

ANALISIS REGRESI *COUNT DATA* UNTUK PEMODELAN JUMLAH KASUS PENYAKIT TUBERKULOSIS DI KABUPATEN BANYUMAS

Kristy

Jurusan Matematika, Universitas Jenderal Soedirman
kristy@mhs.unsoed.ac.id

Jajang

Jurusan Matematika, Universitas Jenderal Soedirman
jajang@unsoed.ac.id

Nunung Nurhayati

Jurusan Matematika, Universitas Jenderal Soedirman

ABSTRACT. *Tuberculosis is an infectious disease caused by Mycobacterium tuberculosis. Banyumas Regency is one of the districts with quite high Tuberculosis cases in Central Java. This study aims to analyze the factors that affect the number of tuberculosis cases in Banyumas Regency using regression analysis of count data. Poisson regression is the simplest count data regression model that has the assumption of equidispersion, that is, the mean value equal to the variance. However, in its application, these assumption is often not fulfilled, for example, there are cases of overdispersion (variance value is greater than the mean). In this study, to overcome the case of overdispersion, an approach was used using Generalized Poisson Regression (GPR) and negative binomial regression. The results showed that the data on the number of tuberculosis cases in Banyumas Regency in 2019 was overdispersion. The data modeling of the number of tuberculosis cases in Banyumas Regency with the negative binomial regression model is better than the GPR model. Meanwhile, the only predictor variable that affects the number of tuberculosis cases in Banyumas Regency is the sex ratio of productive age (15-49 years).*

Keywords: *tuberculosis, count data regression, Generalized Poisson Regression (GPR), negative binomial regression.*

ABSTRAK. Tuberkulosis merupakan salah satu penyakit menular yang disebabkan oleh *Mycobacterium Tuberculosis*. Kabupaten Banyumas merupakan salah satu kabupaten dengan kasus *Tuberculosis* (TB) yang cukup tinggi di Jawa Tengah. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis di Kabupaten Banyumas menggunakan analisis regresi *count data*. Regresi Poisson merupakan model regresi *count data* paling sederhana yang memiliki asumsi equidispersi yaitu nilai *mean* sama dengan variansi. Namun, pada pengaplikasiannya asumsi tersebut sering tidak terpenuhi, misalnya terdapat kasus overdispersi (nilai variansi lebih besar daripada *mean*). Dalam penelitian ini, untuk mengatasi kasus

overdispersi dilakukan pendekatan menggunakan *Generalized Poisson Regression* (GPR) dan regresi binomial negatif.. Hasil penelitian menunjukkan bahwa data jumlah kasus tuberkulosis di Kabupaten Banyumas tahun 2019 mengalami overdispersi. Pemodelan data jumlah kasus tuberkulosis di Kabupaten Banyumas dengan model regresi binomial negatif lebih baik dibandingkan model GPR. Sementara itu, satu-satunya variabel prediktor yang berpengaruh terhadap jumlah kasus tuberkulosis di Kabupaten Banyumas yaitu rasio jenis kelamin usia produktif (15-49 tahun).

Kata kunci: tuberkulosis, regresi *count data*, overdispersi, *Generalized Poisson Regression* (GPR), regresi binomial negatif.

1. PENDAHULUAN

Tuberkulosis merupakan salah satu penyakit menular yang sering menyerang organ paru-paru dan proses penularannya terjadi melalui dahak. Penyakit ini disebabkan oleh adanya infeksi bakteri *Mycobacterium Tuberculosis*. Hingga saat ini kasus tuberkulosis masih menjadi permasalahan global, karena jumlah penderita tuberkulosis terus meningkat termasuk di Indonesia. Menurut data WHO (*World Health Organization*), Indonesia menduduki peringkat kedua dengan jumlah penderita tuberkulosis tertinggi di dunia (Kemenkes, 2020:153). Banyumas merupakan salah satu kabupaten di Jawa Tengah dengan jumlah kasus tuberkulosis yang cukup tinggi. Berdasarkan data Profil Kesehatan Kabupaten Banyumas penderita tuberkulosis pada tahun 2019 mencapai 932 kasus sedangkan pada tahun 2018 mencapai 1056 kasus. Meskipun terjadi penurunan sebesar 124 kasus di tahun 2019, jumlah kasus masih tergolong sangat tinggi. Oleh sebab itu, kasus ini memerlukan penanganan khusus guna meningkatkan kualitas kesehatan masyarakat khususnya dalam pencegahan penyakit tuberkulosis.

