

PEMODELAN KASUS TUBERKULOSIS (TB) DI NUSA TENGGARA BARAT MENGUNAKAN MODEL REGRESI BINOMIAL NEGATIF

¹Wirajaya Kusuma, ²Carina Firstca Utomo, ³Sindy Tervia, ⁴Rifani Nur Sindy Setiawan

¹Universitas Bumigora

wirajaya@universitasbumigora.ac.id

²Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi

carinafirstca@gmail.com

³Badan Strategi Kebijakan Dalam Negeri Kementerian Dalam Negeri

sindytervi@gmail.com

⁴Universitas Mataram

setiawanrifani@unram.ac.id

ABSTRAK

Tuberkulosis merupakan penyakit menular langsung yang disebabkan oleh bakteri (*Mycobacterium Tuberculosis*). Jumlah kasus TB di NTB tahun 2020 menurun sebesar 16,58% dari tahun 2019. Hal ini perlu dilakukan analisis untuk mengetahui faktor apa saja yang berpengaruh terhadap tuberkulosis sehingga jumlah kasus tuberkulosis bisa diminimalisir. Data jumlah kasus TB merupakan data *count* sehingga analisis yang digunakan untuk memodelkan adalah regresi Poisson. Dalam analisis regresi Poisson sering kali terjadi fenomena overdispersi. Jika terjadi overdispersi, regresi Poisson tidak sesuai untuk memodelkan data karena akan menghasilkan estimasi parameter yang bias. Salah satu metode yang digunakan dalam mengatasi overdispersi dalam regresi Poisson adalah regresi Binomial Negatif. Hasil analisis menunjukkan terdapat empat variabel yang signifikan terhadap jumlah kasus TB yaitu kepadatan penduduk (X_1), jumlah puskesmas (X_2), jumlah tenaga keperawatan (X_3), dan persentase rumah tangga yang memiliki akses sanitasi layak (X_6) dengan model $\hat{\mu}_i = \exp(6.5730 + 0.0002X_1 + 0.0704X_2 - 0.0004X_3 + 0.0183X_6)$. Hasil pemilihan model terbaik menunjukkan bahwa model Regresi Binomial Negatif lebih baik daripada model regresi Poisson berdasarkan kriteria kebaikan model *Deviance* dan AIC.

Kata kunci: Regresi Poisson, *Overdispersion*, Regresi Binomial Negatif

ABSTRACT

Tuberculosis is a direct infectious disease caused by bacteria (*Mycobacterium tuberculosis*). The number of TB cases in NTB in 2020 decreased by 16.58% from 2019. This needs to be analyzed to find out what factors influence tuberculosis so that the number of tuberculosis cases can be minimized. Data on the number of TB cases is count data, so the analysis used to model is Poisson regression. In Poisson regression analysis, the phenomenon of overdispersion often occurs. If there is overdispersion, Poisson regression is not suitable for modeling the data because it will produce biased parameter estimates. One of the methods used to overcome overdispersion in Poisson regression is Negative Binomial regression. The results of the analysis show that there are four variables that are significant to the number of TB cases, namely population density (X_1), number of health centers (X_2), number of nursing staff (X_3), and the percentage of households that have access to proper sanitation (X_6) with the model $\hat{\mu}_i = \exp(6.5730 + 0.0002X_1 + 0.0704X_2 - 0.0004X_3 + 0.0183X_6)$. The results of selecting the best model show that the Negative Binomial Regression model is better than the Poisson regression model based on the criteria for the goodness of the *Deviance* and AIC models.

Keywords: Poisson Regression, *Overdispersion*, Negative Binomial Regression.

I. PENDAHULUAN

Tuberkulosis (TB) adalah penyakit menular yang merupakan penyebab utama kesehatan yang buruk, salah satu dari 10 penyebab kematian teratas di dunia dan penyebab utama kematian dari agen infeksi tunggal (peringkat di atas HIV/AIDS). TB

disebabkan oleh basil *mycobacterium tuberculosis*, yang menyebar ketika orang yang sakit TB mengeluarkan bakteri ke udara; misalnya dengan batuk. Penyakit ini biasanya mempengaruhi paru-paru (TB paru) tetapi juga dapat mempengaruhi tempat lain (TB ekstraparu). Sekitar seperempat populasi dunia

terinfeksi *mycobacterium tuberculosis* (WHO, 2020).

