

# **Application of Negative Binomial Regression to Deal with Overdispersion in Identifying Factor Causing Crime Case Counts in Indonesia<sup>1</sup>**

Penerapan Regresi Binomial Negatif untuk Menangani Overdispersi dalam Identifikasi Faktor Penyebab Kasus Kejahatan di Indonesia

Aurally Budi Arini<sup>1‡</sup>, David Fauzan Akmal<sup>1</sup>, Adinda Pratiwi<sup>1</sup>, Rizky Kurniawan<sup>1</sup>, R. Mugni Chairil Arbi Asyari

Akbar Rizki<sup>1</sup>, and Akmal Riza Wibisono<sup>1</sup>

<sup>1</sup>Department of Statistics, IPB University, Indonesia

<sup>‡</sup>corresponding author: aurallyarini@apps.ipb.ac.id

## **Abstract**

Kejahatan adalah tindakan yang melanggar hukum, nilai, dan norma masyarakat. Berdasarkan data dari Global Organized Crime Index tahun 2021, Indonesia memiliki skor kriminalitas sebesar 6,38 dan menduduki peringkat ke-25 di dunia. Skor ini menunjukkan tingginya tingkat kriminalitas di Indonesia dan menjadi tantangan tersendiri dalam menciptakan lingkungan yang aman dan stabil. Kondisi ini mencerminkan belum tercapainya Tujuan Pembangunan Berkelanjutan (TPB) ke-16 yang mengedepankan perdamaian, keadilan, dan kelembagaan yang kuat. Penelitian ini bertujuan untuk menerapkan regresi binomial negatif untuk memodelkan data jumlah kasus kejahatan di Indonesia yang mengalami overdispersi dalam model regresi Poisson. Data yang digunakan berasal dari Badan Pusat Statistik (BPS) tahun 2021 mencakup jumlah kasus kejahatan dan delapan variabel penjelas dari 34 provinsi di Indonesia. Hasil analisis menunjukkan bahwa model regresi binomial negatif lebih baik dalam memodelkan data kejahatan yang mengalami overdispersi dibandingkan dengan regresi Poisson, berdasarkan nilai Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC) yang lebih rendah. Selain itu, didapatkan bahwa rasio jenis kelamin penduduk, tingkat hunian kamar hotel, dan jumlah penduduk miskin berpengaruh signifikan terhadap jumlah kejahatan.

## **1. Pendahuluan**

### **1.1 Latar Belakang**

Kejahatan merupakan masalah yang terjadi hampir di semua negara. Hal ini menyebabkan Perserikatan Bangsa-Bangsa (PBB) menetapkan perdamaian, keadilan, dan kelembagaan yang tangguh sebagai Tujuan Pembangunan Berkelanjutan (TPB) yang ke-16. Tujuan Pembangunan

---

<sup>1</sup> Received: Mei 2024; Reviewed: Mei 2024; Published: Mei 2024

Berkelanjutan telah disepakati pemimpin negara di dunia termasuk Indonesia (United Nations 2015). Selaras dengan Tujuan Pembangunan Berkelanjutan (TPB) yang ke-16, penting untuk mengatasi berbagai bentuk kejahatan yang mengganggu kestabilan masyarakat. Kejahatan yang merajalela tidak hanya merusak rasa aman, tetapi juga menghambat perkembangan institusi yang berfungsi untuk menjaga hukum dan keadilan. Kriminalitas atau tindak kejahatan merupakan suatu tindakan yang bertentangan dengan hukum, nilai, dan norma yang dianut di lingkungan masyarakat. Kejahatan memiliki dampak yang signifikan di kehidupan masyarakat. Banyak faktor yang dapat memengaruhi suatu individu untuk terlibat dalam tindakan kejahatan tersebut. Menurut *Global Organized Crime Index* pada tahun 2021, Indonesia memiliki skor kriminalitas sebesar 6.38 dan menduduki peringkat ke-25 di dunia dengan skor tersebut. Indonesia juga diklasifikasikan sebagai *High Criminality - Low Resilience* yang berarti ketahanan negara masih rendah dengan angka kriminalitas yang masih tinggi sehingga penting untuk dilakukan analisis faktor-faktor menyebabkan tingginya kasus kejahatan di Indonesia agar kebijakan pemerintah dan militer dapat diperbaharui kembali sebagai upaya menurunkan kasus kejahatan tersebut.