Data jumlah kasus tuberkulosis termasuk *count data* (data diskrit), sehingga untuk menganalisis data tersebut digunakan analisis regresi *count data*. Analisis regresi *count data* adalah suatu metode statistika yang digunakan untuk menganalisis hubungan antara variabel respon bertipe diskrit dengan satu atau lebih variabel prediktor. Model regresi *count data* yang paling sederhana adalah model regresi Poisson (Cameron dan Trivedi, 1998:59). Regresi Poisson mengasumsikan data berasal dari distribusi Poisson sehingga memenuhi sifat equidispersi, yaitu nilai *mean* sama dengan variansi. Akan tetapi pada

pengaplikasiannya asumsi equidisersi jarang ditemukan, sehingga sering terjadi pelanggaran asumsi tersebut, misalnya kasus overdispersi (nilai variansi lebih besar dari nilai *mean*). Untuk mengatasi kasus overdispersi pada model regresi Poisson digunakan metode *Generalized Poisson regression* (GPR) dan regresi binomial negatif.

Beberapa penelitian mengenai tuberkulosis telah dilakukan. Yuli dan Indriani (2015) telah melakukan penelitian mengenai pemodelan binomial negatif untuk mengatasi overdispersi data diskrit pada kasus baru Tb di provinsi Jawa Timur. Pada penelitian tersebut diperoleh faktor yang mempengaruhi adalah jumlah penduduk dan persentase UPK pelaksana DOTS. Penelitian lainnya dilakukan oleh Lestari dkk (2014) yang membahas tentang pemodelan faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis di Jawa Timur dengan pendekatan *Generalized Poisson Regression* (GPR) dan *Geographically Weighted Poisson Regression* (GWPR). Penelitian tersebut menunjukkan hasil bahwa pada model GPR faktor yang mempengaruhi adalah persentase penduduk usia produktif dan persentase TUPM sehat sedangkan pada model GWPR faktor yang mempengaruhi adalah persentase penduduk usia produktif, persentase TUPM sehat, dan persentase tenaga kesehatan terdidik tuberkulosis. Berdasarkan paparan sebelumnya, penulis tertarik melakukan penelitian mengenai faktor-faktor yang mempengaruhi jumlah kasus tuberkulosis di Kabupaten Banyumas tahun 2019 menggunakan analisis regresi *count data*.

2. METODE PENELITIAN

Penelitian ini dilakukan dengan metode studi pustaka dari buku teks dan jurnal. Data yang digunakan adalah data sekunder jumlah kasus tuberkulosis, rasio jenis kelamin usia produktif (15-49 tahun), persentase rumah tangga ber-PHBS, persentase balita kurang gizi, persentase keluarga pra-sejahtera, jumlah kasus HIV/AIDS, dan jumlah puskesmas di 27 kecamatan di Kabupaten Banyumas tahun 2019. Data diperoleh dari Dinas Kesehatan Banyumas, Dinas Pengendalian Penduduk Keluarga Berencana Pemberdayaan Perempuan dan Perlindungan Anak

(PPKBP3A) Kabupaten Banyumas, dan Badan Pusat Statistik (BPS) Kabupaten Banyumas. Langkah-langkah yang dilakukan dalam penelitian ini adalah:

1. Mendeskripsikan data.
2. Melakukan uji Kolmogorov-Smirnov untuk memeriksa variabel respon Y mengikuti distribusi Poisson atau tidak.

Menurut Daniel (1989:344) asumsi-asumsi pada uji Kolmogorov-Smirnov yaitu data terdiri atas sampel acak berukuran n (Y_1, Y_2, \dots, Y_n) dari suatu fungsi distribusi yang belum diketahui dan dinyatakan dengan $F(y)$. Statistik uji yang digunakan pada uji Kolmogorov-Smirnov adalah

$$D = \max |S(y) - F_0(y)|.$$

Kriteria pengujian yang digunakan, tolak H_0 apabila $p\text{-value} < \alpha$ atau nilai $D > D_\alpha$, dengan D_α nilai kritis pada tabel Kolmogorov-Smirnov.

