

MODEL REGRESI BINOMIAL NEGATIF DALAM MENGANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI JUMLAH PENDERITA DIARE**Micha Annata Shinami**

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Ir. H. Soekarno, Surabaya, Indonesia
e-mail: mikaannata@gmail.com

Moh. Hafiyusholeh

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Ir. H. Soekarno, Surabaya, Indonesia
*e-mail: hafiyusholeh@uinsa.ac.id

Susilo Ari Wardani

Dinas Kesehatan Profinsi Jawa Timur, Jl. Ahmad Yani No.118, Ketintang, Kec. Gayungan, Surabaya, Indonesia
e-mail: ari_mmko@yahoo.co.id

Hani Khaulasari

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Ir. H. Soekarno, Surabaya, Indonesia
e-mail: hani.khaulasari@uinsa.ac.id

Aris Fanani

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel Surabaya, Jl. Ir. H. Soekarno, Surabaya, Indonesia
e-mail: arisfa@uinsa.ac.id

Abstrak

Kasus diare di Jawa Timur dari tahun ke tahun cenderung tinggi dengan lebih dari 600.000 kasus yang dipengaruhi oleh banyak faktor. Penelitian ini untuk menganalisis faktor yang berpengaruh signifikan terhadap jumlah penderita diare di Jawa Timur. Penelitian ini menggunakan regresi Binomial Negatif sebagai salah satu solusi untuk mengatasi overdispersi pada regresi Poisson. Data diambil dari Dinas Kesehatan Provinsi Jawa Timur dengan variabel terdiri atas akses sanitasi belum layak (X_1), jumlah sarana air minum (X_2) dan pengelolaan sampah rumah tangga (X_3). Hasil analisis regresi Binomial Negatif menunjukkan bahwa seluruh variabel independen secara simultan mempunyai pengaruh signifikan terhadap jumlah penderita diare (Y) sedangkan variabel pengelolaan sampah rumah tangga (X_3) berpengaruh signifikan secara parsial terhadap jumlah penderita diare (Y) dengan nilai AIC sebesar 805,699 dan nilai R-Square sebesar 66,6%, sedangkan 33,4% sisanya dijelaskan oleh faktor lain diluar model.

Kata Kunci: Diare, Overdispersi, Regresi Binomial Negatif.

Abstract

Diarrhea cases in East Java tend to be high from year to year with more than 600,000 cases influenced by many factors. This study aims to analyze factors that significantly influence the number of diarrhea sufferers in East Java. This study uses Negative Binomial regression as one solution to overcome overdispersion in Poisson regression. Data were taken from the East Java Provincial Health Office with variables consisting of inadequate sanitation access (X_1), number of drinking water facilities (X_2), and household waste management (X_3). The results of the Negative Binomial regression analysis show that all independent variables simultaneously have a significant influence on the number of diarrhea sufferers (Y) while the variable household waste management (X_3) has a partial significant effect on the number of diarrhea sufferers (Y) with an AIC value of 805.699 and an R-Square value of 66.6%, while the remaining 33.4% is explained by other factors outside the model.

Keywords: Diarrhea, Overdispersion, Negative Binomial Regression.

PENDAHULUAN

Diare termasuk salah satu masalah kesehatan yang dimana kondisi dalam tubuh seseorang mengalami buang air besar secara berlebihan dalam satu hari dengan mengeluarkan kotoran cair. Penyebab diare karena mengkonsumsi makanan dan minuman yang terkena bakteri. Selain itu faktor lingkungan juga sangat penting seperti fasilitas sanitasi yaitu ketersediaan air bersih, kamar mandi layak, tempai buang air besar layak, pembuangan air limbah dan pembuangan sampah (Sagara *et al.*, 2023). Diare juga dapat menyebabkan dehidrasi yang berbahaya seperti haus, kelelahan hingga pingsan apabila penanganannya tidak dilakukan dengan tepat dapat menyebabkan kematian, terutama pada balita, anak-anak dan lansia (Simadibrata *et al.*, 2024).

Berdasarkan data dari Dinas Kesehatan Provinsi Jawa Timur pada tahun 2024 termasuk dalam kategori tinggi untuk kasus diare yang tersebar di 38 Kabupaten/Kota, seperti Kota Surabaya terdapat 70.069 kasus, Kabupaten Sidoarjo terdapat 44.144 kasus dan Kabupaten Malang 30.159 kasus, artinya penyakit diare masih menjadi tantangan kesehatan masyarakat yang signifikan, meskipun berbagai program penanganan telah dilaksanakan seperti distribusi oralit ke puskesmas, pemberian oralit ketika pasien datang ke puskesmas dengan keluhan diare, serta sosialisasi kepada masyarakat terkait perilaku hidup yang bersih dan sehat akan tetapi jumlah kasus diare di Jawa Timur masih tinggi. Oleh karena itu, perlu ada identifikasi secara menyeluruh terhadap beberapa faktor yang secara dominan dapat mempengaruhi terjadinya diare.

Terdapat beberapa faktor yang diduga berpengaruh terhadap jumlah diare diantaranya yaitu sanitasi dan sarana air minum yang kurang memadai, penyediaan sanitasi dan sarana air minum (Merid *et al.*, 2023). Selain itu, pengelolaan sampah rumah tangga, imunisasi, akses air minum layak dan sanitasi layak juga berpotensi mempengaruhi diare (Jannah *et al.*, 2023)

Namun demikian, beberapa daerah dengan pengelolaan sampah rumah tangga yang tinggi, seperti Malang dan Sidoarjo, angka kasus diarenya sangat tinggi. Hal ini kemungkinan adanya faktor lingkungan lain yang belum teratasi dengan baik, seperti kualitas air minum serta perilaku hidup bersih dan sehat (Muttaqin, 2025).