Faktor-faktor yang memengaruhi kejahatan di Indonesia dapat menjadi fokus utama pemerintah dan militer dalam mengatasi kejahatan serta warga Negara Indonesia untuk meningkatkan kewaspadaan. Penelitian terkait kasus kejahatan sebelumnya pernah dilakukan oleh Maulana (2014) dengan memodelkan faktor-faktor yang dapat memengaruhi tingkat kejahatan pencurian di Kota Semarang menggunakan metode analisis regresi linear berganda dengan salah satu variabel dependen yang signifikan adalah tingkat pendidikan. Jika ditinjau dari faktor ekonomi, terdapat faktor pengangguran terbuka yang pernah diteliti oleh Effendi dan Julia (2021), upah minimum penduduk yang telah diteliti oleh Erlangga (2019), dan jumlah penduduk miskin oleh Lumenta et al. (2012). Faktor-faktor lain yang telah terbukti signifikan dalam penelitian sebelumnya meliputi rasio jenis kelamin penduduk dan tingkat penghunian kamar hotel (Sudarsono 2012), persentase penduduk yang mengakses internet (Millah 2020), serta jumlah penduduk (Handayani 2017).

Kejadian kejahatan merupakan data cacahan sehingga salah satu jenis analisis regresi yang dapat digunakan adalah regresi Poisson. Adapun asumsi yang harus dipenuhi pada regresi Poisson yaitu ekuidispersi. Ekuidispersi terjadi ketika ragam dan nilai tengah dari peubah respon bernilai sama besar. Asumsi ekuidispersi menjadi tidak terpenuhi karena terjadinya overdispersi, yaitu ketika nilai ragam lebih besar daripada nilai tengah pada peubah respon. Penggunaan regresi Poisson pada data overdispersi dapat menyebabkan model yang terbentuk menghasilkan estimasi parameter yang berbias (Lestari dan Puji 2014). Untuk mengatasi masalah ini, regresi binomial negatif dapat menjadi alternatif yang lebih sesuai. Regresi binomial negatif telah menjadi pendekatan yang umum digunakan untuk memodelkan

data yang mengalami overdispersi. Oleh karena itu, dalam penelitian ini, akan diterapkan regresi binomial negatif untuk memodelkan jumlah kasus kejahatan di Indonesia dengan tujuan untuk mengatasi overdispersi yang terjadi pada regresi Poisson.

## 1.2 Tujuan

Tujuan Penelitian ini adalah sebagai berikut

1. Mengidentifikasi faktor yang memengaruhi jumlah kejahatan di Indonesia.
2. Mengetahui performa regresi binomial negatif untuk memodelkan data jumlah kasus kejahatan di Indonesia yang mengalami overdispersi.

## 2. Tinjauan Pustaka

### 2.1 Kriminalitas

Kriminalitas mengacu pada berbagai perilaku yang menyebabkan kerugian ekonomi maupun psikologis, melanggar hukum di Indonesia, serta bertentangan dengan nilai-nilai sosial dan agama. Dengan demikian, tindakan kriminal dapat didefinisikan sebagai segala perilaku yang bertentangan dengan hukum dan norma-norma sosial, yang menimbulkan keberatan dari masyarakat. Terdapat dua jenis faktor yang memengaruhi tingkat kriminalitas, yaitu faktor internal dan eksternal. Faktor internal meliputi kebutuhan ekonomi yang mendesak, kondisi ketenagakerjaan (termasuk pengangguran dan pekerja), serta tingkat kesejahteraan individu. Sementara itu, faktor eksternal mencakup tingkat pendidikan dan pengaruh lingkungan sosial.

### 2.2 Regresi Poisson

Regresi Poisson termasuk salah satu bentuk analisis regresi yang berada dalam kerangka Generalized Linear Model (GLM). GLM merupakan perkembangan dari analisis regresi yang tidak mengharuskan asumsi tentang distribusi normalitas dan dapat diterapkan pada variabel respons yang berasal dari keluarga eksponensial. Regresi Poisson digunakan untuk menemukan hubungan antara variabel penjelas dan variabel respons yang berupa data hitungan (Agresti, 2019). Teknik ini mengacu pada distribusi Poisson yang memiliki fungsi massa peluang.

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

Hilbe (2014) menjelaskan bahwa dalam model regresi Poisson,  $\mu$  mewakili rata-rata jumlah kejadian, sedangkan  $y$  adalah jumlah peristiwa yang terjadi dalam periode waktu tertentu. Model regresi Poisson didefinisikan dengan menggunakan fungsi rata-rata variabel respons  $\mu_i$ , yang dapat dijelaskan sebagai berikut:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = x_i' \beta \quad (2)$$

dengan  $\ln(\mu_i)$  merupakan fungsi penghubung log dari rata-rata respon,  $\beta_0$  merupakan intersep,  $\beta_k$  merupakan koefisien untuk setiap peubah penjelas ke- $k$ .  $x_i$  merupakan vektor peubah penjelas yang berukuran  $(1 \times (k + 1))$  dan  $\beta$  merupakan vektor parameter regresi yang berukuran  $((k + 1) \times 1)$ .