3. Melakukan pemeriksaan multikolinearitas antar variabel prediktor menggunakan VIF. Multikolinearitas adalah sebuah kondisi ketika dua atau lebih variabel prediktor yang digunakan pada regresi berkorelasi sedang atau sangat tinggi (Mendenhall dan Sincich, 2012:363). Deteksi multikolinearitas dapat dilakukan dengan menggunakan VIF. Nilai VIF untuk variable prediktor j dirumuskan sebagai berikut:

$$(\text{VIF})_j = \frac{1}{1-R_j^2}, \quad j = 1, 2, \dots, p$$

dengan R_j^2 merupakan koefisien determinasi berganda. Multikolinearitas terjadi apabila nilai VIF lebih dari 10.

4. Melakukan pemodelan GPR dan pemodelan regresi binomial negative.
5. Melakukan uji signifikansi parameter GPR dan regresi binomial negatif secara serentak dan parsial. Uji serentak digunakan untuk melihat signifikansi parameter terhadap model secara serentak. Pada uji serentak digunakan uji *likelihood ratio* atau LR. Adapun rumus *likelihood ratio* yang didefinisikan (Agresti, 2007:84) adalah sebagai berikut:

$$\text{LR} = -2\{\mathcal{L}_0 - \mathcal{L}_1\}.$$

Jika $LR > \chi^2_{(\alpha;p)}$ maka tolak H_0 atau dengan kata lain sedikitnya ada satu $\beta_j \neq 0$ yang berpengaruh secara signifikan terhadap model. Sementara itu, uji parsial digunakan untuk melihat apakah masing-masing parameter berpengaruh signifikan terhadap model. Uji yang digunakan adalah uji Wald. Menurut Arisandi dkk. (2018) uji Wald dirumuskan sebagai berikut:

$$W_j = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \right)^2, \quad j = 1, 2, \dots, p.$$

Apabila $W_j \geq \chi^2_{(\alpha;1)}$ atau $p\text{-value} < 0,05$ maka tolak H_0 atau dengan kata lain $\beta_j \neq 0$ atau dengan kata lain β_j secara parsial signifikan terhadap model

6. Memilih model terbaik dengan memperhatikan nilai *Akaike Information Criterion* (AIC). Menurut Hilbe (2011:69) AIC diperoleh melalui persamaan berikut :

$$AIC = -2\mathcal{L} + 2k = -2(\mathcal{L} - k)$$

dengan \mathcal{L} adalah nilai *log-likelihood* dari model yang digunakan, dan k banyaknya parameter prediktor. Model terbaik adalah model yang memiliki nilai AIC terkecil.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Variabel respon yang digunakan pada penelitian ini adalah jumlah kasus tuberkulosis di Kabupaten Banyumas tahun 2019. Sementara itu, variabel prediktor yang digunakan yaitu rasio jenis kelamin usia produktif (15-49 tahun), persentase rumah tangga ber-PHBS, persentase balita kurang gizi, persentase keluarga pra-sejahtera, jumlah kasus HIV/AIDS, dan jumlah puskesmas. Hasil analisis data dari seluruh variabel terdapat pada Tabel 1.

Berdasarkan Tabel 1 rata-rata jumlah kasus tuberkulosis berkisar antara 34 hingga 35 kasus dengan nilai standar deviasi sebesar 16,04. Rata-rata rasio jenis kelamin usia produktif (15-49 tahun) sebesar 101,24% dengan standar deviasi

sebesar 1,28. Variabel persentase rumah tangga ber-PHBS memiliki rata-rata sebesar 86,15% dan standar deviasi sebesar 7,13. Kemudian rata-rata untuk variabel persentase balita kurang gizi sebesar 8,57% dan standar deviasi sebesar 2,50. Besar rata-rata persentase keluarga pra-sejahtera adalah 20,75% dengan standar deviasi sebesar 8,58. Rata-rata untuk variabel jumlah kasus HIV/AIDS berkisar antara 5 hingga 6 kasus dan standar deviasi sebesar 3,52. Sementara itu, rata-rata jumlah puskesmas berkisar antara 1 hingga 2 unit puskesmas dengan standar deviasi sebesar 0,51.