Untuk mengidentifikasi beberapa faktor penyebab kasus diare tersebut dapat digunakan beberapa metode antara lain regresi Poisson (Tendriyawati *et al.*, 2023) dan regresi Binomial Negatif (Arum *et al.*, 2023). Regresi Poisson menggambarkan keterkaitan antara variabel dependen berdistribusi Poisson dengan variabel independen untuk data yang berisi jumlah kejadian, namun tidak dapat mengatasi permasalahan data yang variansnya lebih tinggi daripada nilai rata-rata (overdispersi). Untuk mengatasi masalah tersebut, pengolahan datanya dengan menggunakan regresi Binomial Negatif (Ryandi & Rifai, 2024).

Regresi binomial negatif termasuk salah satu bentuk *Generalized Linear Model* (GLM) yang digunakan untuk memodelkan data dengan nilai variansnya lebih besar dari rata-ratanya (overdispersi), parameter model ini di estimasi menggunakan metode *Maximum Likelihood Estimation* (MLE). Metode ini memiliki beberapa keunggulan diantaranya yaitu mampu mengatasi data dengan varians lebih tinggi daripada nilai rata-rata (overdispersi) yang tidak memenuhi asumsi pada regresi Poisson, dalam metode ini varians tidak harus sama dengan nilai rata-rata, dan mudah dalam menunjukkan variabel independen yang berpengaruh signifikan (Winata, 2023).

Beberapa penelitian yang membahas terkait dengan penerapan regresi Poisson, regresi Binomial Negatif maupun yang berhubungan dengan diare antara lain penelitian yang dilakukan oleh Fatmala *et al.*, (2024) dan Fitriani *et al.*, (2021) yang menyimpulkan bahwa model Binomial Negatif memiliki nilai *Akaike Information Criterion* (AIC) lebih rendah dibandingkan dengan model regresi Poisson, artinya kecocokan model yang lebih baik dengan data yang tersedia.

Penelitian ini memiliki tujuan menganalisis faktor yang berpengaruh terhadap jumlah penderita diare di Jawa Timur tahun 2024. Penelitian ini berbeda dengan penelitian terdahulu, penelitian ini mengambil kasus semua usia yang dilakukan di Provinsi Jawa Timur tahun 2024 dan berbeda dengan penelitian kasus diare pada balita oleh Widyaningrum *et al.*, (2021) yang menggunakan data tahun 2019 dengan menyatakan bahwa variabel yang signifikan terhadap jumlah penyakit diare pada balita adalah status gizi kurang, jumlah penduduk miskin dan posyandu.

Penelitian ini menggunakan metode regresi poisson, regresi binomial negatif dan menambahkan beberapa faktor yang berpengaruh terhadap jumlah penderita diare. Penelitian ini diharapkan dapat dimanfaatkan untuk merancang strategi dalam mencari solusi untuk mengatasi masalah kesehatan yang lebih efektif.

KAJIAN TEORI

KASUS DIARE DI JAWA TIMUR

Diare merupakan masalah kesehatan masyarakat yang ditandai dengan buang air besar secara berlebihan dalam satu hari dengan mengeluarkan kotoran yang cair. Penyebab diare bisa dari bakteri pada makanan dan minuman, berkaitan dengan kebersihan. Berdasarkan data dari Dinas Kesehatan Provinsi Jawa Timur, jumlah penderita diare tahun 2023 sebanyak 642.000 jiwa sedangkan di tahun 2024 jumlah penderita diare di Jawa Timur sebanyak 601.074 jiwa artinya di tahun ini mengalami penurunan sebanyak 40,926 jiwa.

Penanganan diare secara efektif dapat memperhatikan berbagai aspek, seperti faktor sanitasi dan faktor kebersihan lingkungan. Tingginya jumlah penderita diare yang mencakup 38 kabupaten/kota di Jawa Timur ini menunjukkan penanganan diare masih menjadi tantangan kesehatan masyarakat. Banyaknya kasus diare yang terjadi di Jawa Timur, Dinas Kesehatan Jawa Timur sudah melaksanakan berbagai upaya untuk mengatasinya seperti memberikan penyuluhan kesehatan masyarakat.

Faktor-faktor yang diidentifikasi berpengaruh terhadap kejadian diare diantaranya yaitu akses sanitasi belum layak, jumlah sarana air minum, dan pengelolaan sampah rumah tangga. Akses sanitasi yang belum layak menjadi salah satu determinan utama terjadinya diare. Rumah tangga yang menggunakan fasilitas buang air besar di tempat terbuka terbukti meningkatkan risiko penularan penyakit diare. Peningkatan kepemilikan jamban layak dan pengelolaan limbah rumah tangga juga berkontribusi nyata dalam menurunkan diare (Abidin *et al.*, 2022), akses terhadap air minum layak juga mampu menurunkan risiko diare hingga 30% (Susanti *et al.*, 2024).

Faktor lain yang juga penting yaitu pengelolaan sampah rumah tangga. Lingkungan dengan tumpukan sampah yang tidak terkelola akan menjadi tempat berkembang biak penyakit, seperti lalat, yang membawa agen penyebab diare. Penelitian di Kupang juga menegaskan pentingnya ketersediaan sarana pembuangan sampah umum dalam menurunkan diare (Adinata *et al.*, 2024). Dengan demikian, peningkatan akses sanitasi yang layak, penyediaan sarana air minum yang aman, pengelolaan sampah rumah tangga yang baik dan pengelolaan limbah domestik rumah tangga merupakan faktor kunci dalam upaya pencegahan dan pengendalian kejadian diare di masyarakat.

REGRESI POISSON

Regresi Poisson adalah salah satu jenis analisis regresi yang diterapkan pada data diskrit dan terdapat asumsi bahwa dalam analisis data, nilai varians harus sama dengan nilai rata-ratanya (equidispersi). Regresi poisson adalah penerapan dari *Generalized Linear Model* (GLM) yang menunjukkan keterkaitan antara variabel dependen data diskrit berdistribusi Poisson dengan beberapa variabel independen (Iin, 2022). Berikut merupakan persamaan model regresi Poisson:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (1)$$

Dimana:

μ_i : nilai rata-rata dari variabel Y pada observasi ke-i.

β_0 : nilai konstanta, yaitu nilai rata-rata μ saat semua variabel X sama dengan nol.