### 2.3 Overdispersi

Overdispersi adalah keadaan di mana variabilitas suatu variabel respons jauh lebih besar daripada yang diharapkan sebanding dengan nilai rata-ratanya. Hal ini dapat terjadi karena beberapa alasan, seperti faktor-faktor yang tidak teramati atau tidak diperhitungkan secara tepat dalam analisis, keberadaan outlier atau data yang hilang pada variabel independen, serta pengelompokan data yang tidak merata dalam populasi. Dalam konteks statistik, overdispersi sering teridentifikasi dalam model-model yang mengasumsikan bahwa varians respons seharusnya sama dengan rata-rata respons (contohnya, dalam model regresi linier), tetapi kenyataannya variansnya jauh lebih besar. Hal ini dapat menyebabkan estimasi yang tidak konsisten dan tidak akurat, serta mengganggu interpretasi hasil analisis statistik. Oleh karena itu, penting untuk mengenali dan memperhitungkan overdispersi dalam analisis data untuk memastikan hasil yang lebih reliabel dan valid.

Kondisi overdispersi bisa terlihat saat deviance dibagi dengan derajat kebebasannya, dan jika nilai  $\phi_1$  melebihi 1, itu menandakan terjadi overdispersi pada data, dengan pengujian statistik tertentu untuk mengkonfirmasi.

$$\chi^2 = \sum_{i=1}^n \frac{y_i - \hat{y}_i}{\hat{y}_i}$$

$$\phi^2 = \frac{\chi^2}{db}$$

dengan  $db = n - k$  di mana  $k$  adalah jumlah parameter regresi,  $n$  adalah jumlah pengamatan,  $y_i$  adalah nilai aktual dari respon pengamatan ke- $i$ ,  $\hat{y}_i$  adalah nilai estimasi dari respon pengamatan ke- $i$ , dan  $\chi^2$  adalah nilai Pearson Chi Square.

### 2.4 Regresi Binomial Negatif

Regresi binomial negatif adalah model regresi yang digunakan untuk menganalisis hubungan antara variabel respon berupa data hitungan dengan satu atau lebih variabel independen. Model ini dapat diterapkan baik dalam kondisi equidispersi maupun overdispersi (Fadil 2021). Parameter dispersi ( $\alpha$ ) memiliki sebaran gamma yang bertujuan mengakomodasi kelebihan ragam dari pengamatan sehingga lebih fleksibel untuk memodelkan data yang mengalami overdispersi. Fungsi massa peluang dari distribusi binomial negatif adalah sebagai berikut:

$$f(y; \mu, \alpha) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \alpha\mu_i}\right)^{y_i}$$

Keterangan:

$f(y; \mu, \alpha)$  = fungsi massa peluang sebaran binomial negatif

$y_i$  = nilai dari peubah respon untuk pengamatan ke-i

$\mu_i$  = rata-rata dari peubah respon Y untuk pengamatan ke-i

$\alpha$  = parameter dispersi

Persamaan model regresi binomial negatif mirip dengan regresi Poisson karena keduanya menggunakan fungsi penghubung (link function) yang sama, yaitu fungsi penghubung log. Model regresi binomial negatif dapat dituliskan sebagai berikut:

$$\ln(\mu_i) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = x_i' \beta$$

## 2.5 Pemilihan Model Terbaik

Satu pendekatan untuk memilih model yang optimal adalah melalui penggunaan Kriteria Informasi Akaike (AIC) dan Kriteria Informasi Bayes (BIC). Sebuah model dianggap unggul jika memiliki nilai AIC dan BIC yang lebih rendah dibandingkan dengan model-model lainnya. Menurut Hilbe (2011), perhitungan nilai AIC dan BIC dapat dilakukan menggunakan rumus sebagai berikut:

$$AIC = -2Lp + 2p \text{ dan } BIC = -2Lp + p \ln(n)$$

Keterangan:

$Lp$  = nilai fungsi kemungkinan maksimum *log - likelihood* dengan  $p - 1$  peubah penjelas

$p$  = banyaknya parameter

$n$  = banyaknya amatan

## 3. Metodologi

### 3.1 Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang didapat dari Badan Pusat Statistik (BPS) mengenai kejahatan, tingkat pendidikan, jenis kelamin, kependudukan, kemiskinan, penghunian hotel, dan ketenagakerjaan. Data terdiri dari satu variabel respon dan delapan variabel penjelas dari setiap provinsi di Indonesia pada tahun 2021. Amatan pada penelitian ini berjumlah sebanyak 34 provinsi di Indonesia. Pemilihan variabel dalam penelitian didasarkan pada penelitian terdahulu yang telah dilakukan. Secara lebih rinci peubah yang digunakan pada penelitian ini disajikan pada tabel 1.