Indonesia sebagai salah satu negara yang masuk dalam kategori negara beban tinggi terhadap Tuberculosis, berada pada peringkat kedua setelah India. Di Provinsi NTB, pada tahun 2019 dilaporkan bahwa jumlah seluruh pasien TB (semua tipe) mencapai 6509 orang, sedangkan untuk tahun 2020, jumlah seluruh pasien TB adalah 5430 orang. Apabila dibandingkan dengan tahun 2019 maka kasus TB pada tahun 2020 mengalami penurunan sebesar 16,58%. TBC banyak ditemukan di daerah pemukiman padat penduduk dengan sanitasi yang kurang bagus. (Dinkes NTB, 2021).

Jumlah kasus TBC merupakan data *count* sehingga untuk mengetahui faktor-faktor yang berpengaruh terhadap kasus TBC digunakan analisis regresi Poisson. Salah satu contoh data kesehatan yang memiliki ciri-ciri percobaan poisson adalah jumlah kasus TB karena berupa data diskrit (*count*), yaitu banyaknya hasil percobaan yang terjadi dalam selang waktu tertentu atau di suatu daerah tertentu, peluang terjadinya suatu hasil percobaan tidak bergantung pada banyaknya hasil percobaan yang terjadi di luar selang waktu atau luar daerah tersebut dan peluang kesuksesan percobaan mengikuti sebaran distribusi Bernoulli (Walpole dan Myers, 1995).

Permasalahan yang muncul dalam pemodelan menggunakan regresi poisson adalah adanya asumsi yang harus terpenuhi yaitu asumsi nilai variabel respons memiliki nilai rata-rata dan varian yang sama (equidispersi). Kenyataan dalam kasus riil di lapangan asumsi equidispersi jarang terpenuhi karena data diskrit seringkali mengalami kasus overdispersi yaitu nilai varian lebih besar dari nilai rata-rata ($Var(y) > \mu$) dan memungkinkan juga terjadi underpersi yaitu nilai varian lebih kecil dari nilai rata-rata (Wang dan Famoye, 1997). Jika terjadi fenomena overdispersi pada data, maka regresi Poisson kurang akurat digunakan untuk analisis karena berdampak pada nilai penaksir simpangan baku koefisien regresi terlalu kecil. Sehingga nilai statistik uji untuk pengujian hipotesis dari pengaruh prediktor menjadi lebih

besar dan menjadikan prediktor terlalu mudah dianggap berpengaruh. Metode analisis regresi yang dapat mengatasi masalah underdispersi atau overdispersi yaitu metode Binomial Negatif (Cameron dan Trivedi, 1998). Distribusi Binomial Negatif merupakan pengembangan dari distribusi Poisson karena dihasilkan dari perpaduan antara distribusi Poisson dan distribusi Gamma. Jika nilai parameter dispersi pada regresi Binomial Negatif sama dengan nol, maka akan menghasilkan nilai varian sama dengan nilai rata-rata sesuai asumsi pada regresi Poisson (Cameron dan Trivedi, 1998).

Penelitian yang dilakukan oleh Banapon, dkk (2020) mengenai kasus Tuberculosis (TB) di Provinsi Jawa Barat Tahun 2017 dengan menggunakan regresi Binomial Negatif dan hasil penelitiannya menyatakan bahwa saat terjadi overdispersi pada data *count*, regresi Binomial Negatif lebih cocok digunakan dibandingkan regresi Poisson karena mempunyai nilai AIC yang lebih kecil. Model binomial negatif lebih tepat dan sesuai apabila digunakan pada data yang mengalami overdispersi atau underdispersi (Astuti, dkk., 2013).

Berdasarkan pemaparan di atas tujuan penelitian ini untuk membentuk model statistik jumlah kasus TB di NTB tahun 2020 dengan metode Binomial Negatif dan mengetahui adanya perbaikan goodness of fit model dari data yang mengalami overdispersi melalui indikator *deviance* dan AIC (*Akaike's Information Criteria*).

II. METODOLOGI PENELITIAN

Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh melalui publikasi data profil kesehatan di Dinas Kesehatan Nusa Tenggara Barat tahun 2021 dan Badan Pusat Statistik (BPS) Provinsi Nusa Tenggara Barat tahun 2021. Jumlah data penelitian yang digunakan adalah sebanyak 10 kabupaten/kota di Provinsi Nusa Tenggara Barat. Variabel-variabel yang digunakan dalam penelitian ini ditampilkan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
Y	Jumlah Kasus TB
X_1	Kepadatan Penduduk
X_2	Jumlah Puskesmas
X_3	Jumlah Tenaga Keperawatan
X_4	Persentase Rumah Tangga Yang Memiliki Akses Air Bersih
X_5	Persentase Penduduk Miskin
X_6	Persentase Rumah Tangga Yang Memiliki Akses Sanitasi Layak

