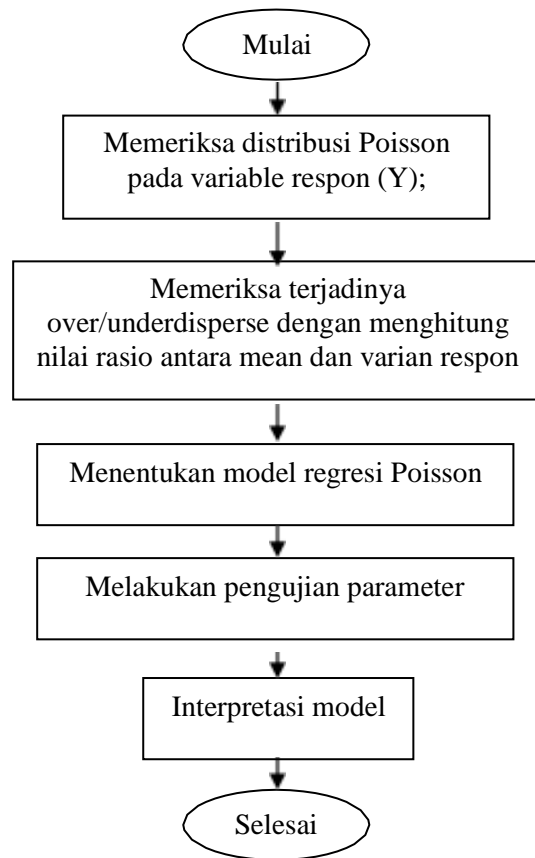
Sumber Data dan Variabel Penelitian

Penulis melakukan analisis data terhadap data sekunder jumlah kasus kematian bayi menurut kabupaten/kota di Provinsi Jawa Tengah diperoleh dari Pusat Data dan Informasi Kesehatan Provinsi Jawa Tengah dan Profil Kesehatan Povinsi Jawa Tengah tahun 2014. Data yang digunakan sebanyak 35 kabupaten atau kota di Provinsi Jawa Tengah.

Variabel-variabel yang digunakan dinyatakan sebagai berikut :

- y = Jumlah kasus kematian bayi
- x₁ = Jumlah sarana kesehatan puskesmas pada tiap kabupaten/kota
- x₂ = Ratio ketersediaan bidan desa pada tiap kabupaten/kota
- x₃ = Persentase persalinan ditolong tenaga kesehatan
- x₄ = Cakupan pelayanan kesehatan bayi
- x₅ = Jumlah penanganan komplikasi kebidanan

Adapun langkah- langkah teknik analisis datanya adalah:



Analisis data pada penelitian ini menggunakan bantuan software R.

4. HASIL DAN PEMBAHASAN

Uji Asumsi

Uji kolmogorov smirnov

Tabel 1. One-Sample Kolmogorov-Smirnov Test

	Y
N	35
Poisson Parameter ^a Mean	7.8857
Most Extreme Differences	
Absolute	.134
Positive	.134
Negative	-.088
Kolmogorov-Smirnov Z	.795
Asymp. Sig. (2-tailed)	.552

a. Test distribution is Poisson.

Nilai T untuk uji Kolmogorov-Smirnov adalah 0.134, dengan nilai $w_{0,97}$ adalah 0.269. Nilai T kurang dari $w_{0,97}$ sehingga terima H_0 yang artinya variabel dependent atau variabel respon berdistribusi Poisson.

Uji Kolinieritas

Tabel 2. Nilai Variance Inflation Factor

Variabel	x_1	x_2	x_3	x_4	x_5
ke-j					
VIF _j	2.450	1.189	1.349	1.236	4.385

Diketahui bahwa tidak ada kondisi kolinieritas antar variabel independent karena nilai VIF kurang dari 10 (VIF < 10)

Nilai Eigen

Tabel 3. Nilai Eigen/*Eigen Value*

Variabel ke-j	<i>Eigenvalue_j</i>
x_1	.439
x_2	.063
x_3	.057
x_4	.042
x_5	.028

Pada tabel nilai eigen (*eigen value*) tersebut dapat diketahui bahwa terdapat nilai yang kurang dari 0.05, maka ada efek kolinieritas pada variabel independen. Variabel yang mengalami kolinieritas tidak digunakan pada analisis berikutnya. Variabel yang dapat digunakan adalah x_1, x_2 dan x_3 .