Tabel 1: Daftar Peubah yang Digunakan

Kode	Peubah	Satuan	Referensi
Y	Jumlah Kejahatan yang Dilaporkan	Jumlah	(Handayani 2017)
X1	Tingkat Penyelesaian Pendidikan SMA	Persen	(Maulana 2014)
X2	Pengangguran Terbuka	Persen	(Effendi dan Julia 2021)
X3	Upah Minimum Penduduk	Rupiah	(Erlangga 2019)
X4	Persentase Penduduk Mengakses Internet	Persen	(Millah 2020)
X5	Rasio Jenis Kelamin Penduduk	Rasio	(Sudarsono 2012)
X6	Tingkat Penghunian Kamar Hotel	Persen	(Sudarsono 2012)
X7	Jumlah Penduduk	Jiwa	(Handayani 2017)
X8	Jumlah Penduduk Miskin	Jiwa	(Lumenta <i>et al.</i> 2012)

### 3.2 Metode Penelitian

Prosedur analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Mengeksplorasi data untuk melihat relasi antar variabel respon dan variabel penjelas. Eksplorasi data dilakukan dengan membuat grafik seperti histogram, scatter plot, dan matriks korelasi (Julaeha *et al.* 2020).
2. Melakukan pemeriksaan multikolinearitas antar peubah penjelas dengan menghitung nilai *Variance Inflation Factor* (VIF). Multikolinearitas dapat dideteksi ketika nilai VIF lebih besar dari 10 (Montgomery *et al.* 2012).
3. Melakukan pemodelan regresi Poisson menggunakan tahapan berikut:
  - a. Melakukan pendugaan parameter  
Pendugaan parameter dikerjakan dengan metode kemungkinan maksimum dengan memanfaatkan formula kemungkinan maksimum (*log-likelihood*) yang diterapkan sebagai berikut:

$$\ln L(y, B) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln(y_i!)$$

Mengestimasi parameter  $\beta$  dapat melalui pengoptimalan persamaan log-likelihood tersebut dengan iterasi metode Newton-Raphson (Putri *et al.* 2021).

b. Pengujian hipotesis dugaan parameter model

Hipotesis parameter model diuji menggunakan uji parsial dengan statistik uji Wald. Hipotesis yang digunakan dalam pengujian parameter secara parsial adalah sebagai berikut:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Dengan statistik uji yang digunakan, yaitu :

$$Z^2 = \left( \frac{\hat{\beta}}{ASE} \right)^2 \approx X^2$$

Pengujian hipotesis dilakukan pada taraf nyata 5%. Tolak  $H_0$  apabila nilai dari *p-value* yang didapat lebih kecil dari taraf nyata 5% (Al'afi *et al.* 2020) atau apabila nilai dari statistik uji yang didapat lebih besar dari nilai kritis yang ditentukan.

4. Menguji overdispersi terhadap data

Pengujian dikerjakan dengan mendapatkan nilai Z dari  $Z^2$  dalam uji Wald yang asimtotik dengan Chi-Square. Uji Wald biasanya dilengkapi dengan nilai *p-value* ketika dijalankan dalam perangkat lunak. Apabila nilai *p-value* yang didapat kurang dari 0,5 atau 0,1 maka dapat dikatakan terjadi overdispersi pada data (McCullagh dan Nelder 1989). Jika terjadi overdispersi maka dilanjutkan ke-langkah 5, selainnya dilanjutkan ke-langkah 6.

5. Melakukan pemodelan regresi binomial negatif menggunakan tahapan berikut:

a. Melakukan pendugaan parameter

Pendugaan parameter dikerjakan dengan metode kemungkinan maksimum dengan memanfaatkan formula kemungkinan maksimum (*log-likelihood*) yang diterapkan sebagai berikut:

$$\ln L(y, B) = \sum_{i=1}^n y_i \ln(\mu_i) - \sum_{i=1}^n \mu_i - \sum_{i=1}^n \ln(y_i!)$$

Mengestimasi parameter  $\beta$  dapat melalui pengoptimalan persamaan log-likelihood tersebut dengan iterasi metode Newton-Raphson (Putri *et al.* 2021)

b. Pengujian hipotesis dugaan parameter model

Hipotesis parameter model diuji menggunakan uji parsial dengan statistik uji Wald. Hipotesis yang digunakan dalam pengujian parameter secara parsial adalah sebagai berikut:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Dengan statistik uji yang digunakan, yaitu :

$$Z^2 = \left( \frac{\beta^c}{ASE} \right)^2 \approx \chi^2$$

Pengujian hipotesis dilakukan pada taraf nyata 5%. Tolak  $H_0$  apabila nilai dari  $p$ -value yang didapat lebih kecil dari taraf nyata 5% (Al'afi *et al.* 2020) atau apabila nilai dari statistik uji yang didapat lebih besar dari nilai kritis yang ditentukan.