Metode Analisis

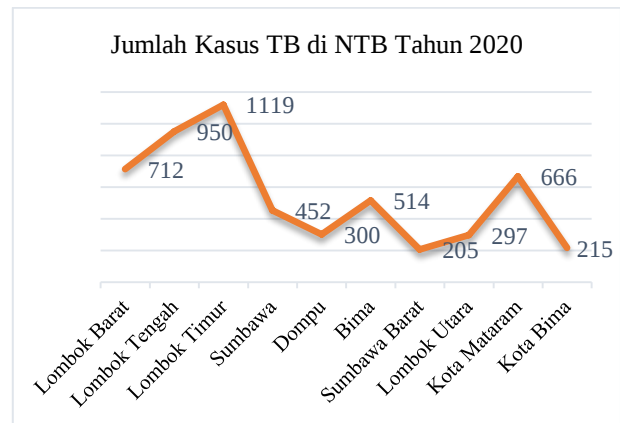
Langkah-langkah yang dilakukan dalam pengolahan data sekunder untuk mendapatkan pola hubungan terbaik dalam analisis regresi Binomial Negatif untuk mengatasi overdispersion pada regresi Poisson pada data jumlah kasus TB di NTB Tahun 2020 adalah sebagai berikut.

1. Mendeskripsikan karakteristik jumlah TBC di Provinsi NTB Tahun 2020
2. Mendeteksi kasus multikolineritas menggunakan kriteria nilai VIF
3. Pemodelan menggunakan regresi poisson
4. Mendeteksi adanya overdispersi pada regresi poisson
5. Pemodelan dengan menggunakan regresi binomial negatif dengan tahapan:
 - a. Melakukan penaksiran parameter model regresi binomial negatif
 - b. Melakukan pengujian signifikansi parameter pada model regresi binomial negatif
 - c. Menginterpretasikan hasil model regresi binomial negatif
6. Menarik kesimpulan.

III. HASIL DAN PEMBAHASAN

Karakteristik Penderita Penyakit TBC di Provinsi NTB

Jumlah seluruh pasien TB (semua tipe) di Provinsi NTB tahun 2020 dilaporkan mencapai 5430 orang dan sebanyak 250 orang kasus TB Anak usia 01-14 tahun. Jumlah semua kasus TB lebih banyak ditemukan pada laki-laki sebesar 59.7% atau 3239 orang. Distribusi jumlah penderita di tiap kabupaten/kota dapat dilihat pada Gambar 1.



Gambar 1. Grafik kasus TB di NTB Tahun 2020

Berdasarkan Gambar 1, daerah dengan jumlah kasus TB tertinggi yaitu Kabupaten Lombok Timur yaitu sebanyak 1119 kasus, disusul daerah Kabupaten Lombok Tengah sebesar 950 kasus. Sedangkan kasus TB terendah di NTB adalah Kabupaten Lombok Utara yaitu sebanyak 205 kasus.

Deteksi Multikolineritas

Sebelum melakukan analisis regresi poisson, perlu dilakukan pengecekan pada data untuk mengetahui apakah terjadi kasus multikolineritas atau tidak. Pemeriksaan multikolineritas pada data merupakan salah satu asumsi yang harus dipenuhi sebelum melakukan analisis regresi poisson, diantara variabel prediktor harus independen atau tidak ada hubungan antar variabel prediktor. Salah satu cara untuk mendeteksi multikolineritas yaitu dengan melihat nilai VIF (*Variance Inflation Factor*), apabila terdapat nilai VIF lebih besar dari 10 maka dapat disimpulkan terdapat multikolineritas pada data. Berikut adalah nilai VIF antar variabel prediktor.

Tabel 2. Nilai VIF Variabel Prediktor

Variabel	VIF	Variabel	VIF
X_1	1.889	X_4	1.766
X_2	1.209	X_5	1.601
X_3	1.048	X_6	1.791

Berdasarkan Tabel 2 tidak terdapat nilai VIF yang lebih dari 10, sehingga dapat disimpulkan bahwa tidak terdapat kasus multikolineritas pada data sehingga semua variabel layak diikutsertakan dalam pembentukan model.

Regresi Poisson

Setelah melakukan uji multikolinieritas, selanjutnya dilakukan analisis untuk mengetahui hubungan antar variabel prediktor terhadap jumlah kasus TB di Provinsi NTB dengan menggunakan analisis regresi poisson. Berikut adalah hasil estimasi parameter menggunakan regresi poisson.

