



Prediksi Jumlah Sepeda yang Melintasi Williamsburg Bridge Menggunakan Regresi Binomial Negatif Berdasarkan Variabel Cuaca: Suhu dan Curah Hujan

Luh Sukma Mulyani^{1*}, Stefani Putri Wulandari², Marcellina Layata³, Ni Kadek Trisnawati⁴, I Wayan Sumarjaya⁵

¹⁻⁵Program Studi Matematika, FMIPA, Universitas Udayana, Indonesia

*Penulis Korespondensi: fnusukma@gmail.com

Abstract. *Negative Binomial Regression is a statistical modeling approach used to analyze count data with overdispersion, where the variance exceeds the mean. This study applies the method to examine the influence of weather factors on the daily number of cyclists crossing the Williamsburg Bridge in New York City. The independent variables used in the analysis include maximum temperature, minimum temperature, and precipitation. The dataset was obtained from the NYC Department of Transportation through the Kaggle platform and covers the period from April 1 to April 30, 2016. The analysis began with a Poisson Regression model; however, the presence of overdispersion was identified, indicated by a high AIC value of 8598.19, suggesting that the model was not suitable. The alternative Negative Binomial Regression model was then employed and produced a significantly lower AIC value of 518.77, demonstrating a superior fit. The findings indicate that maximum temperature has a positive effect on the number of cyclists, while precipitation shows a significant negative effect. Conversely, minimum temperature does not exhibit a meaningful influence. These results highlight the importance of considering weather conditions when planning bicycle-based transportation systems and support the development of sustainable mobility strategies in urban environments.*

Keywords: *Cyclist Mobility; Negative Binomial Regression; Overdispersion; Statistical Modeling; Weather*

Abstrak. Regresi binomial negatif merupakan metode analisis yang digunakan untuk memodelkan data cacah yang mengalami overdispersi, yaitu kondisi ketika variansi data lebih tinggi dibandingkan nilai rata-rata. Penelitian ini menerapkan model tersebut untuk menganalisis faktor cuaca yang memengaruhi jumlah pesepeda harian yang melintasi Williamsburg Bridge di New York City. Variabel yang digunakan dalam penelitian ini meliputi suhu maksimum, suhu minimum, dan curah hujan. Data yang digunakan merupakan data sekunder yang diperoleh dari NYC Department of Transportation melalui platform Kaggle untuk periode 1 April hingga 30 April 2016. Analisis diawali dengan estimasi model regresi Poisson sebagai model awal, namun hasil uji menunjukkan indikasi overdispersi dengan nilai AIC sebesar 8598,19 sehingga model tersebut dinilai kurang sesuai. Alternatif model regresi binomial negatif kemudian diterapkan dan menghasilkan nilai AIC yang jauh lebih rendah yaitu 518,77, menunjukkan kinerja prediksi yang lebih baik. Hasil analisis menunjukkan bahwa suhu maksimum berpengaruh positif terhadap jumlah pesepeda, sedangkan curah hujan memiliki pengaruh negatif yang signifikan. Sementara itu, suhu minimum tidak memberikan pengaruh yang berarti terhadap jumlah pesepeda. Temuan ini memberikan implikasi bahwa kondisi cuaca perlu dipertimbangkan dalam perencanaan transportasi berbasis sepeda terutama pada wilayah perkotaan yang mendukung mobilitas berkelanjutan.

Kata kunci: Cuaca; Mobilitas Pesepeda; Overdispersi; Pemodelan Statistik; Regresi Binomial Negatif

1. LATAR BELAKANG

Keterkaitan antara sebuah variabel dependen (respons) dengan satu atau lebih variabel independen (prediktor) dapat dikaji menggunakan pendekatan statistika yang dikenal sebagai analisis regresi (Montgomery et al., 2012). Apabila variabel dependen berbentuk diskret, sedangkan variabel independen dapat berupa diskret, kontinu, kategorik, maupun gabungan dari ketiganya, maka model yang sesuai untuk diterapkan adalah regresi Poisson (Tendriyawati et al., 2023). Model ini umumnya digunakan pada data yang mengikuti distribusi Poisson, misalnya jumlah kejadian dalam periode waktu tertentu atau di wilayah tertentu.