6. Memilih model terbaik berdasarkan nilai AIC dan BIC.

Model yang terbentuk dari regresi Poisson dan regresi binomial negative dievaluasi dengan melihat nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criteria* (BIC). Apabila suatu model memiliki nilai AIC dan BIC terkecil di antara model lainnya dapat dikatakan sebagai model yang baik (Hilbe 2011).

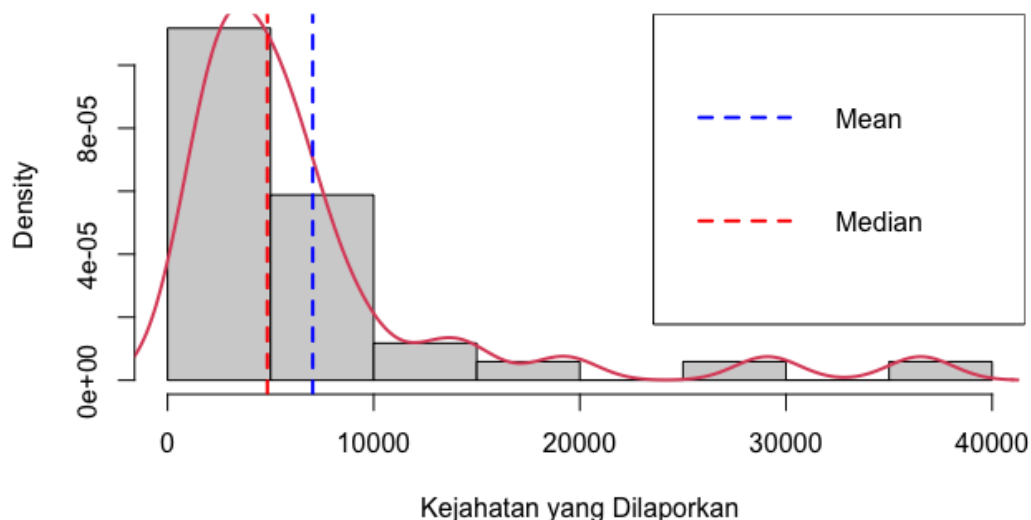
7. Membuat interpretasi berdasarkan model terbaik

Interpretasikan peubah-peubah yang berpengaruh signifikan pada model yang terpilih terhadap jumlah kejahatan di Indonesia dengan taraf nyata 5%.

## 4. Hasil dan Pembahasan

### 4.1 Eksplorasi Data

Histogram jumlah kejahatan (Gambar 1) yang dilaporkan menunjukkan bahwa sebarannya menjuur ke kiri dengan nilai mean lebih besar dari median sehingga peubah Y menyebar Poisson. Dari eksplorasi ini, terlihat bahwa dapat digunakan analisis regresi Poisson dengan peubah responnya adalah variabel jumlah kejahatan yang dilaporkan (Y).



Gambar 1: Histogram peubah respon

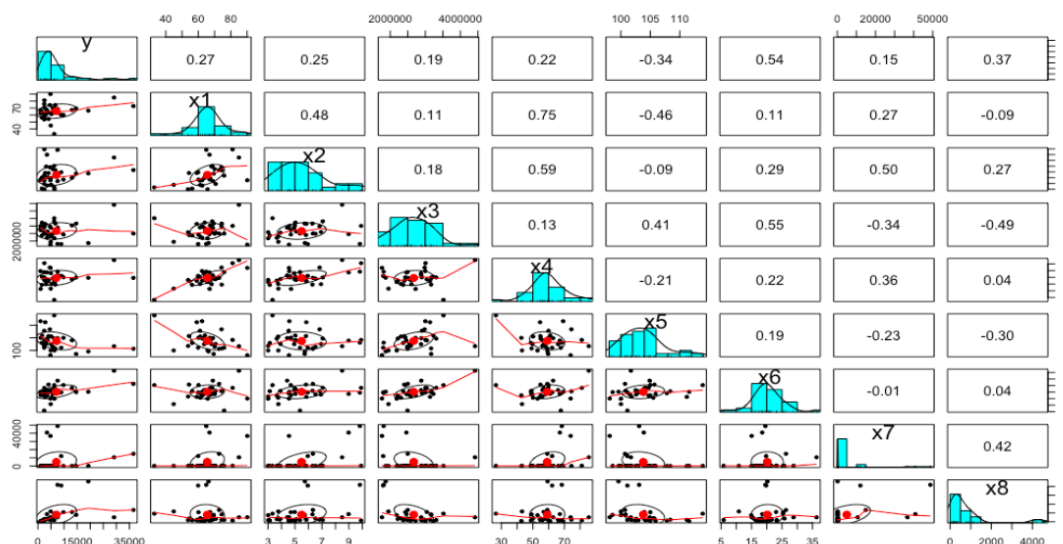
Statistika deskriptif dari data peubah respon dapat dilihat dari tabel 2. Jumlah laporan kejahatan tertinggi pada tahun 2021 ada di angka 36534 di