$\beta_1, \beta_2, \dots, \beta_k$: koefisien regresi untuk variabel X ke-1 sampai k.

X_1, X_2, \dots, X_k : nilai variabel X ke-1 sampai k.

Untuk mengestimasi parameter dari model regresi Poisson menggunakan metode *Maximum Likelihood Estimation* (MLE) sebagai berikut:

$$\begin{aligned} \ln L(y, \beta) &= \ln \left(\prod_{i=1}^n \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \right) \\ &= \sum_{i=1}^n y_i x_i^T \beta - \sum_{i=1}^n e^{x_i^T \beta} - \sum_{i=1}^n \ln(y_i!) \end{aligned} \quad (2)$$

Dimana:

$\ln L(y, \beta)$: fungsi log-likelihood dari parameter β

μ_i : nilai rata-rata dari distribusi Poisson

y_i : nilai respon

x_i : vektor baris dari variabel X

Untuk menganalisis menggunakan regresi poisson, yang harus dipastikan adalah tidak terjadinya multikolinearitas atau korelasi tinggi antar variabel independen yang umumnya digunakan pada semua jenis regresi termasuk regresi Poisson dan Binomial Negatif. Deteksi multikolinearitas yang digunakan yaitu *Variance Inflation Factor* (VIF) yang melibatkan beberapa variabel independen (Jeswanti *et al.*, 2022). Rumus VIF adalah sebagai berikut:

$$VIF = \frac{1}{1 - R_j^2} \quad (3)$$

Multikolinearitas antar variabel independen terjadi apabila nilai VIF > 10. Selanjutnya, asumsi utama pada regresi Poisson yaitu varians harus sama dengan nilai rata-rata atau disebut dengan equidispersi:

$$Var(Y_i) = \mu_i \quad (4)$$

Asumsi tersebut sering kali tidak terpenuhi dalam model ini, karena data dengan jumlah kejadian suatu peristiwa yang diolah menggunakan model regresi Poisson akan menghasilkan varians > nilai rata-rata ataupun sebaliknya dan kejadian itu disebut dengan overdispersi. Berikut merupakan rumus pengujian overdispersi:

1. Deviance Residual

$$\phi_1 = \frac{D^2}{db}$$

$$D^2 = 2 \sum_{i=1}^n \left\{ y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\} \quad (5)$$

Dengan db = n - k, dimana:

D^2 : jumlah *deviance residual*

k : jumlah parameter

n : jumlah observasi

2. Pearson Chi-Square

$$\phi_1 = \frac{x^2}{db}$$

$$x^2 = \sum_{i=1}^n \frac{(y_i - \mu_i)^2}{var(y_i)} \quad (6)$$

Dengan db = n - k, dimana:

x^2 : nilai *pearson chi-square*

k : jumlah parameter

n : jumlah observasi

Jika salah satu atau kedua nilainya > 1, maka model regresi Poisson tidak dapat dilakukan karena telah melanggar asumsi dasarnya. Jika terjadi overdispersi, maka salah satu solusi lain dalam menangani overdispersi yaitu model regresi Binomial Negatif.

REGRESI BINOMIAL NEGATIF

Regresi binomial negatif adalah salah satu alternatif dari model regresi Poisson yang mengalami *overdispersi*. Regresi Binomial Negatif adalah model penerapan dari *Generalized Linear Model* (GLM) yang menghubungkan antara variabel dependen dengan variabel independen yang dapat diterapkan pada kondisi equidispersi atau overdispersi (Palinoan *et al.*, 2024). Berikut merupakan persamaan model regresi Binomial Negatif:

$$\ln(\mu_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (7)$$

Dimana:

μ_i : nilai rata-rata dari variabel Y pada observasi ke-i.

β_0 : nilai konstantan, yaitu nilai rata-rata μ saat semua variabel X sama dengan nol.

$\beta_1, \beta_2, \dots, \beta_k$: nilai koefisien variabel X ke-1 sampai k.

X_1, X_2, \dots, X_k : nilai variabel X ke-1 sampai k.

Model ini fleksibel ketika data menunjukkan terjadinya overdispersi. Strukturnya sama dengan Poisson tetapi model ini menambahkan parameter dispersion (α) untuk menangani kelebihan variasi:

$$Var(Y_i) = \mu_i + \alpha \mu_i^2 \quad (8)$$

Estimasi parameter dilakukan dengan metode *Maximum Likelihood Estimation* (MLE) sebagai berikut:

$$L(\mu, y, \alpha) = \prod_{i=1}^n \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(\frac{1}{\alpha}) \Gamma(y_i + 1)} \left(\frac{1}{1 + \alpha \mu_i} \right)^{\frac{1}{\alpha}} \left(\frac{\alpha \mu_i}{1 + \alpha \mu_i} \right)^{y_i} \quad (9)$$

Dimana:

$L(\mu, y, \alpha)$: fungsi *likelihood* berdasarkan nilai μ, y, α

μ_i : nilai rata-rata

y_i : nilai respon

α : parameter dispersi

$\frac{1}{1 + \alpha \mu_i}$: faktor penyesuaian karena adanya overdispersi

$\left(\frac{1}{1 + \alpha \mu_i} \right)^{\frac{1}{\alpha}}$: bagian dari fungsi probabilitas Binomial Negatif

$\left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i}$:komponen probabilitas berdasarkan banyaknya kejadian

UJI SIGNIFIKANSI PARAMETER

UJI SIMULTAN

Dilakukannya uji simultan yaitu untuk menentukan pengaruh variabel independen terhadap variabel dependen secara keseluruhan dengan taraf signifikansi 0,05.

Hipotesis:

$H_0: \beta_j = 0$, tidak ada variabel X yang berpengaruh secara simultan terhadap variabel Y.

$H_1: \beta_j \neq 0$, minimal ada satu variabel X yang mempengaruhi variabel Y secara simultan.