Tabel 1 Statistik deskriptif data penelitian

Variabel	Rata-rata	Standar Deviasi	Minimum	Median	Maksimum
Jumlah kasus tuberkulosis	34,52	16,04	12	32	66
Rasio jenis kelamin usia produktif (15-49 tahun)	101,24	1,28	97,95	99,96	102,12
Persentase rumah tangga ber-PHBS	86,15	7,13	66,00	86,06	98,80
Persentase balita kurang gizi	8,57	2,50	0,50	8,90	12,10
Persentase keluarga pra-sejahtera	20,75	8,58	5,79	20,82	35,56
Jumlah kasus HIV/AIDS	5,82	3,52	0,00	5,00	14,00
Jumlah puskesmas	1,44	0,51	1,00	1,00	2,00

3.2. Uji Distribusi Poisson

Sebelum melakukan pemodelan, terlebih dahulu akan dilakukan uji Kolmogorov-Smirnov terhadap jumlah kasus tuberkulosis untuk memeriksa apakah mengikuti distribusi Poisson. Adapun hipotesis yang digunakan sebagai berikut:

H_0 : Data jumlah kasus tuberkulosis berdistribusi Poisson.

H_1 : Data jumlah kasus tuberkulosis tidak berdistribusi Poisson.

Berdasarkan hasil analisis menggunakan uji Kolmogorov-Smirnov diperoleh nilai statistik uji Kolmogorov-Smirnov yaitu $D = 0,33197$. Nilai tabel

Kolmogorov-Smirnov dengan taraf signifikansi 5% dan derajat bebas 27 yaitu $D_{(0,05;27)} = 0,245$. Karena $D > D_{(0,05;27)}$ maka tolak H_0 yang artinya data jumlah kasus tuberkulosis di Kabupaten Banyumas tidak berdistribusi Poisson.

3.3 Uji Multikolinearitas

Salah satu metode yang digunakan untuk mendeteksi multikolinearitas adalah *Variance Inflation Factors* (VIF). Hasil uji VIF disajikan pada Tabel 2. Berdasarkan Tabel 3, dapat dilihat bahwa seluruh variabel prediktor memiliki nilai VIF kurang dari 10, sehingga dapat disimpulkan bahwa tidak terjadi multikolinearitas antar variabel prediktor.

Tabel 2 Hasil uji *Variance Inflation Factors* (VIF)

Variabel	VIF
Rasio jenis kelamin usia produktif (15-49 tahun) (X_1)	1,46
Persentase rumah tangga ber-PHBS (X_2)	1,40
Persentase balita kurang gizi (X_3)	1,61
Persentase keluarga pra-sejahtera (X_4)	1,13
Jumlah kasus HIV/AIDS (X_5)	1,50
Jumlah puskesmas (X_6)	1,31

3.4 *Generalized Poisson Regression* (GPR)

3.4.1 Pembentukan Model *Generalized Poisson Regression* (GPR)

Sebelum melakukan pemodelan GPR terlebih dahulu akan dicari estimasi parameter masing-masing variabel prediktor. Hasil estimasi parameter GPR disajikan pada Tabel 3.

Tabel 3 Hasil estimasi parameter GPR

Parameter	Estimasi	<i>Standard Error</i>
β_0	-16,4412	7,0497
β_1	0,1890	0,0654
β_2	0,0020	0,0118
β_3	0,0456	0,0351
β_4	- 0,0035	0,0091
β_5	0,0548	0,0289
β_6	0,1294	0,1522
AIC	222,5	

Berdasarkan hasil estimasi parameter yang diperoleh maka model GPR dapat dituliskan sebagai berikut:

$$\hat{\mu}_i = \exp(-16,4412 + 0,1890X_1 + 0,0020X_2 + 0,0456X_3 - 0,0035X_4 + 0,0548X_5 + 0,1294X_6).$$

3.4.2 Uji Signifikansi Parameter Secara Serentak

Selanjutnya melakukan uji signifikansi parameter secara serentak menggunakan metode uji *likelihood ratio*. Hipotesis yang digunakan adalah:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_6 = 0$$

$$H_1: \text{Paling sedikit ada satu } \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, 6.$$

Berdasarkan hasil perhitungan, diperoleh nilai LR yaitu 14,6 yang artinya nilai LR lebih besar dibandingkan nilai tabel *Chi-Square* yaitu $\chi_{(0,05;6)} = 12,59158$. Oleh karena itu, diperoleh keputusan tolak H_0 , sehingga dapat disimpulkan bahwa paling sedikit ada satu variabel prediktor memiliki pengaruh terhadap jumlah kasus tuberkulosis di Kabupaten Banyumas.

3.4.3 Uji Signifikansi Parameter Secara Parsial

Dengan menggunakan hipotesis berikut:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, 6$$

diperoleh keputusan hasil uji parsial yang dirangkum pada Tabel 4.