provinsi Sumatera Utara dan jumlah laporan kejahatan terendahnya ada di provinsi Kalimantan Utara dengan jumlah laporan 971. Jumlah laporan kejahatan dalam suatu provinsi memiliki rata-rata 7044 dan median 4843 sedangkan ragamnya 59321685. Nilai ragam yang lebih besar daripada nilai mean mengindikasikan adanya overdispersi pada model regresi Poisson.

Tabel 2: Statistika Deskriptif	
Statistika Deskriptif	Nilai
Nilai Minimum	971
Kuartil Pertama	2557
Median	4842
Rata-rata	7044
Kuartil Ketiga	7289
Nilai Maksimum	36534
Ragam	59321685

Peubah Y tidak berkorelasi tinggi secara linear dengan semua peubah X. Peubah X1 dan X4 mempunyai nilai korelasi tinggi, yaitu 0,75. Selain X1 dan X4 tidak ada peubah X yang saling berkorelasi tinggi secara linear. Tingginya nilai korelasi antar peubah penjelas mengindikasikan adanya multikolinearitas sehingga perlu dipastikan dengan nilai VIF.



Gambar 2: Scatter plot, histogram, dan matriks korelasi

#### 4.2 Pemeriksaan Multikolinearitas

Digunakan analisis regresi linear berganda dan dihitung nilai VIF untuk memastikan ada atau tidaknya multikolinearitas. Nilai VIF > 10 menunjukkan

adanya multikolinearitas dan perlu ditangani dengan baik sehingga tidak mengganggu model. Dari tabel 3, didapatkan tidak ada peubah yang memiliki nilai  $VIF > 10$ , sehingga tidak terdapat multikolinearitas. Analisis regresi Poisson dapat dilanjutkan tanpa penanganan apapun dan semua peubah dapat dimasukkan ke dalam model.

Tabel 3: Hasil Pemeriksaan Multikolinearitas

Parameter	Koefisien	Nilai t	Nilai p	VIF
$\beta_0$	0,0005958	1,379	0,1802	
$\beta_1$	0,0128	0,720	0,4780	3,672870
$\beta_2$	-0,07417	-0,891	0,3816	2,414976
$\beta_3$	0,005195	1,697	0,1021	3,061541
$\beta_4$	-0,536	-0,364	0,7187	2,849970
$\beta_5$	-0,08002	-2,087	0,0472 *	2,018470
$\beta_6$	0,06347	2,563	0,0168 *	1,787325
$\beta_7$	0,04493	0,418	0,6795	1,860813
$\beta_8$	3,082	2,478	0,0203 *	2,174965

#### 4.3 Pemodelan Regresi Poisson

Dengan menggunakan regresi Poisson, pemodelan dilakukan karena peubah respon merupakan data cacahan. Setelah dilakukan pemodelan didapatkan semua peubah penjelas berpengaruh signifikan terhadap peubah respon sedangkan nilai AIC dan BIC besar, yaitu nilai AIC sebesar 46715 dan nilai BIC 46728.36. Dari tabel didapatkan nilai dugaan koefisien sehingga terbentuk model regresi Poisson sebagai berikut:

$$\ln(\mu) = 0,02037 + 0,01266 X_1 - 0,04272 X_2 + 0,0000001641 X_3 - 0,02794 X_4 - 0,1332 X_5 - 0,1332 X_6 + 0,000008324 X_7 + 0,0002511 X_8$$

Tabel 4: Pemodelan Regresi Poisson

Parameter	Koefisien	Nilai Z	Nilai p
$\beta_0$	0,02037	159,24	$2 \times 10^{16} ***$
$\beta_1$	0,01266	29,67	$2 \times 10^{16} ***$
$\beta_2$	-0,04272	-20,38	$2 \times 10^{16} ***$
$\beta_3$	0,0000001641	25,92	$2 \times 10^{16} ***$
$\beta_4$	-0,02794	-84,22	$2 \times 10^{16} ***$
$\beta_5$	-0,1332	-120,23	$2 \times 10^{16} ***$
$\beta_6$	0,1178	184,52	$2 \times 10^{16} ***$
$\beta_7$	0,000008324	34,01	$2 \times 10^{16} ***$
$\beta_8$	0,0002511	104,19	$2 \times 10^{16} ***$

#### 4.4 Pengujian Overdispersi

Dari Uji Z, didapatkan nilai Z sebesar 3,5224 dan *p-value* 0,0002138. Nilai *p-value* yang kurang dari 0,05 dan sangat kecil menunjukkan bahwa  $H_0$  dapat ditolak artinya nilai ragam lebih besar dari rataannya. Nilai dispersi 1405,488, hal ini menunjukkan bahwa ragam data lebih besar sekitar 1405 lipat dari rataannya sehingga sangat jauh dari asumsi Poisson standar.