Statistik Uji Likelihood Ratio (uji G):

$$G^2 = -2\ln\frac{1}{2}\left(\frac{L_0}{L_1}\right) = -2[\ln(L_0) - (L_1)] \quad (10)$$

Dimana:

L_0 : nilai *Likelihood* dari model tanpa variabel X.

L_1 : nilai *Likelihood* dari model lengkap dengan semua variabel X.

Daerah kritis:

Tolak H_0 jika statistik uji $G^2 > X^2_{(\alpha;k)}$, nilai $X^2_{(\alpha;k)}$ didapatkan dari tabel *chi-square* dengan $\alpha = 0,05$. Maka artinya seluruh variabel X berpengaruh signifikan secara simultan terhadap variabel Y dan paling tidak harus ada satu yang berpengaruh signifikan.

UJI PARSIAL

Dilakukannya uji parsial yaitu untuk mengetahui pengaruh parameter β secara individu terhadap variabel dependen dengan taraf signifikansi 0,05.

Hipotesis:

$H_0: \beta_j = 0$, tidak ada pengaruh signifikan antara variabel X terhadap variabel Y.

$H_1: \beta_j \neq 0$, ada pengaruh signifikan antara variabel independen X terhadap variabel Y.

Statistik uji Wald:

$$W = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right)^2 \quad (11)$$

Dimana:

$\hat{\beta}_j$: koefisien parameter hasil estimasi

$se(\hat{\beta}_j)$: standard error dari koefisien $\hat{\beta}_j$

W : statistik uji (wald statistik)

Daerah kritis:

Tolak H_0 jika *Wald Chi-Square* $\geq 3,841$ atau p-value $< 0,05$, maka artinya variabel X berpengaruh signifikan secara parsial terhadap variabel Y.

PEMILIHAN MODEL TERBAIK

Pemilihan model terbaik adalah proses untuk memilih model terbaik diantara beberapa model yang akan dibandingkan. Model terbaik yang dipilih yaitu nilai *Akaike Information Criterion (AIC)* terkecil.

$$AIC = -2\ln L(\beta) + 2k \quad (12)$$

Dimana,

$L(\beta)$: nilai *log-likelihood* dari model yang sedang diuji dengan parameter (β)

k : jumlah parameter

KOEFISIEN DETERMINASI (R^2)

R-Square digunakan untuk menunjukkan seberapa besar variasi dalam variabel dependen dapat dijelaskan oleh satu atau lebih variabel independen dalam suatu model regresi. Nilai *R Square* berkisaran antara 0 sampai 1, nilai yang mendekati 1 maka menunjukkan model regresi yang dapat menjelaskan sebagian besar variasi data dengan baik, nilai yang lebih kecil menunjukkan model tersebut kurang menjelaskan variasi variabel dependen. Berikut rumus untuk menghitung nilai R^2 :

$$R^2 = 1 - \frac{RSS}{TSS} \quad (13)$$

R^2 : koefisien determinasi

RSS : jumlah kuadrat residual

TSS : jumlah kuadrat total

METODE

DATA PENELITIAN

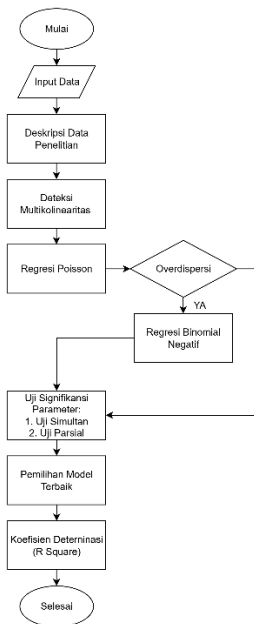
Peneitian ini menggunakan data yang diperoleh dari Dinas Kesehatan Provinsi Jawa Timur tahun 2024. Data tersebut mencakup informasi mengenai jumlah penderita diare semua usia di 38 kabupaten/kota Jawa Timur tahun 2024 dan beberapa variabel independen sebagai pendukung dari farmasi, program kesehatan dan kondisi

lingkungan. Variabel penelitian ini terdiri dari satu variabel dependen (Y) yaitu jumlah penderita diare semua usia dan beberapa variabel independen yaitu akses sanitasi belum layak (X_1), jumlah sarana air minum (X_2), dan pengelolaan sampah rumah tangga (X_3). Penelitian ini menggunakan metode regresi Binomial Negatif. Berikut merupakan rincian sampel datanya sebagai berikut:

Tabel 1. Data

Y	X_1	X_2	X_3
4.875	21.780	90	127.863
9.863	9.851	83	195.188
8.239	26.210	118	149.322
⋮	⋮	⋮	⋮
70.069	0	1	889.757
2.314	8.956	11	33.286

Untuk mengetahui rangkaian penelitian yang dilakukan, berikut diagram alur penelitiannya:



Gambar 1. Diagram Alur Penelitian

Berdasarkan diagram alur tersebut, maka tahapan penelitian yang dilakukan adalah:

1. Mengumpulkan data yang berupa data sekunder pada Dinas Kesehatan Provinsi Jawa Timur tahun 2024.
2. Melakukan deskripsi data untuk mengetahui terkait penderita diare serta faktor-faktor yang diduga berpengaruh.

3. Sebelum mengolah data menggunakan regresi poisson perlu melakukan deteksi multikolinearitas.
4. Setelah melakukan deteksi multikolinearitas, selanjutnya melakukan analisis model regresi Poisson, terdapat asumsi bahwa nilai varians harus sama dengan nilai rata-rata, jika asumsi tersebut tidak terpenuhi maka perlu dilakukan uji overdispersi dengan melihat nilai *deviance/df* dan *pearson chi-square/df*. Jika salah satu atau kedua nilainya > 1
5. Melakukan analisis model regresi Binomial Negatif untuk mengatasi adanya overdispersi.
6. Melakukan uji signifikansi terhadap parameter yaitu uji simultan dan uji parsial
7. Pemilihan model terbaik menggunakan nilai AIC terkecil.