Salah satu asumsi penting dalam regresi Poisson adalah kesamaan antara nilai rata-rata dan variansi. Namun, dalam praktiknya, asumsi tersebut sering kali tidak terpenuhi. Banyak data diskret menunjukkan variansi yang lebih besar dibandingkan rata-ratanya, kondisi ini dikenal sebagai overdispersi (Rahmadeni & Sari, 2018). Ketika data menunjukkan variansi yang tidak setara dengan rata-ratanya, regresi binomial negatif menjadi pendekatan alternatif yang sesuai untuk menangani permasalahan tersebut (Rahmayanti & Rizki, 2018). Model ini mengasumsikan bahwa variabel dependen mengikuti distribusi binomial negatif.

Sejumlah penelitian sebelumnya telah membuktikan keunggulan regresi binomial negatif di berbagai konteks. Misalnya penelitian Widodo & Ariani (2018) tentang faktor-faktor penyebab penyakit DBD di Jawa Tengah menunjukkan bahwa regresi binomial negatif menghasilkan nilai AIC yang lebih kecil dibandingkan regresi Poisson. Studi lain mengenai jumlah kasus COVID-19 di Indonesia juga memperlihatkan bahwa model binomial negatif lebih tepat dalam memodelkan data berdasarkan nilai AIC (Fitrial & Fatikhurizqi, 2020). Selain itu, penelitian Rahmayanti & Rizki (2018) mengenai data klaim asuransi kendaraan bermotor menemukan bahwa regresi binomial negatif memberikan hasil yang lebih baik daripada regresi Poisson.

Selain itu, penerapan regresi binomial negatif juga telah banyak digunakan pada berbagai bidang penelitian mutakhir. Studi mengenai kriminalitas di Indonesia menunjukkan bahwa regresi binomial negatif mampu menangani masalah overdispersi pada data tindak kejahatan sehingga menghasilkan model yang lebih andal dibandingkan regresi Poisson (Dani, Fathurahman, Ni'matuzzahroh, Permata, & Putra, 2025). Pendekatan serupa juga diterapkan dalam bidang kesehatan masyarakat, misalnya untuk menganalisis faktor-faktor yang memengaruhi kasus stunting di Lombok Tengah, di mana model binomial negatif terbukti lebih sesuai karena karakteristik data yang mengalami overdispersi (Ananda et al., 2024). Dalam konteks pariwisata, penelitian Ramadan & Rantini (2024) di Jawa Barat membandingkan regresi Poisson dan binomial negatif untuk memodelkan jumlah kunjungan wisatawan, dan hasilnya menunjukkan bahwa model binomial negatif memberikan kinerja yang lebih baik dalam menjelaskan variasi jumlah pengunjung. Lebih lanjut, studi internasional terkait pola bersepeda di Amerika Utara menemukan bahwa cuaca berperan penting dalam memengaruhi jumlah pesepeda, dan analisis dengan model regresi, termasuk binomial negatif, digunakan untuk memahami dinamika tersebut selama periode sebelum, selama, dan setelah pandemi COVID-19 (Valenzuela et al., 2024).

Penerapan regresi binomial negatif telah meluas dalam berbagai bidang, seperti kesehatan, ekologi, dan transportasi. Dalam bidang transportasi, model ini relevan untuk

menganalisis pengaruh faktor lingkungan, khususnya kondisi cuaca, terhadap jumlah pengguna sepeda. Williamsburg Bridge di New York City, sebagai salah satu jalur utama pesepeda, menjadi lokasi penelitian yang ideal untuk penerapan metode ini. Jumlah pesepeda yang melintasi jembatan tersebut diketahui dipengaruhi oleh faktor cuaca, terutama suhu dan curah hujan.

Tujuan penelitian ini adalah memprediksi jumlah pesepeda yang melintasi Williamsburg Bridge dengan menggunakan variabel suhu dan curah hujan melalui model regresi binomial negatif. Penerapan pendekatan ini diharapkan tidak hanya memberikan pemahaman yang lebih baik mengenai pengaruh cuaca terhadap aktivitas bersepeda, tetapi juga menghasilkan model prediksi yang akurat dan aplikatif bagi perencanaan transportasi.