#### 4.5 Pemodelan Regresi Binomial Negatif

Setelah dibuktikan ada overdispersi dalam model regresi Poisson, dilakukan pemodelan dengan regresi binomial negatif. Dari model regresi binomial negatif, didapatkan 3 peubah penjelas berpengaruh signifikan terhadap jumlah kejahatan yang dilaporkan yaitu X5 berpengaruh negatif terhadap Y, X6 berpengaruh positif terhadap Y, dan X8 berpengaruh positif terhadap Y. Model ini memiliki nilai AIC 645.3 dan BIC 660.5654. Dari tabel didapatkan nilai dugaan koefisien sehingga terbentuk model regresi binomial negatif sebagai berikut:

$$\ln(\mu) = 15,93 + 0,01503 X_1 - 0,05163 X_2 + 0,0000004053 X_3 - 0,01703 X_4 - 0,0954 X_5 + 0,0732 X_6 + 0,000002652 X_7 + 0,000375 X_8$$

Tabel 5: Pemodelan Regresi Binomial Negatif

Parameter	Koefisien	Nilai Z	Nilai p
$\beta_0$	15,93	4,316	0,0000159 ***
$\beta_1$	0,01503	0,991	0,321811
$\beta_2$	-0,05163	-0,726	0,467686
$\beta_3$	0,0000004053	1,551	0,120850
$\beta_4$	-0,01703	-1,355	0,175273
$\beta_5$	-0,0954	-2,914	0,003571 **
$\beta_6$	0,0732	3,462	0,000536 ***
$\beta_7$	0,000002652	0,289	0,772563
$\beta_8$	0,000375	3,533	0,000411 ***

#### 4.6 Perbandingan Model

Tabel 6: Perbandingan Model

Metode	AIC	BIC
Regresi Poisson	46715	46728,36
Regresi Binomial Negatif	645,3	660,5654

Perbandingan model dilakukan dengan membandingkan nilai AIC dan BIC kedua model. Adanya overdispersi pada model regresi Poisson menyebabkan semua peubah penjelas berpengaruh terhadap peubah respon. Hal ini disebabkan oleh adanya underestimate pada ragam dugaan parameter yang menyebabkan uji parsial cenderung tolak  $H_0$ . Nilai AIC dan BIC lebih kecil pada model regresi binomial negatif menunjukkan bahwa model regresi binomial negatif lebih baik (Chandra dan Budi 2020) digunakan untuk mengatasi overdispersi pada regresi Poisson dan memodelkan faktor-faktor yang mempengaruhi jumlah kejahatan di Indonesia.

#### 4.7 Interpretasi Model

Dari permodelan regresi binomial negatif, didapatkan beberapa faktor yang terbukti berpengaruh signifikan terhadap jumlah kejahatan. Faktor-faktor

tersebut adalah rasio jenis kelamin penduduk (X5), tingkat hunian kamar hotel (X6), dan jumlah penduduk miskin (X8). Model regresi yang didapatkan sebagai berikut:

$$\ln(\mu) = 15,93 + 0,01503 X1 - 0,05163 X2 + 0,0000004053 X3 - 0,01703 X4 - 0,0954 X5 + 0,0732 X6 + 0,000002652 X7 + 0,000375 X8$$

Interpretasi dari hasil permodelan regresi binomial negatif pada peubah rasio jenis kelamin penduduk adalah setiap peningkatan 1 perbandingan laki-laki terhadap perempuan akan diikuti oleh penurunan rata-ran jumlah kejahatan sebesar  $\exp(-0,0954) = 0,909$  kali lipat dengan diasumsikan peubah lainnya tetap sedangkan interpretasinya adalah setiap peningkatan 1% tingkat hunian hotel akan cenderung mengalami peningkatan rata-ran jumlah kejahatan sebesar  $\exp(0,0732) = 1,076$  kali lipat dengan diasumsikan peubah lainnya tetap. Pada peubah jumlah penduduk miskin, peningkatan 1 jiwa penduduk miskin diiringi oleh peningkatan rata-ran jumlah kejahatan sebesar  $\exp(0,000375) = 1,001$  kali lipat dengan diasumsikan peubah lainnya tetap.