Tabel 3. Estimasi Parameter Regresi Poisson

Parameter	Estimasi	Std. Error	Z value	P-value
Intercept	7.0303	0.5714	12.304	0.000
β_1	0.0002	0.0001	13.902	0.000
β_2	0.0753	0.0042	17.870	0.000
β_3	-0.0005	0.0001	-6.431	0.000
β_4	-0.0073	0.0057	-1.269	0.204
β_5	0.0109	0.0046	2.397	0.017
β_6	-0.0184	0.0026	-6.963	0.000

Berdasarkan Tabel 3 model regresi poisson yang terbentuk adalah :

$$\hat{\mu}_i = \exp(7.0303 + 0.0002X_1 + 0.0753X_2 - 0.0005X_3 - 0.0073X_4 + 0.0109X_5 - 0.0184X_6)$$

Dimana i menunjukkan banyaknya kabupaten dan kota di provinsi NTB.

Selanjutnya akan dilakukan pengujian parameter untuk mengetahui variabel prediktor mana yang berpengaruh terhadap variabel respon. Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 ; j = 1, 2, 3, 4, 5, 6$$

Berdasarkan hasil pengujian parameter dengan taraf signifikansi 5% terdapat lima parameter yang signifikan ($p\text{-value} < 0,05$) yaitu β_1 , β_2 , β_3 , β_5 dan β_6 sehingga dapat disimpulkan bahwa tolak H_0 yang artinya variabel X_1 , X_2 , X_3 , X_5 dan X_6 memberikan pengaruh yang signifikan terhadap model.

Pada analisis regresi poisson, salah satu asumsi yang harus dipenuhi yaitu keadaan *equisdispersion*. *Equisdispersion* adalah keadaan dimana taksiran dispersi yaitu nilai devians yang dibagi dengan derajat bebas adalah sama dengan 1. Apabila nilai taksiran dispersi diperoleh lebih dari satu maka terjadi overdispersi pada data. Diperoleh nilai devians pada model regresi poisson sebesar 51.651

dengan derajat bebas sebesar 3. Sehingga diperoleh nilai taksiran dispersi yang lebih dari 1, maka dapat disimpulkan bahwa terjadi kasus overdispersi pada data sehingga perlu dilakukan pemodelan lebih lanjut menggunakan regresi binomial negatif.

Regresi Binomial Negatif

Langkah awal dalam pemodelan regresi binomial negatif adalah penentuan nilai *initial* θ . Berdasarkan hasil *trial error initial theta* didapatkan *initial theta* sebesar 126.2 sehingga dilakukan pemodelan regresi Binomial Negatif dengan *initial theta* sebesar 126.2. hasil analisis regresi binomial negatif adalah sebagai berikut.

Tabel 4. Estimasi Parameter Regresi Binomial Negatif

Parameter	Estimasi	Std. Error	Z value	P-value
Intercept	7.1100	1.3760	5.166	0.000
β_1	0.0002	0.0004	5.585	0.000
β_2	0.0717	0.0098	7.300	0.000
β_3	-0.0004	0.0002	-2.401	0.016
β_4	-0.0080	0.0136	-0.589	0.556
β_5	0.0830	0.0087	0.945	0.345
β_6	-0.01724	0.0059	-2.942	0.003

Berdasarkan Tabel 4 model regresi binomial negatif yang terbentuk adalah

$$\hat{\mu}_i = \exp(7.1100 + 0.0002X_1 + 0.0717X_2 - 0.0004X_3 - 0.0080X_4 + 0.0830X_5 - 0.0172X_6)$$

Dimana i menunjukkan banyaknya kabupaten dan kota di provinsi NTB.

Selanjutnya akan dilakukan pengujian parameter untuk mengetahui variabel prediktor mana yang berpengaruh terhadap variabel respon. Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0 ; j = 1, 2, 3, 4, 5, 6$$

Berdasarkan hasil pengujian parameter dengan taraf signifikansi 5% terdapat empat parameter yang signifikan ($p\text{-value} < 0,05$) yaitu β_1 , β_2 , β_3 dan β_6 sehingga dapat disimpulkan bahwa tolak H_0 yang artinya variabel X_1 , X_2 , X_3 dan X_6 memberikan pengaruh yang signifikan terhadap model. Selanjutnya dilakukan penentuan model

regresi Binomial Negatif berdasarkan variabel prediktor yang signifikan yaitu X_1 , X_2 , X_3 dan X_6 . Hasil analisis regresi Binomial Negatif berdasarkan variabel yang signifikan disajikan dalam tabel berikut ini.