2. KAJIAN TEORITIS

Analisis regresi merupakan pendekatan statistika yang digunakan untuk menjelaskan keterkaitan antara suatu variabel terikat dengan satu atau beberapa variabel penjelas (Montgomery et al., 2012). Pada data yang berupa hitungan (*count data*), salah satu model yang paling banyak digunakan adalah regresi Poisson. Model ini berasumsi bahwa variabel respons mengikuti distribusi Poisson dengan karakteristik utama berupa kesetaraan antara mean dan varians (Tendriyawati et al., 2023).

Dalam praktik penelitian, asumsi tersebut kerap tidak terpenuhi karena data hitungan sering menunjukkan varians yang jauh lebih besar dibandingkan nilai rata-rata. Kondisi ini dikenal sebagai overdispersi dan menjadi salah satu indikator bahwa model Poisson kurang tepat diterapkan (Rahmadeni & Sari, 2018). Untuk menangani permasalahan tersebut, model regresi binomial negatif dikembangkan dengan penambahan parameter dispersi agar mampu mengakomodasi heterogenitas data dan menghasilkan estimasi yang lebih stabil (Rahmayanti & Rizki, 2018).

Sejumlah penelitian empiris telah menunjukkan bahwa regresi binomial negatif lebih unggul dibandingkan regresi Poisson dalam kondisi overdispersi. Widodo dan Ariani (2018) melaporkan bahwa model binomial negatif memberikan nilai AIC yang lebih kecil dalam kasus pemodelan penyakit menular. Hasil serupa juga ditemukan pada penelitian yang berkaitan dengan epidemiologi, kriminalitas, serta analisis data mobilitas populasi, di mana model binomial negatif memberikan performa estimasi yang lebih baik dan sensitif terhadap variabilitas data (Fitrial & Fatikhurriqzi, 2020; Dani et al., 2025). Selain itu, kajian terkait penggunaan sepeda menunjukkan bahwa faktor lingkungan, termasuk suhu dan curah hujan, memiliki kontribusi signifikan dalam perubahan intensitas aktivitas bersepeda. Pada penelitian

tersebut, regresi binomial negatif digunakan untuk mengakomodasi struktur data yang terindikasi mengalami overdispersi (Valenzuela et al., 2024).

Berdasarkan landasan teoritis dan temuan penelitian terdahulu, regresi binomial negatif dipandang sebagai model yang lebih tepat untuk diterapkan dalam analisis data cacah dengan variabilitas tinggi, khususnya ketika asumsi dasar model Poisson tidak terpenuhi karena adanya overdispersi.

3. METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif dengan metode observasional berbasis data sekunder. Analisis dilakukan untuk memodelkan hubungan antara faktor kondisi cuaca terhadap jumlah pesepeda yang melintas di beberapa jembatan East River menggunakan regresi pada data cacah (*count data*). Data penelitian merupakan data sekunder yang diperoleh dari platform Kaggle dengan judul New York City NYC East River Bicycle Crossings yang berasal dari New York City Department of Transportation. Dataset berisi data harian jumlah sepeda yang melintas pada empat jembatan di East River, yaitu Brooklyn Bridge, Manhattan Bridge, Williamsburg Bridge, dan Queensboro Bridge. Periode data yang digunakan adalah 1 April 2016 hingga 30 April 2016.

Populasi penelitian adalah seluruh lalu lintas pesepeda di East River, New York City. Sampel penelitian merupakan seluruh data yang tersedia dalam periode satu bulan pada dataset, yakni 30 data pengamatan harian. Penelitian ini melibatkan tiga variabel prediktor (X) dan satu variabel target (Y). Penjabaran lengkap mengenai masing-masing variabel dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel Target	Y	Williamsburg Bridge
Variabel Prediktor	X1	<i>High.Temp...F</i>
	X2	<i>Low.Temp...F</i>
	X3	<i>Precipitation</i>