## 5. Simpulan

Model regresi binomial negatif menghasilkan nilai AIC yang jauh lebih kecil dibandingkan dengan nilai AIC pada model regresi Poisson sehingga model regresi binomial negatif lebih baik daripada regresi Poisson. Faktor yang memengaruhi angka kejahatan di Indonesia berdasarkan hasil analisis adalah rasio jenis kelamin penduduk, tingkat hunian kamar hotel, dan jumlah penduduk miskin. Rasio jenis kelamin penduduk berpengaruh negatif terhadap jumlah kejahatan sedangkan tingkat hunian kamar hotel dan jumlah penduduk miskin berpengaruh negatif terhadap jumlah kejahatan. Model regresi Binomial Negatif menghasilkan model yang lebih baik karena memiliki nilai AIC dan BIC yang lebih kecil dibandingkan regresi Poisson.

## Daftar Pustaka

- Agresti A. 2019. *An Introduction to Categorical Data Analysis Third Edition*. New York (NY): John Wiley & Sons.
- Al'afi AM, Widiarti, Kurniasari D, Usman M. 2020. Peramalan data time series seasonal menggunakan metode analisis spektral. *Jurnal Siger Matematika*. 1(1): 10-15. DOI: 10.23960%Fjsm.v1i1.2484
- Hilbe JM. 2011. *Negatif Binomial Regression Second Edition*. Cambridge: Cambridge University Press.
- Chandra C, Budi S. 2020. Analisis komparatif ARIMA dan prophet dengan studi kasus dataset pendaftaran mahasiswa baru. *Jurnal Teknik Informatika dan Sistem Informasi*. 6(2): 278-287. DOI: 10.28932/jutisi.v6i2.2676

- Effendi I, Julia A. 2021. Faktor ekonomi yang mempengaruhi kejahatan properti di Pulau Jawa tahun 2014-2019. *Journal Riset Ilmu Ekonomi*. 1(1): 41-47. DOI: 10.29313/jrieb.v1i1.172
- Erlangga LT. 2019. Analisis tingkat pengangguran terbuka di Indonesia menggunakan regresi non parameter B-Spline. [skripsi]. Semarang: Universitas Muhammadiyah Semarang.
- Handayani R. 2017. Analisis dampak kependudukan terhadap tingkat kriminalitas di Provinsi Banten. *Jurnal Administrasi Publik*. 8(2): 149-169. DOI: 10.31506/jap.v8i2.3312
- Julaeha S, Kustian N, Parulian D. 2020. Pemetaan tabel relationship dalam visualisasi diagram relasi untuk eksplorasi data pada database. *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*. 5(2): 126-133. DOI: 10.30998/string.v5i2.6653
- Lestari SP, Wulandari SP. 2014. Pemodelan faktor – faktor yang mempengaruhi jumlah kasus tetanus neonatorum (TN) di Jawa Timur dengan metode regresi zero-inflated generalized Poisson. *Jurnal Sains dan Seni Pomits*. 3(2):116-121. McCullagh P, Nelder JA. 1989. *Generalized Linear Models Second E*.
- Lumenta CY, Kekenusa JS, Hatidja D. 2012. Analisis jalur faktor-faktor penyebab kriminalitas di Kota Manado. *Jurnal Ilmiah Sains*. 12(2): 77–83. DOI: 10.35799/jis.12.2.2012.556
- Maulana T. 2014. Pengaruh umur, pendidikan, pendapatan, dan jumlah tanggungan keluarga terhadap tingkat kejahatan pencurian dengan pendekatan ekonomi. [skripsi]. Semarang: Universitas Diponegoro.
- McCullagh P, Nelder JA. 1989. *Generalized Linear Models Second Edition*. London: Chapman & Hall.
- Millah IA. 2020. Penanggulangan kejahatan di masa pandemi covid-19 (dalam perspektif kriminologi dan viktimologi). *Jurnal Komunikasi Hukum (JHK)* Universitas Pendidikan Ganesha. 6(2): 497-513. DOI:10.23887/jkh.v6i2.28099
- Montgomery DC, Peck EA, Vining GG. 2012. *Introduction to Linear Regression Analysis, Fifth Edition*. New Jersey (US): John Wiley and Sons, Inc.
- Putri DLW, Mariani S, Sunarmi. 2021. Peningkatan ketepatan klasifikasi model regresi logistik biner dengan metode bagging (bootstrap aggregating). *Indonesian Journal of Mathematics and Natural Sciences*. 44(2): 61-72. DOI:10.15294/ijmns.v44i2.33144

Sudarsono DB. 2012. Identifikasi faktor-faktor yang berpengaruh terhadap banyaknya kejahatan dengan pendekatan analisis spasial. [skripsi]. Bogor: Institut Pertanian Bogor.

United Nations. 2015. Promote Peaceful and Inclusive Societies for Sustainable Development, Provide Access to Justice for All and Build Effective, Accountable and Inclusive Institutions at All Levels. Department of Economic and Social Affairs Sustainable Development.