HASIL DAN PEMBAHASAN

DESKRIPSI DATA PENELITIAN

Provinsi Jawa Timur memiliki 38 Kabupaten/Kota. Jumlah penderita diare sangat bervariasi antar kabupaten/kota, ada daerah yang kasusnya sangat tinggi dan ada yang rendah. Sebelum melakukan analisis regresi, maka perlu dilakukan pendeskripsian data penelitian sebagai berikut:

Tabel 2. Deskripsi Data Penelitian

Variabel	Rata-rata	St. Deviasi	Min	Max
Jumlah Penderita Diare (Y)	15817,74	13586,129	2005	70069
Akses Sanitasi Belum Layak (X_1)	18006,89	21368,964	0	601074
Jumlah Sarana Air Minum (X_2)	90,67	78,240	1	413
Pengelolaan Sampah Rumah Tangga (X_3)	210974,24	173251,469	28528	889757

Berdasarkan tabel 2 jumlah penderita diare (Y) memiliki nilai rata-rata sebesar 15.817,74 kasus diare. Selanjutnya nilai rata-rata pada akses sanitasi belum

layak (X_1) memiliki nilai rata-rata sebesar 18006,89 rumah tangga yang belum memiliki sanitasi layak di tiap kabupaten/kota. Jumlah sarana air minum (X_2) memiliki nilai rata-rata sebesar 90.67 unit sarana air minum misalnya sumur, sistem perpipaan, atau sarana air bersih lainnya. Pengelolaan sampah rumah tangga (X_2) memiliki nilai rata-rata sebesar 210974,24 rumah tangga yang sudah melakukan pengelolaan sampah.

Tahap selanjutnya adalah melakukan estimasi jumlah kejadian diare adalah dengan melakukan pendeteksian multikolinearitas.

DETEKSI MULTIKOLINEARITAS

Multikolinearitas merupakan adanya hubungan yang tinggi antara sesama variabel independen. Penelitian ini menggunakan *Variance Inflation Factors* (VIF) untuk mendeteksi multikolinearitas. Sebelum melakukan analisis menggunakan regresi Poisson, maka perlu melakukan deteksi multikolinearitas dalam rangka untuk memenuhi asumsi dari regresi Poisson. Berikut hasil deteksi multikolinearitas menggunakan nilai VIF:

Tabel 3. Deteksi Multikolinearitas menggunakan VIF

Variabel	VIF
Akses Sanitasi Belum Layak (X_1)	1,093
Jumlah Sarana Air Minum (X_2)	1,075
Pengelolaan Sampah Rumah Tangga (X_3)	1,018

Berdasarkan tabel 3 menunjukkan nilai VIF dari setiap variabel independen bernilai < 10 yang artinya semua variabel independen tidak terdeteksi adanya multikolinearitas.

PENGUJIAN DISTRIBUSI POISSON

Selanjutnya, setelah diketahui bahwa data tidak terdeteksi adanya multikolinearitas maka dapat dilakukan analisis model regresi Poisson. Berikut hasil estimasi parameter modelnya:

Tabel 4. Estimasi Parameter Regresi Poisson

Param	Estimasi	SE	p-value
β_0	8,741	0.0031	0,000
β_1	-0,000003910	0,000000061035	0,000
β_2	0,003	0,000015466	0,000

β_3	0,000002743	0,0000000054194	0,000
-----------	-------------	-----------------	-------

Berdasarkan Tabel 4, diperoleh model pada regresi Poisson:

$$\ln(\mu_i) = 8,741 - 0,000003910X_1 + 0,003X_2 + 0,000002743X_3$$

Berdasarkan hasil estimasi parameter dengan taraf signifikansi 0,05 menunjukkan seluruh variabel X berpengaruh signifikan terhadap jumlah penderita diare Y, ditunjukkan dengan nilai *p-value* 0,000 < 0,05. Model Poisson menunjukkan variabel akses sanitasi belum layak berpengaruh negatif terhadap jumlah kasus diare, sedangkan jumlah sarana air minum dan pengelolaan sampah rumah tangga berpengaruh positif, namun pengaruh ketiga variabel relatif sangat kecil.

Selanjutnya, dilakukan deteksi terjadinya overdispersi menggunakan nilai *deviance* dan *pearson chi-square* dibagi dengan nilai derajat bebas (df). Jika salah satu atau kedua nilainya > 1, maka model Poisson tidak lagi tepat digunakan karena asumsi dasarnya dilanggar. Berikut hasil uji overdispersi:

Tabel 5. Uji Overdispersi

Goodness of Fit	Value	df	Value/df
Deviance	126320,176	34	3715,299
Pearson Chi-Square	139411,345	34	4100,334

Berdasarkan Tabel 5, nilai *value deviance*/df sebesar 3715,299 dan *value Pearson Chi-Square*/df sebesar 4100,334, artinya nilai keduanya > 1 maka terjadi overdispersi. Maka model regresi Poisson tidak tepat digunakan, alternatif lain dengan menggunakan model regresi Binomial Negatif.

MODEL REGRESI BINOMIAL NEGATIF

Regresi Binomial Negatif bertujuan untuk memberikan solusi dalam mengatasi overdispersi. Berikut hasil estimasi parameter modelnya:

Tabel 6. Estimasi Parameter Regresi Binomial Negatif

Parameter	Estimasi	SE	p-value
β_0	8,596	0,3485	0,000
β_1	-0,000002756	0,0000077877	0,723
β_2	0,003	0,0024	0,160
β_3	0,000003042	0,0000010276	0,003

Berdasarkan Tabel 6, diperoleh model pada regresi Binomial Negatif:

$$\ln(\mu_i) = 8,596 - 0,000002756X_1 + 0,003X_2 + 0,000003042X_3$$

Berdasarkan model yang didapatkan dari hasil estimasi di atas, menunjukkan bahwa nilai β_0 signifikan, ketika seluruh variabel independen bernilai nol, maka rata-rata jumlah penderita diare adalah 8,596. β_1 bernilai negatif, peningkatan akses sanitasi belum layak cenderung menurunkan rata-rata jumlah penderita diare, namun pengaruhnya tidak signifikan karena nilai $p\text{-value} = 0,723 > 0,05$. β_2 bernilai positif, semakin banyak sarana air minum maka jumlah penderita diare cenderung meningkat namun pengaruhnya tidak signifikan karena nilai $p\text{-value} = 0,160 > 0,05$. β_3 bernilai positif dan signifikan dengan nilai $p\text{-value} = 0,003 < 0,05$, artinya terdapat pengaruh antara pengelolaan sampah rumah tangga terhadap jumlah penderita diare, menunjukkan bahwa setiap kenaikan 1 satuan pada pengelolaan sampah rumah tangga akan meningkatkan rata-rata jumlah penderita diare sekitar 0,0003%. Namun, untuk memastikan variabel mana yang berpengaruh signifikan secara statistik, perlu dilakukan pengujian simultan dan parsial.