Pengumpulan data dilakukan melalui proses pengunduhan dataset dari Kaggle. Data kemudian diekstraksi dan diolah menggunakan Microsoft Excel dan dianalisis menggunakan *software* R studio. Analisis dilakukan untuk memperoleh model terbaik yang mampu memprediksi jumlah pesepeda berdasarkan variabel cuaca pada wilayah East River, New York City. Analisis data dilakukan melalui beberapa tahapan sebagai berikut:

Uji Multikolinearitas

Salah satu cara untuk mendeteksi masalah multikolinearitas pada model regresi adalah dengan menggunakan *Variance Inflation Factor* (VIF). Apabila nilainya di atas 10, maka

model tersebut terindikasi memiliki multikolinearitas (Hair et al., 2018). Rumus statistik VIF dapat dituliskan sebagai berikut (Pratama & Wulandari, 2015):

$$VIF = \frac{1}{1 - R^2}.$$

Regresi Poisson

Regresi Poisson adalah teknik statistika yang diterapkan untuk memodelkan hubungan antara variabel independen dan variabel dependen dalam bentuk data cacah (*count data*). Variabel dependen diasumsikan mengikuti distribusi Poisson dengan fungsi probabilitas (Syafiqoh et al., 2024):

$$f(y, \mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots, \text{ dan } \mu > 0$$

Peubah acak y yang berdistribusi Poisson memiliki rata-rata sama dengan variansinya, yaitu sebesar μ . Dalam regresi Poisson, fungsi penghubung yang diterapkan adalah logaritma natural, yang dinyatakan sebagai berikut (Syafiqoh et al., 2024):

$$g(\mu_i) = \ln(\mu_i) = (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i)$$

$$\hat{\mu}_i = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}).$$

Uji Overdispersi

Overdispersi merupakan kondisi ketika varians variabel dependen lebih besar dari nilai rata-ratanya, yang umumnya muncul pada data dengan dominasi nilai nol (Syafiqoh et al., 2024). Identifikasi overdispersi dapat dilakukan melalui rasio devians terhadap derajat kebebasan, serta rasio *Pearson chi-square* terhadap derajat kebebasan. Model dikatakan sesuai apabila nilai keduanya lebih besar dari 1. Statistik uji dituliskan sebagai berikut (Sauddin et al., 2020):

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

Sedangkan untuk nilai *Pearson chi-square* diperoleh dari persamaan berikut (Sauddin et al., 2020):

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}.$$

Distribusi Binomial Negatif

Distribusi binomial negatif digunakan untuk memodelkan jumlah percobaan yang dibutuhkan hingga diperoleh keberhasilan pada percobaan yang saling bebas. Peluangnya dapat dituliskan sebagai berikut (Sauddin et al., 2020):

$$f(x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \quad x = k, k+1, k+2, \dots$$

Model Regresi Binomial Negatif

Model regresi binomial negatif dibangun dengan asumsi bahwa variabel dependen mengikuti distribusi binomial negatif. Bentuk persamaan model ditunjukkan sebagai berikut (Syafiqoh et al., 2024).

$$\ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (4)$$

$$\hat{\mu}_i = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i) \quad (5)$$

Estimasi Model Regresi Binomial Negatif

Parameter β dan α dalam model diestimasi menggunakan metode *maximum likelihood estimation* (MLE). Prosesnya meliputi (Saudidin et al., 2020):

- Menyusun fungsi *likelihood* dan kemudian mengubahnya menjadi fungsi *log-likelihood* untuk mempermudah perhitungan:

$$\ln L(\beta, \alpha) = \ln \prod_{i=1}^n \left\{ \prod_{r=0}^{y_i-1} (1 + \alpha r) \left(\frac{1}{y_i!} \right) \left(\frac{\exp(x_i' \beta)}{1 + \alpha \exp(x_i' \beta)} \right)^{y_i} \left(\frac{1}{1 + \alpha \exp(x_i' \beta)} \right)^{\frac{1}{\alpha}} \right\}$$

- Fungsi *log-likelihood* dimaksimalkan dengan mencari turunan parameter β dan α kemudian menyamakannya dengan nol. Estimasi nilai optimum β dan α dilakukan secara iteratif, salah satunya melalui metode Newton-Raphson yang banyak digunakan pada regresi binomial negatif.