Tabel 4 Nilai Estimasi Parameter dan Devians pada Tiap Kemungkinan Model Regresi Poisson

Model	Nilai Dugaan Parameter			
	β_0	β_1	β_2	β_3
$\mu = \exp \beta_0 + x_1$	2.2649	-0.008		
$\mu = \exp \beta_0 + x_2$			0.0021	
$\mu = \exp \beta_0 + x_3$				

$\mu = \exp \beta_0$	1.9519	-	-	-	25.2112
$+ x_3$				0.0039	
$\mu = \exp \beta_0 + x_1$	2.1883	-	0.0011	-	23.8513
$+ x_2$		0.0075			*
$\mu = \exp \beta_0 + x_1$	2.3344	-	-	-	23.9542
$+ x_3$		0.0079		0.0011	
$\mu = \exp \beta_0 + x_2$	2.3344		0.0022	-	24.7595
$+ x_3$				0.0043	*
$\mu = \exp \beta_0 + x_1$	2.2815	-0.007	0.0012	-0.001	23.8227
$+ x_2$					*
$+ x_3$					

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 :

- $\mu = \exp \beta_0 + x_1$
- $\mu = \exp \beta_0 + x_1 + x_2$
- $\mu = \exp \beta_0 + x_2 + x_3$
- $\mu = \exp \beta_0 + x_1 + x_2 + x_3$

Statistik uji kesesuaian model seperti pada persamaan berikut :

$$G = 2 \sum_{i=1}^n y_i \ln \frac{y_i}{\mu} - y_i - \mu$$

Pada tabel 5 dapat dilihat nilai devians dan $\chi^2_{(n-k-1, \alpha)}$ untuk 4 model regresi Poisson.

Tabel 5. Model Regresi Poisson

Model	Devians	df	$\chi^2_{(df)}$	Devians / db
$\mu = \exp \beta_0 + x_1$	23.959	3	47.399	0.72604
	6	3	8	8
$\mu = \exp \beta_0 + x_1 + x_2$	23.851	3	46.196	0.74535
	3	2	2	3
$\mu = \exp \beta_0 + x_2 + x_3$	24.759	3	46.196	0.77373
	5	2	2	4
$\mu = \exp \beta_0 + x_1 + x_2 + x_3$	23.822	3	44.985	0.76845
	2	1	3	8

Berdasarkan ke-4 model Regresi Poisson tersebut memiliki nilai Devians yang seluruhnya lebih kecil dari nilai *Chi-square* tabel, yang artinya seluruh model tidak signifikan pada tingkat signifikansi 0.05, sehingga dapat disimpulkan bahwa ke-4 model tidak layak digunakan. Model-model tersebut mengalami kondisi *underdispersi* karena berdasarkan nilai Devians yang dibagi dengan derajat bebasnya kurang dari 1, sehingga penggunaan Regresi Poisson tidak layak digunakan karena tidak memenuhi asumsi *equidispersi*. Alternatif pemodelan yang dapat digunakan adalah Model Regresi *Negative Binomial* atau *Generalized Poisson Regression*.

5. SIMPULAN

1. Pola hubungan dalam analisis regresi Poisson Kasus Kematian Bayi Menurut Kabupaten/Kota di Provinsi Jawa Tengah didapatkan melalui *software R* dengan 3 variabel prediktor yang signifikan adalah jumlah sarana kesehatan puskesmas pada tiap kabupaten/kota, ratio ketersediaan bidan desa pada tiap kabupaten/kota, dan persentase persalinan ditolong tenaga kesehatan
2. Terdapat tiga kelompok model berdasarkan jumlah variabel independent yang dimasukkan ke dalam model, yaitu kelompok model dengan satu variabel, dua variabel dan tiga variable

3. Model Regresi Poisson yang di dapat mengalami kondisi *underdispersi* karena berdasarkan nilai Devians yang dibagi dengan derajat bebasnya kurang dari 1.

2. REFERENSI

[1] Agresti, A.(2002). *Categorical Data Analysis Second Edition*, John Wiley & Sons, New York.

[2] Arriaga, A.E.(1979). Infant and Child Mortality in Selected Asian Countries. *Proceeding of The Meeting on Socio economic Determinants of Consequences of Mortality*, Mexico City.

[3] Astuti, E.T & Yanagawa, T.(2002).Testing Trend for Count Data with Extra-Poisson Variability. *Biometrics*, 58, 398-402

[4] Badan Penelitian & Pengembangan Kesehatan.(1995). *Survei Kesehatan Rumah Tangga 1995*. Jakarta: Departemen Kesehatan RI.

[5] Badan Pusat Statistik.(2001). *Estimasi Fertilitas, Mortalitas dan Migrasi Hasil Sensus Penduduk Tahun 2000*. Jakarta :Badan Pusat Statistik.

[6] Cameron, A.C. and P.K. Trivedi. (1998). *Regression Analysis of Count Data*. Cambridge University Press

[7] Casella, G and Berger, R. L., *Statistical Inference*, Wadsworth, Inc., California, 1990.

[8] Indraswari, Made Susy. (2011). “Faktor-faktor Penyebab Kematian Bayi di Kabupaten Karangasem dengan Pendekatan Regresi Poisson”. *Skripsi*. Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Udayana. Bukit Jimbaran.

[9] McCullagh, P. & Nelder, J.A. (1989). *Generalized Linier Models, second edition*. London: Chapman & Hall.

[10] Myers, R.H., Montgomery, D.C. Vining, G.G., & Robinson, T.J. (2010). *Generalized Linier Models with Aplications in Engineering and the Sciences*. Canada : A John Wiley & Sons, Inc., Publication.