Selanjutnya dilakukan pengujian simultan pada model regresi Binomial Negatif menggunakan nilai uji *Likelihood Ratio* (uji G) sebagai berikut:

Tabel 7. Uji G

<i>Likelihood Ratio Chi-Square</i>	df	<i>p-value</i>
13,139	3	0,004

Berdasarkan uji simultan (serentak) yang dilakukan pada Tabel 7 dengan model regresi Binomial Negatif didapatkan hasil nilai $G^2 -2 \log \text{likelihood}$ sebesar 13,139 dengan derajat bebas (df) = 3, artinya lebih besar dari nilai kritis $\chi^2_{(0,05;3)} = 7,815$ yang dapat dilihat pada tabel *Chi-Square* pada $p\text{-value}$ 0,05. Maka tolak H_0 artinya semua variabel independen berpengaruh signifikan secara simultan terhadap jumlah penderita diare semua usia (Y).

Setelah melakukan uji simultan, maka dilakukan pengujian parsial pada model regresi Binomial Negatif dengan menggunakan nilai uji *Wald* sebagai berikut:

Tabel 8. Uji *Wald*

Parameter	<i>Wald</i>	<i>p-value</i>
β_0	608,536	0,000
β_1	0,125	0,723
β_2	1,974	0,160

β_3	8,760	0,003
-----------	-------	-------

Berdasarkan uji parsial (masing-masing) yang dilakukan pada Tabel 8 dengan uji *Wald* pada model regresi Binomial Negatif didapatkan hasil bahwa hanya pada parameter β_3 dengan variabel pengelolaan sampah rumah tangga (X_3) yang menunjukkan pengaruh signifikan terhadap jumlah penderita diare semua usia (Y), dengan nilai *Wald* sebesar 8,760 dan nilai $p\text{-value}$ $0,003 < 0,05$, maka tolak H_0 artinya berpengaruh signifikan secara parsial. Artinya mengindikasikan bahwa pengelolaan limbah yang tidak sesuai standar dapat meningkatkan kejadian diare, karena sampah yang tidak terkelola dengan baik bisa menjadi sarang berkembang biaknya serangga dan mikroba. Serangga tersebut menjadi pembawa mikroorganisme berbahaya menuju makanan atau minuman, sehingga risiko terjadinya diare pun meningkat.

Sedangkan variabel X_1 dan X_2 memiliki nilai $p\text{-value} > 0,05$, maka artinya variabel tersebut tidak berpengaruh signifikan secara parsial terhadap jumlah penderita diare semua usia (Y), variabel tersebut belum menunjukkan hubungan yang terlalu kuat terhadap kasus diare di Jawa Timur.

PEMILIHAN MODEL TERBAIK

Pemilihan model terbaik didasarkan dengan nilai AIC terkecil. Berikut hasil nilai AIC kedua model regresi yang telah digunakan dalam penelitian ini:

Tabel 11. Pemilihan Model Terbaik

Model	AIC
Regresi Poisson	126753,836
Regresi Binomial Negatif	805,699

Berdasarkan tabel 11, diperoleh nilai AIC pada regresi poisson sebesar 126753,836, sedangkan nilai AIC pada regresi Binomial Negatif lebih kecil yaitu 805,699. Artinya bahwa nilai AIC pada model Binomial Negatif yang lebih kecil menunjukkan bahwa model tersebut menjelaskan variasi data dan estimasi yang lebih baik.

KOEFISIEN DETERMINASI (R^2)

Nilai R^2 yang semakin mendekati angka 1 menunjukkan bahwa variabel independen mampu menjelaskan sebagian besar variasi yang terjadi pada variabel dependen. Berikut nilai R^2 dalam penelitian ini:

Tabel 9. R Square

R	R Square	Adjusted R Square
0,816 ^a	0,666	0,636

Berdasarkan Tabel 9, diperoleh nilai R^2 sebesar 0,666, artinya kemampuan variabel independen dalam menjelaskan variabel dependen yaitu sebesar 66,6%, sedangkan 33,4% sisanya dijelaskan oleh faktor lain diluar model.

PEMBAHASAN

Hasil penelitian ini menggunakan model regresi Binomial Negatif untuk menganalisis pengaruh akses sanitasi belum layak (X_1), jumlah sarana air minum (X_2) dan pengelolaan sampah rumah tangga (X_3) terhadap jumlah penderita diare (Y) di kabupaten/kota Jawa Timur tahun 2024. Berdasarkan hasil analisis, variabel pengelolaan sampah rumah tangga berpengaruh signifikan terhadap jumlah penderita diare yang menunjukkan bahwa peningkatan kualitas pengelolaan limbah rumah tangga perlu menjadi prioritas utama dalam upaya pencegahan penyakit diare. Nilai AIC model Binomial Negatif yang didapatkan yaitu 805,699, dan nilai koefisien determinasi (R^2) sebesar 66,6%.

Hasil penelitian ini memiliki kesamaan maupun perbedaan dengan beberapa penelitian terdahulu. Jika dibandingkan dengan penelitian Widyaningrum *et al.*, (2021) menggunakan regresi *Robust* dan menghasilkan R^2 sebesar 100%, jauh lebih tinggi daripada penelitian ini. Perbedaan ini dikarenakan penelitian Widyaningrum fokus pada balita, sehingga faktor sosial ekonomi seperti status gizi kurang, jumlah penduduk miskin dan posyandu memiliki pengaruh signifikan. Sementara itu, penelitian ini menggunakan data semua usia, sehingga faktor lingkungan justru lebih dominan.