Seleksi Model Regresi

Akaike Information Criterion (AIC) merupakan salah satu kriteria yang digunakan untuk mengevaluasi dan membandingkan kualitas relatif dari beberapa model statistika. Kriteria ini membantu menilai kecocokan model dengan mempertimbangkan jumlah parameter yang digunakan. Uji ini bertujuan membandingkan kinerja regresi Poisson dengan regresi binomial negatif, dengan rumus berikut (Famoye, 2010; Greene, 2012):

$$AIC = -2 \log L + 2v$$

Model yang dikembangkan bertujuan menjelaskan pengaruh kondisi cuaca terhadap jumlah pesepeda harian yang melintas di jembatan-jembatan East River, serta menentukan model regresi yang paling sesuai antara Poisson dan Binomial Negatif.

4. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil pengolahan dan interpretasi data mengenai jumlah pesepeda yang melintasi jembatan-jembatan di kawasan East River, New York City, selama periode 1–30 April 2016. Analisis ini bertujuan untuk mengkaji pengaruh faktor cuaca terhadap jumlah pesepeda setiap hari serta menentukan model prediksi yang paling tepat berdasarkan karakteristik variabel cuaca yang tersedia. Data yang digunakan merupakan data sekunder yang diperoleh dari NYC Department of Transportation melalui platform Kaggle.

Dataset yang dianalisis berisi informasi jumlah pesepeda harian yang melintas melalui Williamsburg Bridge sebagai variabel target penelitian, serta data pendukung mengenai kondisi cuaca yang terdiri dari suhu maksimum, suhu minimum, dan curah hujan harian. Williamsburg Bridge dipilih sebagai fokus analisis karena memiliki volume pesepeda yang relatif tinggi dibandingkan jembatan lainnya, sehingga dianggap representatif untuk mengevaluasi pengaruh faktor cuaca terhadap pergerakan pesepeda. Tahapan analisis diawali dengan penyajian statistik deskriptif terhadap seluruh variabel untuk memperoleh gambaran mengenai karakteristik dan distribusi data sebelum dilakukan pemodelan regresi.

Statistika Deskriptif

Statistik deskriptif untuk Williamsburg Bridge menunjukkan bahwa suhu maksimum berkisar antara 39,90°F hingga 81,00°F, dengan rata-rata 60,58°F dan median 62,10°F. Suhu minimum tercatat antara 26,10°F hingga 66,00°F, dengan rata-rata 46,41°F dan median 46,90°F. Presipitasi Sebagian besar tercatat 0.00 *inci*, dengan rata-rata 0.03667 *inci* dan maksimum 0,24000 *inci*. Jumlah pesepeda berkisar antara 1.440 hingga 7.834 pesepeda, dengan rata-rata 4.864 pesepeda. Fluktuasi suhu moderat, curah hujan rendah, dan variasi jumlah pesepeda tercatat selama periode pengamatan.

Tabel 2. Statistik Deskriptif

Variabel	Min	Median	Mean	Max
Y	1440	5194	4862	7834
X1	39,90	62,10	60,58	81,00
X2	26,10	46,90	46,41	66,00
X3	0,00000	0,00000	0,03667	0,24000

Uji Multikolinearitas

Tabel 3. Hasil Uji Multikolinearitas

Variabel	VIF
<i>High.Temp...F</i>	3,84
<i>Low.Temp...F</i>	3,73
<i>Precipitation</i>	1,24

Hasil analisis multikolinearitas pada model awal menunjukkan bahwa tidak terdapat masalah multikolinearitas yang signifikan. Variabel *Precipitation* memiliki nilai VIP sebesar