Tabel 4. Hasil uji parsial model GPR

Parameter	Uji Wald	Keputusan
β_1	8,362	Tolak H_0
β_2	0,028	Gagal Tolak H_0
β_3	1,693	Gagal Tolak H_0
β_4	0,145	Gagal Tolak H_0
β_5	3,593	Gagal Tolak H_0
β_6	0,723	Gagal Tolak H_0

Berdasarkan Tabel 4 diperoleh hanya satu parameter yang menolak H_0 yaitu β_1 sehingga dapat disimpulkan bahwa secara parsial variabel rasio jenis kelamin usia produktif (15-49 tahun) (X_1) memiliki pengaruh signifikan terhadap variabel respon. Parameter lainnya dengan keputusan gagal menolak H_0 yaitu β_2 , β_3 , β_4 , β_5 dan β_6 . Hal tersebut dapat diartikan bahwa variabel persentase keluarga ber-PHBS (X_2), persentase balita kurang gizi (X_3), persentase keluarga pra-sejahtera (X_4), jumlah kasus HIV/AIDS (X_5), dan jumlah puskesmas (X_6) secara parsial tidak memiliki pengaruh signifikan terhadap variabel respon.

Karena terdapat variabel yang tidak signifikan maka dilakukan regresi kembali dengan mengkombinasikan variabel prediktor. Adapun hasil yang diperoleh yaitu model GPR dengan satu variabel prediktor merupakan model GPR terbaik untuk data jumlah kasus tuberkulosis di Kabupaten Banyumas. Model GPR terbaik yang diperoleh adalah

$$\hat{\mu}_i = \exp(-14,4160 + 0,1789X_1)$$

dengan nilai AIC sebesar 219,40.

3.5 Regresi Binomial Negatif

3.5.1 Pembentukan Model Regresi Binomial Negatif

Pada bagian ini akan dilakukan pembentukan model regresi binomial negatif. Berikut ini hasil estimasi parameter model regresi binomial negatif yang diperoleh:

Tabel 5 Hasil estimasi parameter model regresi binomial negatif

Parameter	Estimasi	<i>Standard Error</i>
β_0	-15,9993	7,0198
β_1	0,1851	0,0654
β_2	0,0027	0,0115
β_3	0,0392	0,0350
β_4	-0,0035	0,0087
β_5	0,0465	0,0237
β_6	0,1268	0,1563
AIC	222,97	

Berdasarkan estimasi parameter diperoleh maka model regresi binomial negatif dapat dituliskan sebagai berikut:

$$\hat{\mu}_i = \exp(-15,9993 + 0,1851X_1 + 0,0027X_2 + 0,0392X_3 - 0,0035X_4 + 0,0465X_5 + 0,1268X_6).$$

3.5.2 Uji Signifikansi Parameter Secara Serentak

Pada tahap ini, digunakan nilai *likelihood ratio* yang akan dibandingkan dengan nilai tabel *Chi-Square* dengan taraf signifikansi 5% dan derajat bebas 6 yaitu $\chi_{(0,05;6)} = 12,59158$. Adapun hipotesis yang digunakan adalah:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_6 = 0$$

$$H_1: \text{Paling sedikit ada satu } \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, 6.$$

Nilai *likelihood ratio* yang diperoleh dengan bantuan R 3.6.3 sebesar 18,644. Karena $LR > \chi_{(0,05;6)}$, maka tolak H_0 , sehingga dapat disimpulkan bahwa paling sedikit ada satu variabel prediktor yang berpengaruh secara signifikan dalam model.

3.5.3 Uji Signifikansi Parameter Secara Parsial

Hipotesis yang digunakan dalam uji signifikansi parameter secara parsial adalah sebagai berikut:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0, \text{ untuk } j = 1, 2, \dots, 6.$$

Keputusan uji parsial model regresi binomial negatif dirangkum dalam Tabel 6.