Pada penelitian Sulasih *et al.*, (2021) yang menggunakan model *Spatial Error Model* (SEM) menghasilkan R^2 sebesar 57,69%, nilai tersebut sedikit lebih rendah daripada penelitian ini. Faktor signifikan yang ditemukan berbeda, yaitu kepadatan penduduk dan jumlah sarana sanitasi yang menunjukkan bahwa penyebab diare dapat bervariasi antar wilayah. Penelitian Pratiwi, (2023) menghasilkan nilai AIC sebesar 267,558 dan R^2

sebesar 96,01% pada model SEM. Nilai AIC tersebut jauh lebih kecil dibandingkan dengan penelitian ini. Faktor signifikan yang ditemukan juga berbeda, yaitu jumlah sarana air minum dan jumlah sarana sanitasi. Namun, secara umum hasil penelitian Pratiwi menunjukkan bahwa ketika faktor spasial diperhitungkan, kemampuan model menjelaskan kasus diare menjadi lebih tinggi.

Pada penelitian Putri *et al.*, (2023) menggunakan *Geographically Weighted Regression* (GWR), diperoleh R^2 sebesar 79,29%, lebih tinggi dari penelitian ini. Variabel yang signifikan pada penelitian Putri adalah pemberian vitamin A pada balita, sehingga fokus penelitiannya berbeda dengan penelitian ini yang menekankan faktor lingkungan. Penelitian Faqih *et al.*, (2021) juga menggunakan *Geographically Weighted Regression* (GWR) dan mendapatkan nilai AIC yaitu -364,033 pada *kernel bisquare* serta R^2 sebesar 48,69%, dan nilai R^2 ini lebih rendah daripada penelitian ini. Variabel yang signifikan berbeda namun masih termasuk dalam faktor lingkungan yaitu PHBS dan rumah sehat yang berbeda pada penelitian ini.

Kesamaan utama beberapa penelitian tersebut yaitu faktor lingkungan dan sanitasi menjadi faktor penting dalam kejadian diare. Penelitian Widyaningrum, Sulasih, Pratiwi, Putri, maupun Faqih sama-sama menunjukkan bahwa kondisi rumah tangga, kualitas infrastruktur, kebersihan lingkungan, serta perilaku kesehatan masyarakat berperan besar dalam mendorong angka kejadian diare, meskipun jenis variabel yang signifikan berbeda pada setiap penelitian. Selain itu, hampir semua penelitian menunjukkan faktor yang diduga berpengaruh terhadap kejadian diare melibatkan interaksi antara lingkungan, perilaku kesehatan masyarakat, sanitasi, dan kondisi sosial ekonomi masyarakat.

Secara keseluruhan, kelima penelitian yang dibandingkan menunjukkan bahwa meskipun variabel signifikan berbeda, terdapat pola umum bahwa faktor lingkungan seperti sanitasi, kebersihan rumah tangga, maupun perilaku hidup bersih dan sehat merupakan faktor utama kejadian diare di Indonesia. Penelitian ini memperkuat bukti tersebut dengan menunjukkan bahwa pengelolaan sampah rumah tangga menjadi faktor paling berpengaruh di Jawa Timur tahun 2024. Oleh karena itu, intervensi peningkatan kualitas pengelolaan limbah rumah

tangga perlu dijadikan prioritas utama dalam upaya penanggulangan diare.

PENUTUP

SIMPULAN

Hasil dari riset ini menunjukkan bahwa hasil analisis regresi Binomial Negatif yang menangani masalah overdispersi menjelaskan bahwa variabel yang berpengaruh secara signifikan secara parsial terhadap jumlah kasus diare di semua usia (Y) adalah pengelolaan sampah rumah tangga (X_3) dengan p -value 0,003 lebih kecil dari 0,05. Diperoleh nilai AIC untuk regresi Binomial Negatif yang lebih kecil yaitu 805,699 dan nilai R-Square mencapai 0,666 atau 66,6%, sementara sisanya 33,4% dijelaskan oleh faktor lain yang tidak ada dalam model ini.