1,24, yang mengindikasikan bahwa variabel ini tidak berkorelasi kuat dengan variabel prediktor lainnya, sehingga bisa digunakan dalam model. Di sisi lain, variabel *High.Temp...F* dan *Low.Temp...F* menunjukkan nilai VIF masing-masing sebesar 3,84 dan 3,73. Meskipun nilai ini mengindikasikan adanya korelasi yang cukup kuat antara kedua variabel tersebut, keduanya masih berada dalam ambang batas toleransi multikolinearitas yang dapat diterima. Dalam konteks regresi Negatif Binomial, ambang batas VIF umumnya mengikuti kriteria umum, yaitu nilai $VIF \leq 5$ menunjukkan tidak adanya multikolinearitas yang bermasalah, nilai $5 < VIF \leq 10$ menunjukkan indikasi korelasi kuat namun masih dapat ditoleransi, sedangkan nilai $VIF > 10$ mengindikasikan multikolinearitas serius yang perlu ditangani (Hair et al., 2019). Oleh karena itu, tidak terdapat multikolinearitas yang mengganggu, dan model regresi Poisson dapat diterapkan pada tahap analisa berikutnya.

Model Regresi Poisson

Tabel 4. Nilai Estimasi Parameter Model Regresi Poisson

Parameter	Estimasi	Std. Error	Z value	sig
Y	7,064	0,1639	431,02	***
X1	0,026	0,0004	57,80	***
X2	-0,003	0,0005	-5,10	***
X3	-1,833	0,0503	-36,40	***

Berdasarkan Tabel 4 dapat dilihat hasil analisis menggunakan regresi Poisson menunjukkan bahwa semua variabel prediktor, yaitu *High.Temp...F*, *Low.Temp...F*, dan *Precipitation*, secara signifikan memengaruhi jumlah sepeda yang melintasi jembatan ($p\text{-value} < 0,001$). Berdasarkan hasil tersebut, persamaan regresi Poisson yang diperoleh adalah sebagai berikut:

$$\ln(\mu_i) = 7,064 + 0,026 \text{ High.Temp}_i - 0,003 \text{ Low.Temp}_i - 1,833 \text{ Precipitation}_i$$

$$\hat{\mu}_i = \exp(7,064 + 0,026 \text{ High.Temp}_i - 0,003 \text{ Low.Temp}_i - 1,833 \text{ Precipitation}_i)$$

Model yang dibangun menunjukkan performa yang cukup baik dalam menjelaskan variasi yang terdapat dalam data. Hal ini ditunjukkan oleh nilai *null deviance* sebesar 22227,4 dan *residual deviance* sebesar 8283,1, yang menandakan bahwa model berhasil mengurangi *deviance* secara signifikan setelah proses pemodelan dilakukan. Dengan kata lain, model mampu menangkap Sebagian besar variasi dalam data. Selain itu, nilai AIC sebesar 8,598.2 mengindikasikan bahwa model memiliki keseimbangan yang baik antara tingkat kompleksitas dan kemampuan prediksi, mengingat bahwa model dengan nilai AIC yang lebih rendah umumnya dianggap lebih optimal. Untuk memastikan kesesuaian lebih lanjut, langkah berikutnya adalah melakukan pemeriksaan terhadap kemungkinan adanya overdispersi dalam model.

Uji Overdispersi

Tabel 5. Hasil Uji Overdispersi

Statistik	Nilai
Overdispersion	318,5808
Pearson Chi-Square	300,102

Namun, nilai Overdispersion sebesar 318,58 dan Pearson chi-square sebesar 300,10, yang jauh lebih besar dari 1, menunjukkan adanya overdispersi dalam model. Hal ini mengindikasikan bahwa variabilitas data tidak sepenuhnya dapat dijelaskan oleh model Poisson. Sebagai Solusi terhadap gejala overdispersi yang ditemukan, pendekatan regresi binomial negatif dapat diterapkan sebagai alternatif yang lebih tepat guna menghasilkan model yang lebih stabil dan akurat.

Regresi Binomial Negatif

Berdasarkan hasil uji overdispersi yang mengindikasikan adanya variansi data yang melebihi nilai rata-rata, maka regresi binomial negatif dipandang sebagai pendekatan yang tepat untuk memodelkan jumlah sepeda yang melintasi East River Bridge. Model ini dipilih karena mampu mengakomodasi karakteristik data *count* dengan dispersi yang lebih besar dari ekspektasi, sehingga memberikan estimasi parameter yang lebih reliabel dibandingkan regresi Poisson.