Parameter	Uji Wald	Keputusan
β_1	8,008	Tolak H_0
β_2	0,055	Gagal Tolak H_0
β_3	1,255	Gagal Tolak H_0
β_4	0,168	Gagal Tolak H_0
β_5	3,871	Tolak H_0
β_6	0,658	Gagal Tolak H_0

Hasil uji parsial regresi binomial negatif memperlihatkan bahwa sebanyak dua parameter dengan keputusan menolak H_0 yaitu β_1 dan β_5 . Oleh karena itu, dapat disimpulkan secara parsial variabel prediktor rasio jenis kelamin usia produktif (15-49 tahun) (X_1) dan jumlah kasus HIV/AIDS (X_5) memiliki pengaruh signifikan terhadap variabel respon. Parameter lainnya yaitu β_2 , β_3 , β_4 dan β_6 memperoleh keputusan gagal menolak H_0 , sehingga dapat disimpulkan secara parsial variabel prediktor rumah tangga ber-PHBS (X_2), persentase balita kurang gizi (X_3), persentase keluarga pra-sejahtera (X_4), dan jumlah puskesmas (X_6) tidak memiliki pengaruh signifikan terhadap variabel respon.

Karena masih terdapat variabel prediktor yang tidak memiliki pengaruh signifikan terhadap variabel respon pada model regresi binomial negatif maka dilakukan regresi kembali dengan mengkombinasikan variabel prediktor. Adapun hasil yang diperoleh yaitu model regresi binomial negatif dengan satu variabel prediktor merupakan model regresi binomial terbaik dengan nilai AIC sebesar 219,31. Model regresi binomial negatif terbaik dapat dituliskan sebagai berikut:

$$\hat{\mu}_i = \exp(-14,6343 + 0,1811X_1).$$

3.6 Pemilihan Model Terbaik

Model terbaik ditentukan dengan memperhatikan nilai AIC terkecil dari model GPR dan regresi binomial negatif. Nilai AIC dari kedua model regresi dirangkum dalam Tabel 7.

Model	Nilai AIC
GPR	219,40
Regresi binomial negatif	219,31

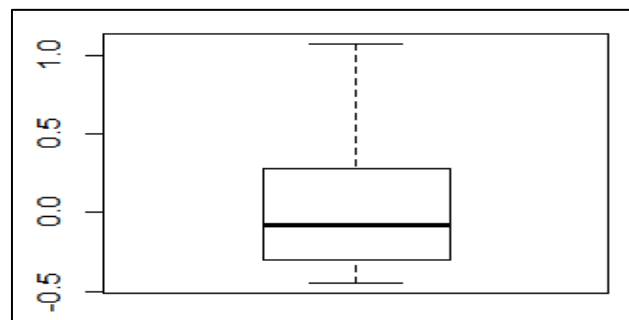
Tabel 7 menunjukkan bahwa model regresi dengan nilai AIC terkecil adalah model regresi binomial negatif yaitu sebesar 219,31. Oleh karena itu, dapat disimpulkan model terbaik untuk memodelkan jumlah kasus tuberkulosis di Kabupaten Banyumas adalah model regresi binomial negatif. Adapun model regresi binomial negatif adalah sebagai berikut:

$$\hat{\mu}_i = \exp(-14,6343 + 0,1811X_1).$$

Berdasarkan model terbaik yang diperoleh yaitu model regresi binomial negatif, maka dapat diinterpretasikan untuk variabel prediktor X_1 memiliki koefisien regresi bernilai positif dan nilai *odd ratio* sebesar $\exp(0,1811) = 1,1985$. Hal tersebut dapat diartikan, untuk setiap kenaikan 1% rasio jenis kelamin usia produktif (15-49 tahun) akan meningkatkan jumlah kasus tuberkulosis 1,1985 kali dari jumlah semula.

Menurut data Dinas Kesehatan Bayumas tahun 2019, pasien tuberkulosis usia 14 tahun keatas yang berjenis kelamin laki-laki lebih banyak dibandingkan perempuan. Dari seluruh kecamatan di Kabupaten Banyumas tercatat tahun 2019 terdapat 487 pasien laki-laki dan 386 pasien perempuan. Dalam hal ini, dapat dikatakan jumlah kasus tuberkulosis pada laki-laki 1,3 kali dibandingkan perempuan. Manalu (2010) menyatakan pada jenis kelamin laki-laki penyakit ini berisiko lebih tinggi karena kebiasaan merokok dan minum beralkohol sehingga menurunkan sistem pertahanan tubuh. Jika sistem pertahanan tubuh menurun maka lebih mudah terinfeksi bakteri tuberkulosis.

3.7 Argumentasi (Analisis Hasil Uji Signifikansi Model Terbaik)



Gambar 3.1 Boxplot nilai *residual*

Berdasarkan hasil model yang diperoleh yaitu model regresi binomial negatif, dilakukan analisis boxplot terhadap nilai *residual* model untuk mengetahui penyebab sedikitnya variabel prediktor yang signifikan. Berikut ini boxplot dari nilai *residual* model yang diperoleh dengan bantuan R 3.6.3.