Tabel 5. Nilai Parameter Yang Signifikan

Parameter	Estimasi	p-value
Intercept	6.5730	0.000
β_1	0.0002	0.000
β_2	0.0704	0.000
β_3	-0.0004	0.043
β_6	0.0183	0.004

Berdasarkan Tabel 5 model regresi binomial negatif yang dihasilkan dapat dituliskan sebagai berikut.

$$\hat{\mu}_i = \exp(6.5730 + 0.0002X_1 + 0.0704X_2 - 0.0004X_3 + 0.0183X_6)$$

Penentuan model yang lebih baik antara model Regresi Poisson dan model Regresi Binomial Negatif pada data yang digunakan dapat dilihat dari hasil uji model terbaik dengan menggunakan nilai AIC dan nilai Deviance/df,. Model yang paling baik adalah model dengan nilai AIC terkecil.

Tabel 6. Kriteria Model

Model Regresi	AIC	Deviance/df
Poisson	145.43	17.217
Binomial Negatif	120.96	3.161
Binomial Negatif dengan variabel signifikan (X_1 , X_2 , X_3 , X_6)	119.08	1.940

Tabel 6 menunjukkan model terbaik adalah model regresi binomial negatif dengan variabel yang signifikan berdasarkan nilai deviance/df yang lebih mendekati 1 dan nilai AIC yang lebih kecil. Berdasarkan hasil analisis regresi binomial negatif dapat diketahui bahwa faktor yang berpengaruh terhadap jumlah penderita TBC di NTB tahun 2020 adalah kepadatan penduduk (X_1), jumlah puskesmas (X_2), jumlah tenaga keperawatan (X_3), dan persentase rumah tangga yang memiliki akses sanitasi layak (X_6).

Peningkatan atau penurunan dari jumlah kasus TB di NTB tahun 2020 tergantung dari koefisien variabel yang berpengaruh. Setiap

penambahan kepadatan penduduk sebesar 1 jiwa/km² maka akan menambah jumlah penderita TBC di NTB sebesar $\exp(0.0002) = 1.0002$ kali lipat. Setiap penambahan 1 unit puskesmas maka akan menurunkan jumlah penderita TBC di NTB sebesar $\exp(0.0704) = 1.073$ kali lipat, setiap penambahan 1 tenaga keperawatan maka akan menurunkan jumlah penderita TBC di NTB sebesar $\exp(0.0004) = 1.0004$ kali lipat dan setiap penambahan 1% rumah tangga yang memiliki akses sanitasi layak maka akan menurunkan jumlah kasus TBC di NTB sebesar $100(e^{0.0183} - 1)\% = 1.85\%$.

IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan dapat disimpulkan bahwa faktor yang mempengaruhi jumlah kasus TB di NTB Tahun 2020 berdasarkan model regresi binomial negatif adalah kepadatan penduduk, jumlah puskesmas, jumlah tenaga keperawatan, dan persentase rumah tangga yang memiliki akses sanitasi layak. Model terbaik untuk kasus TB di NTB tahun 2020 adalah model yang dihasilkan oleh metode *Binomial Negative* melalui indikator nilai AIC terkecil dengan model:

$$\hat{\mu}_i = \exp(6.5730 + 0.0002X_1 + 0.0704X_2 - 0.0004X_3 + 0.0183X_6).$$

Penelitian kedepannya, untuk mengatasi overdispersi pada regresi Poisson diharapkan dapat menggunakan model regresi yang lain seperti model regresi non linier dan model regresi quasi poisson serta menambah jumlah variabel prediktor yang diduga berpengaruh terhadap kejadian kasus TB.

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 : Universitas Brawijaya.
- Banapon, A., Putra, M. L. P., & Widodo, E. (2020). *Penerapan Regresi Binomial Negatif Untuk Mengatasi Pelanggaran*

Overdispersi Pada Regresi Poisson (Studi Kasus Penderita Tuberculosis Di Provinsi Jawa Barat Tahun 2017). Journal Biastatistics. Departemen Statistika Fmipa Universitas Padjadjaran, 14(1), 53-63.

Cameron, A.C. and Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

Dinkes NTB. 2020. *Profil Kesehatan Provinsi Nusa Tenggara Barat Tahun 2020*. Dinas Kesehatan Nusa Tenggara Barat: Dinkes NTB.

Walpole, R. E and Myers. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan, Terbitan ke-2*. Bandung: ITB.

Wang, W. dan Famoye, F., 1997. Modeling Household Fertility Decision with Generalized Poisson Regression. *Journal of Population Economics*, 10 (3) : 273-283.

WHO. (2020). *Global Tuberculosis Report*. World Health Organization.