Tabel 6. Nilai Estimasi Parameter Model Regresi Binomial Negatif

Parameter	Estimasi	Std. Error	Z value	sig
Y	6,753	0,272	24,838	***
X1	0,031	0,008	3,811	***
X2	-0,004	0,009	-0,394	
X3	-1,719	0,794	-2,166	*

Berdasarkan Tabel 6 diperoleh persamaan regresi binomial negatif sebagai berikut:

$$\ln(\mu_i) = 6,753 + 0,031 \text{ High.Temp}_i - 1,719 \text{ Precipitation}_i + \epsilon_i$$

$$\hat{\mu}_i = \exp(6,753 + 0,031 \text{ High.Temp}_i - 1,719 \text{ Precipitation}_i)$$

Berdasarkan persamaan di atas dapat dijelaskan bahwa intersep model sebesar 6,753 menunjukkan bahwa rata-rata jumlah sepeda yang melintasi jembatan diprediksi sebesar $e^{6.753} \approx 856,15$ ketika semua variabel prediktor bernilai nol. Variabel *High.Temp...F* memiliki pengaruh positif yang signifikan dengan koefisien 0,031, yang berarti setiap peningkatan 1°F pada suhu maksimum akan meningkatkan rata-rata jumlah sepeda sebesar 883. Sebaliknya, *Precipitation* menunjukkan pengaruh negatif yang signifikan dengan koefisien -1,719, yang mengindikasikan bahwa setiap kenaikan 0,1 inci presipitasi akan menurunkan rata-rata jumlah sepeda sebesar 721. Variabel *Low.Temp* tidak menunjukkan pengaruh signifikan terhadap jumlah sepeda.

Model ini memiliki *residual deviance* sebesar 30,354, jauh lebih rendah daripada *null deviance* 80,918, yang menunjukkan bahwa penambahan variabel prediktor meningkatkan kemampuan model dalam memprediksi data. Nilai AIC sebesar 518,77 menunjukkan efisiensi model, sementara parameter *disperse* sebesar 14,92 dengan kesalahan standar 3,83 mengonfirmasi bahwa model regresi binomial negatif sesuai untuk menangani overdispersi pada data. Secara keseluruhan, model ini menunjukkan bahwa suhu tinggi dan curah hujan berpengaruh signifikan terhadap jumlah sepeda yang melintasi jembatan, sementara suhu rendah tidak memiliki pengaruh yang signifikan.

Seleksi Model

Tabel 7. Perbandingan Nilai AIC

Model	AIC
Model Poisson	8598,195
Model Binomial Negatif	518,7659

Berdasarkan hasil perbandingan nilai AIC yang ditampilkan pada Tabel 7, model regresi Poisson memiliki nilai AIC sebesar 8598,195, sedangkan model regresi binomial negatif menunjukkan nilai AIC yang jauh lebih rendah, yaitu sebesar 518,7659. Perbedaan yang signifikan ini menunjukkan bahwa model binomial negatif memiliki kinerja yang lebih baik dalam memodelkan dan memprediksi jumlah sepeda yang melintasi Williamsburg Bridge berdasarkan variabel suhu dan cuaca. Dengan demikian, model binomial negatif dipandang lebih sesuai karena memberikan keseimbangan yang lebih optimal antara akurasi dan kompleksitas model.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis pengaruh variabel cuaca terhadap jumlah pesepeda yang melintasi Williamsburg Bridge serta menentukan model pemodelan yang paling sesuai untuk memprediksi jumlah pesepeda harian. Berdasarkan hasil analisis, diperoleh bahwa model regresi binomial negatif merupakan model terbaik, ditunjukkan oleh nilai AIC yang jauh lebih rendah dibandingkan model Poisson. Variabel suhu maksimum berpengaruh positif dan signifikan terhadap jumlah pesepeda, sedangkan curah hujan memiliki pengaruh negatif yang signifikan. Sementara itu, suhu minimum tidak menunjukkan pengaruh signifikan. Temuan ini menggambarkan bahwa kondisi cuaca memiliki keterkaitan kuat dengan pola penggunaan sepeda di wilayah tersebut.