Gambar 3.1 menunjukkan bahwa posisi median mendekati kuartil bawah (Q1), sehingga dapat dikatakan pola distribusi *residual* miring ke kanan atau *skewness* positif. Oleh karena itu, pada penelitian ini penyebab variabel prediktor yang signifikan dalam model hanya sedikit adalah pola distribusi *residualnya* yang *skewness* positif.

4. KESIMPULAN DAN SARAN

Berdasarkan analisis yang telah dilakukan diperoleh variabel respon jumlah kasus tuberkulosis (Y) tidak mengikuti distribusi Poisson, karena adanya pengaruh faktor overdispersi. Overdispersi diatasi dengan menggunakan model *Generalized Poisson regression* (GPR) dan regresi binomial negatif. Model GPR yang diperoleh untuk data jumlah kasus tuberkulosis di Kabupaten Banyumas tahun 2019 adalah

$$\hat{\mu}_i = \exp(-14,4160 + 0,1789X_1),$$

sedangkan model regresi binomial negatif yang diperoleh adalah

$$\hat{\mu}_i = \exp(-14,6343 + 0,1811X_1).$$

Nilai AIC model GPR dan binomial negatif berturut-turut adalah 219,40 dan 219,31. Dilihat dari nilai AIC maka dapat disimpulkan model regresi binomial negatif lebih baik dibandingkan model GPR dengan nilai AIC sebesar 219,31. Sementara itu, satu-satunya faktor yang berpengaruh pada jumlah kasus tuberkulosis di Kabupaten Banyumas adalah rasio jenis kelamin usia produktif (15-49 tahun) (X_1). Pada penelitian ini, model yang diperoleh hanya memiliki satu variabel yang signifikan sehingga model belum representatif. Sedikitnya variabel prediktor yang signifikan diakibatkan oleh standar deviasi yang tinggi, pola distribusi *residual* model yang *skewness* positif, dan ukuran sampelnya yang kecil. Oleh sebab itu, untuk penelitian selanjutnya, masalah sampel yang kecil disarankan menggunakan metode *resampling bootstrap* supaya sifat-sifat teorema limit pusat terpenuhi dan variansi sampelnya kecil.

DAFTAR PUSTAKA

Agresti, A., *An Introduction to Categorical Data Analysis*. John Wiley & Sons,

- Inc., Hoboken, 2007.
- Arisandi, A., Herdiani, E. T., Sahriman, S., Aplikasi Generalized Poisson Regression dalam Mengatasi Overdispersi pada Data Jumlah Penderita Demam Berdarah Dengue, *Statistika Journal of Theoretical Statistics and Its Applications*, **18**(2) (2018), 123–130.
- Cameron, A. C., Trivedi, P. K., *Regression Analysis of Count Data*, Cambridge University Press, Cambridge, 1998.
- Daniel, W. W., *Statistika Nonparametrik Terapan*, Gramedia, Jakarta, 1989..
- Hilbe, J. M., *Negative Binomial Regression*, 2nd Edition, Cambridge University Press, 2011..
- Kemenkes, *Profil Kesehatan Indonesia Tahun 2019*, Kementerian Kesehatan Republik Indonesia, Jakarta, 2020.
- Lestari, R. D., Wulandari, S. P., Purhadi, *Pemodelan Faktor-Faktor yang Mempengaruhi Jumlah Kasus Penyakit Tuberkulosis di Jawa Timur dengan Pendekatan Generalized Poisson Regression dan Geographically Weighted Poisson Regression*, *Jurnal Sains Dan Seni Pomits*, **3**(2) (2014), D-188-D-193.
- Manalu, H. S. P., *Faktor Faktor Yang Mempengaruhi Kejadian Tb Paru Dan Upaya Penanggulangannya*, *Jurnal Ekologi Kesehatan*, **9**(4) (2010), 1340–1346.
- Mendenhall, W., Sincich, T., *A Second Course in Statistics Regression Analysis*, 7th, Prentice Hall, Boston, 2012.
- Yuli, D., Indriani, D., *Pemodelan Binomial Negatif untuk Mengatasi Overdispersi Data Diskrit pada Kasus Baru Tb di Provinsi Jawa Timur*, *Jurnal Biometrika Dan Kependudukan*, **4**(2) (2015), 134–142.