DAFTAR PUSTAKA

- Abidin, K., Ansariadi, A., & Thaha, I. L. M. (2022). Faktor Air, Sanitasi, Dan Higiene Terhadap Kejadian Diare Pada Balita Di Permukiman Kumuh Kota Makassar. *Hasanuddin Journal Of Public Health*, 3(3), 301–311. <https://doi.org/10.30597/Hjph.V3i3.22002>
- Adinata, M. N., Willy, T., & Fatah, M. (2024). Analisa Hubungan Antara Sanitasi Lingkungan Dan Kegiatan Mencuci Tangan Dengan Kejadian Diare Di Wilayah Puskesmas Alak Kota Kupang Analysis Of The Correlation Between Environmental Sanitation And Handwashing Activities With The Incidence Of Diarrhea In T. *Media Gizi Kesmas*, 13(2), 665–671.
- Arum, P. R., Manfaati Nur, I., Jihan Syafiqoh, A., & Rizky Utami, H. (2023). Permodelan Jumlah Kasus Tuberkulosis Di Kabupaten Purbalingga Tahun 2022 Menggunakan Regresi Binomial Negatif. *Journal Of Data Insights*, 1(2), 44–50. <https://doi.org/10.26714/Jodi.V1i2.273>
- Faqih, I. N., Aulia, F. A., Nurhidayat, H., & Arisanti, R. (2021). *Pemodelan Incident Rate Diare Di Kota Bandung Tahun 2020 Menggunakan Geographically Weighted Regression*.
- Fatmala, C. T., Hayati, M., Sari, R. P., & Hudori, M. (2024). *Pemodelan Jumlah Kasus Hiv / Aids Di Provinsi Lampung Menggunakan Regresi Binomial Negatif*. 6(2), 168–177. <https://doi.org/10.31605/Jomta.V6i2.4069>
- Fitrial, N. H., & Fatikhurriqzi, A. (2021). Pemodelan Jumlah Kasus Covid-19 Di Indonesia Dengan Pendekatan Regresi Poisson Dan Regresi Binomial Negatif. *Seminar Nasional Official Statistics*, 2020(1), 65–72. <https://doi.org/10.34123/Semnasoffstat.V2020i1.465>
- Iin, S. (2022). Regresi Poisson Dan Penerapannya Untuk Memodelkan Hubungan Usia Dan Perilaku Merokok Terhadap Jumlah Kematian Penderita Penyakit Kanker Paru - Paru. *Matematika Unaid*, 1(1), 71–76. <http://jmua.fmipa.unand.ac.id/index.php/jmua/article/download/20/17>
- Jannah, M., Susanti, R., & Ismail Ab. (2023). Faktor Yang Memengaruhi Kejadian Diare Pada Balita Di Indonesia Dengan Menggunakan Regresi Data Panel. *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, 3(1), 168–182. <http://jurnal.fmipa.unmul.ac.id/index.php/snmsa/article/view/1181>
- Mawar Jeswanti, Ruslan, Bahridin Abapihi, Baharuddin, Agusrawati, M. (2022). Penanganan Multikolinearitas Dalam Analisis Regresi Linear Berganda. *Departemen Statistika Fmipa Universitas Padjadjaranpenanganan Multikolinearitasdalam Analisis Regresi Linear Berganda Dengan Metode Partial Least Square(Studi Kasus: Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Di Sulawesi Tenggara)*Mawar Jes, 1–17.
- Merid, M. W., Alem, A. Z., Chilot, D., Belay, D. G., Kibret, A. A., Asratie, M. H., Shibabaw, Y. Y., & Aragaw, F. M. (2023). Impact Of Access To Improved Water And Sanitation On Diarrhea Reduction Among Rural Under-Five Children In Low And Middle-Income Countries: A Propensity Score Matched Analysis. *Tropical Medicine And Health*, 51(1). <https://doi.org/10.1186/S41182-023-00525-9>
- Muttaqin, A. (2025). *Hubungan Sanitasi Lingkungan Dengan Kejadian Diare Pada Balita Di Kelurahan Angke Kecamatan Tambora Kota Administrasi Jakarta Barat Tahun 2024*. 9(1), 104–116.
- Palinoan, K. A., Jaya, A. K., & Islamiyati, A. (2024). Pemodelan Regresi Binomial Negatif Menggunakan Estimator Jackknife Negative Binomial Ridge Regression Pada Data Angka

- Kematian Bayi Provinsi Sulawesi Selatan. *Basis: Jurnal Ilmiah Matematika*, 3(2), 1–8.
- Pratiwi, L. P. S. (2023). *Perbandingan Metode Regresi Spatial Auto Regressive Model Dan Spatial Error Model Untuk Kejadian Diare Di*. 9(1), 1–6.
- Putri, F. C., Suciptawati, N. L. P., & Susilawati, M. (2023). Implementasi Metode Geographically Weighted Regression (Gwr) Pada Kasus Diare Balita Di Provinsi Jawa Timur. *E-Jurnal Matematika*, 12(2), 92. <https://doi.org/10.24843/Mtk.2023.V12.I02.P405>
- Ryandi, A., & Rifai, N. A. K. (2024). Pemodelan Regresi Binomial Negatif Pada Kasus Jumlah Kematian Bayi Di Indonesia Tahun 2021. *Bandung Conference Series: Statistics*, 4(1), 258–264. <https://doi.org/10.29313/Bcss.V4i1.11841>
- Sagara, V., Aramico, B., & Arifin, V. N. (2023). Faktor Risiko Kejadian Diare Pada Anak Usia 1-4 Tahun Di Wilayah Kerja Puskesmas Batoh Kota Banda Aceh Tahun 2022. *Jurnal Fakultas Kesehatan Masyarakat*, 4(3), 2756–2762.
- Simadibrata, M., Prof, F., Syam, A. F., Susilo, A., Purnomo, H. D., Mariadi, K., Pribadi, R. R., Muzellina, V. N., Nursyirwan, S. A., Idrus, M. F., Shakinah, S., & Pasaribu, A. (2024). *Pasien Dewasa Di Indonesia Tahun 2024*.
- Sulasih, I. G. A. D., Susilawati, M., & Suciptawati, N. L. P. (2021). Pemodelan Kasus Diare Di Provinsi Bali Dengan Metode Analisis Regresi Spasial. *E-Jurnal Matematika*, 10(2), 95. <https://doi.org/10.24843/Mtk.2021.V10.I02.P327>
- Susanti, N., Rasyid, Z., Hasrianto, N., Redho, A., & Fadhli, R. (2024). Analisis Penyakit Diare Di Desa Cipang Kiri Hulu Dan Faktor Lingkungan Fisik Yang Mempengaruhinya. *Jurnal Kesehatan Lingkungan Indonesia*, 23(3), 374–381. <https://doi.org/10.14710/Jkli.23.3.374-381>
- Tendriyawati, Wibawa, G. N. A., & Abapihi, B. (2023). Pemodelan Regresi Poisson Terhadap Faktor-Faktor Yang Mempengaruhi Terjadinya Hipertensi Di Kota Kendari. *Jurnal Matematika Komputasi Dan Statistika*, 3(1), 255–262. <https://doi.org/10.33772/Jmks.V3i1.35>
- Widyaningrum, A. R., Susanti, Y., & Slamet, I. (2021). Pemodelan Penyakit Diare Balita Di Jawa Timur Menggunakan Regresi Robust. *Prosiding Seminar Nasional Sains*, 2(1), 522–528.
- Winata, H. M. (2023). Mengatasi Overdispersi Dengan Regresi Binomial Negatif Pada Angka Kematian Ibu Di Kota Bandung. *Jurnal Gaussian*, 11(4), 616–622. <https://doi.org/10.14710/J.Gauss.11.4.616-622>