Penelitian ini memiliki keterbatasan karena hanya menggunakan data satu bulan pada satu jembatan, sehingga generalisasi hasil harus dilakukan secara hati-hati. Untuk penelitian selanjutnya, disarankan memperluas cakupan data pada periode waktu yang lebih panjang dan

mempertimbangkan variabel tambahan seperti kecepatan angin, hari libur, dan data aktivitas transportasi lainnya untuk meningkatkan akurasi model prediksi.

DAFTAR REFERENSI

- Ananda, E. Y. P., Annas, S., Ihsan, H., Sukarna, & Aswi, A. (2024). Negative binomial regression analysis of factors influencing stunting cases in Central Lombok Regency. *Inferensi*, 7(3), 167–175. <https://doi.org/10.12962/j27213862.v7i3.21436>
- Dani, A. T. R., Fathurahman, M., Ni'matuzzahroh, L., Permata, R. P., & Putra, F. B. (2025). Exploring crime problems from a statistical point of view with negative binomial regression. *Varian*, 8(2), 199–208. <https://doi.org/10.30812/varian.v8i2.4445>
- Famoye, F. (2010). On the bivariate negative binomial regression model. *Journal of Applied Statistics*, 37(6), 969–981. <https://doi.org/10.1080/02664760902984618>
- Fitrial, N. H., & Fatikhurriqzi, A. (2020). 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>
- Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson Education Limited.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis*. <https://doi.org/10.1002/9781119409137.ch4>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). John Wiley & Sons. <https://doi.org/10.2307/2348362>
- Pratama, W., & Wulandari, S. P. (2015). Pemetaan dan pemodelan jumlah kasus penyakit tuberculosis (TBC) di Provinsi Jawa Barat dengan pendekatan geographically weighted negative binomial regression.
- Rahmadeni, & Sari, N. (2018). Solusi overdispersi menggunakan generalized Poisson regression (Studi Kasus: Penderita HIV di Provinsi Riau). *Jurnal Sains Matematika dan Statistika*, 4(2).
- Rahmayanti, D., & Rizki, S. W. (2018). Penanganan overdispersi dengan model binomial negatif pada data klaim asuransi kendaraan bermotor roda empat. *Buletin Ilmiah Math.Stat. dan Terapannya (Bimaster)*, 7.
- Ramadan, A., & Rantini, D. (2024). Application of negative binomial regression model in West Java tourism. *Journal of Advanced Technology and Multidiscipline*, 3(1), 9–12. <https://doi.org/10.20473/jatm.v3i1.57281>
- Sauddin, A., Auliah, N. I., & Alwi, W. (2020). Pemodelan jumlah kematian ibu di Provinsi Sulawesi Selatan menggunakan regresi binomial negatif. *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, 8(2), 42. <https://doi.org/10.24252/msa.v8i2.17409>
- Syafiqoh, A. J., Mahardika, R., Amaria, S., Winaryati, E., & Al Haris, M. (2024). Pemodelan regresi binomial negatif untuk mengevaluasi faktor-faktor yang mempengaruhi kasus tuberculosis di Provinsi Jawa Barat. *Jurnal Matematika dan Statistika serta Aplikasinya*, 12(1), 15-23. <https://doi.org/10.24252/msa.v12i1.39450>

- Tendriyawati, Wibawa, G. N. A., & Abapihi, B. (2023). Pemodelan regresi Poisson terhadap faktor-faktor yang mempengaruhi terjadinya hipertensi di Kota Kendari. *Jurnal Matematika dan Statistika*, 3(1), 255-262. <https://doi.org/10.33772/jmks.v3i1.35>
- Valenzuela, E. A., Barban, P., Beitel, D., Moreno, L. F. M., & Nguyen, V. T. Van. (2024). Analyzing the behavior and growth of cycling in four North American cities before, during, and after the COVID-19 pandemic. *Transportation Research Record*, 2678(12), 420–433. <https://doi.org/10.1177/03611981231157396>
- Widodo, E., & Ariani, P. M. (2018). Analisis faktor penyebab penyakit DBD di Jawa Tengah menggunakan regresi binomial negatif. *Jurnal Kesehatan Vokasi*, 3(1). <https://doi.org/10.22146/jkesvo.33870>